Faculty of Electronics and Computer Engineering

LOCAL-BASED STEREO MATCHING ALGORITHM USING MULTI-COST PYRAMID FUSION, HYBRID RANDOM AGGREGATION AND HIERARCHICAL CLUSTER-EDGE REFINEMENT

Ahmad Fauzan Kadmin

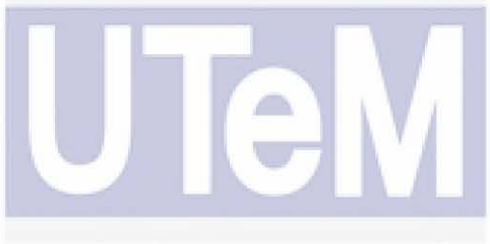
Doctor of Philosophy

2023

LOCAL-BASED STEREO MATCHING ALGORITHM USING MULTI-COST PYRAMID FUSION, HYBRID RANDOM AGGREGATION AND HIERARCHICAL CLUSTER-EDGE REFINEMENT

AHMAD FAUZAN KADMIN

A thesis submitted
in fulfilment of the requirements for the degree of Doctor of Philosophy



Faculty of Electronics and Computer Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I declare that this thesis entitled "Local-Based Stereo Matching Algorithm using Multi-Cost Pyramid Fusion, Hybrid Random Aggregation and Hierarchical Cluster-Edge Refinement" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name : Ahmad Fauzan Kadmin

Date : 16 August 2023

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : .....

Supervisor Name : Dr. Rostam Affendi Hamzah

Date : 18 August 2023

DEDICATION

This thesis is dedicated to God Almighty, my creator, my pillar of strength, and my source of inspiration, wisdom, knowledge, and comprehension. He has been my source of strength throughout this journey, and I have only soared on His wings. This work is also dedicated to my parents, Kadmin and Halijah, who have always loved me unconditionally and whose exemplary behaviour has inspired me to work diligently towards my goals. To my wife, Nor Idayu, who has been a steady source of support and encouragement. I am very thankful for having you in my life. To my children, Ahmad Imran, Sofea Nafeesa, Sofea Imani, and Sofea Aralynn, whose existence have been profoundly impacted by this endeavour. Thank you very much. My affection for you all can never be quantified. God bless you.

ABSTRACT

The estimation of Stereo Matching Algorithm (SMA) is one of the extensive research topics for obtaining the disparity map from two images. The depth measurement provided by the stereo matching framework is used to rebuild the three-dimensional coordinates of point and object detection. Similar to how human eyes and binocular vision perceive depth, the visual disparity information obtained from this pair of images captured by the cameras represents the impression of perceived depth. The stereo vision algorithm computes disparity using local, global, and semiglobal optimisation methods established by the researcher. However, the computation needed for the creation of SMA is more difficult, particularly for images comprising complex scenes. The influencing factors include low-texture regions, repetitive patterns, illumination variation, depth discontinuity, and occlusion. Several issues have been challenges to researchers, especially for local methods, such as producing an accurate correspondence between pixels that lie around the boundaries due to different illumination conditions. Besides that, window-based approaches and pixel-based intensity comparison between central pixels and neighbour pixels may cause problems at incorrect disparities, while similar matching costs at low textures cannot be efficiently solved with increasing window aggregation size or implementation of global optimisation. Therefore, this thesis proposes a local-based SMA that enhances the accuracy of complex regions detection by focusing on these issues. The four stages of the proposed SMA were centred on the matching cost computation. The first stage comprised of Tr uncated Absolute Differences (TAD), Gradient Magnitude CLAHE (GMC), and Modified Census Edge (MCE), which were then combined through Planar Pyramid Fusion (PPF) to obtain the initial cost volume. Then, a new proposed cost aggregation based on the Hybrid Random Aggregation (HRA) was implemented that utilized modified Iterative Non-Local Guided Filter (iNLGF), Simple Linear Iterative Clustering (SLIC), Graph Segmentation (GS) and Extended Restart Random Walk (eRWR) for error reduction. Next, a Winner-Take-All (WTA) approach was used to select the location of minimum aggregated value corresponding to the disparity value for each pixel. During the refinement stage, the Left-Right (LR) consistency checking process and the Confidence Disparity Filling (CDF) were conducted. Then, the K-means clustering, and Side Window Filter (SWF) were used to recover the low texture and to remove the remaining noises. In this thesis, the accuracy of the proposed algorithm was evaluated using two standard online benchmarking database systems. For the quantitative and qualitative assessments, systems from the Middlebury Stereo, the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI), and actual images from UTeMLab-Stereo were applied. Once accurate results were achieved, the proposed SMA's disparity maps were generated to be used for the 3D surface reconstruction. As a result, the proposed SMA was able to deliver accurate validation process with findings from the Middlebury system showing 5.11% *nonocc* error and 9.02% *all* error and the KITTI system showing 7.90% *nonocc* error and 7.07% *all* error. Therefore, the proposed framework is proven to be competitive with other established methods and can be used as a complete algorithm

ALGORITMA PEMADANAN STEREO BERASASKAN-SETEMPAT MENGGUNAKAN BERBILANG-KOS LAKURAN PIRAMID, PENGAGREGATAN RAWAK HIBRID DAN PENGHALUSAN KELOMPOK-SISI BERHIERARKI

ABSTRAK

Anggaran Algoritma Pemadanan Stereo (APS) adalah salah satu topik penyelidikan yang meluas untuk mendapatkan peta perbezaan daripada dua imej. Pengukuran kedalaman rangka kerja pemadanan stereo digunakan untuk membina semula koordinat tiga dimensi pengesanan titik dan objek. Sama seperti cara mata manusia dan penglihatan binokular melihat kedalaman, maklumat perbezaan visual yang diperoleh daripada sepasang imej yang ditangkap oleh kamera ini mewakili kesan kedalaman yang dirasakan. Algoritma pemadanan stereo menentukan peta perbezaan menggunakan kaedah pengoptimum setempat, global dan separa global yang ditetapkan oleh penyelidik. Walau bagaimanapun, pengiraan yang diperlukan untuk penciptaan APS adalah lebih sukar, terutamanya untuk imej yang terdiri daripada adegan yang kompleks. Faktor seperti kawasan bertekstur rendah, corak berulang, variasi pencahayaan, ketakselanjaran kedalaman dan oklusi. Beberapa isu telah menjadi cabaran kepada penyelidik-penyelidik terutamanya pada kaedah-kaedah setempat bagi mendapatkan persamaan yang tepat di antara piksel-piksel yang berada di antara sempadan disebabkan keadaan perbezaan pencahayaan. Selain itu, pendekatan berasakan-tetingkap dan berasaskan-piksel di antara piksel pusat dan piksel kejiranan mungkin menyebabkan peta pembezaan yang salah, manakala kos pemadanan yang serupa di tekstur rendah tidak dapat diselesaikan secara cekap melalui peningkatan saiz tetingkap pengagregatan atau perlaksanaan pengoptimum global. Oleh itu, tesis ini mencadangkan APS berasaskan setempat yang meningkatkan ketepatan pengesanan wilayah kompleks dengan mengfoukuskan isu-isu ini. APS yang dicadangkan terdiri daripada empat peringkat, bermula dengan pengiraan kos pemadanan. Langkah pertama terdiri daripada Perbezaan Mutlak Terpenggal (PMT), Magnitud Kecerunan CLAHE (MKC) dan Pinggir Banci Terubahsuai (PBT), yang kemudiannya digabungkan melalui Piramid Lakuran Satah (PLS) untuk mendapatkan isipadu kos permulaan. Kemudian, cadangan baharu pengagregatan kos berdasarkan Pengagregatan Rawak Hibrid (PRH) dilaksanakan yang menggunakan Penapis Berpandu Bukan Setempat Berulang (PBBSB), Pengelompokan Lelaran Linear Mudah (PLLM), Peruasan Rajah (PR) dan Jalan Mula Semula Rawak Tambahan (JMSRT) untuk pengurangan ralat. Selepas itu, pendekatan Pemenang-Ambil-Semua (PAS) dilaksanakan untuk memilih lokasi nilai agregat minimum yang sepadan dengan nilai jurang bagi setiap piksel. Kemudian, peringkat pemurnian dilaksanakan melalui proses semakan konsistensi Kiri-Kanan (KK) dan Pengisian Ketaksamaan Keyakinan (PKK). Kelompok K-means, dan Penapis Tepi Tetingkap (PTT) dilaksanakan untuk memulihkan tekstur rendah dan mengeluarkan kebisingan yang tinggal. Dalam tesis ini, ketepatan algoritma yang dicadangkan dinilai menggunakan dua sistem pangkalan data penanda aras piawai dalam talian. Untuk penilaian kuantitatif dan kualitatif, sistem ini adalah daripada Middlebury Stereo, Institut Teknologi Karlsruhe dan Institut Teknologi Toyota (KITTI) dan imej sebenar dari UTeMLab-Stereo. Setelah keputusan yang tepat dicapai, peta perbezaan APS yang dicadangkan dijana untuk digunakan untuk pembinaan semula permukaan 3D. Kesimpulannya, APS yang dicadangkan menyampaikan penemuan proses pengesahan yang tepat. Hasilnya ialah 5.11% untuk ralat bukan nonocc dan 9.02% untuk semua ralat dari Middlebury dan 7.90% untuk ralat bukan nonocc dan 7.07% untuk semua ralat daripada KITTI. Ia menunjukkan rangka kerja yang dicadangkan boleh

digunakan sebagai algoritma yang lengkap dan berdaya saing dengan kaedah-kaedah yang sedia ada.

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious, the Most Merciful.

First and foremost, I would like to thank and praise Allah the Almighty, my Creator and Sustainer, for everything I received since the beginning of my life. I would like to extend my appreciation to the Universiti Teknikal Malaysia Melaka (UTeM) for providing the research platform. Thank you also to the Malaysian Ministry of Higher Education (MOHE) for the financial assistance through the Skim Latihan Bumiputera (SLAB) scholarship.

My utmost appreciation goes to my main supervisor, Dr. Rostam Affendi Bin Hamzah, Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka (FTKEE, UTeM), for all his support, advice, and inspiration. His unwavering patience in guiding and providing invaluable insights will be remembered forever. Also, to my co-supervisor, Assoc. Prof. Dr. Nurulfajar Bin Abd Manap, Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (FKEKK, UTeM), who constantly supported my journey. My special thanks go to my research group and colleagues for all the help and support I received from them.

Last but not least, from the bottom of my heart, I am grateful to my beloved wife, Dr. Nor Idayu Binti Ahmad Ruslan, for her encouragement and for being the pillar of strength in all my endeavours. My eternal love also goes to all my children, Ahmad Imran Aqil, Sofea Nafeesa, Sofea Imani, and Sofea Aralynn, for their patience and understanding. I would also like to thank my beloved parents for their endless support, love, and prayers. Finally, thank you to all the individual(s) who provided me with assistance, support, and inspiration to embark on my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATION

1D	-	1 Dimensional
2D	-	2 Dimensional
3D	-	3 Dimensional
AGF	-	Adaptive Support Weight
AI	-	Artificial Intelligence
ANCC	-	Adaptive Normalized Cross Correlation
ASW	-	Adaptive Support Weight
BF	-	Bilateral Filter
BMP	-	Bad Matched Pixels
BP	-	Belief Propagation
CA	-	Cross Aggregation
CDF	-	Confidence Disparity Filling
CLAHE	-	Contrast Limited Adaptive Histogram Equalization
CNN	-	Convolutional Neural Network
CRF	-	Conditional Random Field
CT	-	Census Transform
CTF	-	Coarse-To-Fine
DL	-	Deep Learning
DP	-	Dynamic Programming
ELAS	-	Efficient Large-Scale Stereo

ERWR	-	Extended Random Walk Restart
FPGA	-	Field Programmable Gate Array
GA	-	Genetic Algorithm
GF	-	Guided Filter
GP	-	Gaussian Pyramid
GCP	-	Graph Cut Programming
GCS	-	Ground Control Surfaces
GMC	-	Gradient Magnitude CLAHE
GMD	-	Gradient Magnitude Differences
GMM	-	Gaussian Mixture Model
GPU	-	Graphics Processing Unit
GS	-	Graph Segmentation
HRA	-	Hybrid Random Aggregation
IGCM	-	Intensity Guided Cost Metric
IGF	-	Iterative Guided Filter
IGG-MSB	-	Iterative Guided Gaussian Multi-Baseline
IGF	-	Iterative Non-Local Guided Filter
IVBP	-	Initial Value Belief Propagation
MBS	-	Multi-Baseline
MC	-	Matching Cost
MCE	-	Modified Census Edge
ML	-	Machine Learning
MPF	-	Multi-Cost Pyramid Fusion
MST	-	Minimum Spanning Tree
NCC	-	Normalized Cross Correlation

ND	-	Nonlinear Diffusion
OLT	-	Oriented Linear Tree
PGIF	-	Pervasive Guided Image Filter
PLSP	-	Probabilistic Laplacian Surface Propagation
PPF	-	Planar Pyramid Fusion
QCT	-	Quaternary Census Transform
RAM	-	Random Access Memory
RGB	-	Red Green Blue
RL	-	Reinforcement Learning
RWR	-	Random Walk Restart
SA	-	Simulated Annealing
SAD	-	Sum of Absolute Differences
SCT	-	Star Census Transform
SDDT	-	Self-Adapting Dissimilarity Data Term
SLIC	-	Simple Linear Iterative Clustering
SGM	-	Semiglobal Matching
SMA	-	Stereo Matching Algorithm
SMV	-	Single-View Videos
SR	-	Spearman Rank
SWF	-	Side Window Filter
TAD	-	Truncated Absolute Differences
WLS	-	Weighted Least Squares
WTA	-	Winner-Takes-All

LIST OF SYMBOLS

w	-	Support Window
r	-	Radius of Window
I_l	-	Left Pixel Intensity
I_r	-	Right Pixel Intensity
d	-	Disparity Range
σ_{AD}	-	Pixel Intensity Differences Weightage
τ_{TAD}	-	Truncated Value TAD
p	-	Coordinate of (x, y)
∇_x	-	Horizontal Directional Gradient Magnitude
∇_y	-	Vertical Directional Gradient Magnitude
W_{GM}	-	Sobel Operator Support Window
τ_{GM}	-	Truncated Value GMC
τ_{edge}	-	MCE Threshold Parameter
w_{CE}	-	MCE Support Window
τ_{diff}	-	MCE Error Threshold
ξ	-	MCE Mapping Function
τ_{CN}	-	Truncated Value MCE
σ_{CN}	-	Census Texture Weightage
σ_{ED}	-	Census Edge Weightage

k	-	Pyramid Layers
ω	-	Pyramid Weighting Function
σ_{LM}	-	MCE Balancing Parameter
σ_{LG}	-	GMC Balancing Parameter
σ_{LT}	-	TAD Balancing Parameter
Ω	-	Area of Cost Volume
n_{NL}	-	iNLGF Iteration
$f(p,q)$	-	iNLGF Weighting Function
sd	-	iNLGF Standard Deviation
w_q	-	iNLGF Comparison Support Window
w_p	-	iNLGF Search Support Window
w_g	-	iNLGF Main Support Window
ε	-	iNLGF Smoothness Term
n_s	-	Number of Pixels in SLIC
s	-	Number of SLIC Clusters
W	-	Graph Matrix
σ_e	-	Graph Segmentation Weightage
τ_e	-	Graph Segmentation Truncated Value
D	-	Diagonal Matrix
τ_Υ	-	Penalty Function Truncated Value
σ_Υ	-	Penalty Function Scaling Parameter
γ	-	Segment Cost Weightage
τ_{LR}	-	Left Right Error Threshold
w_p	-	CDF Median Window

τ_{CF}	-	Outlier Threshold
h	-	K-means Iteration
w_h	-	K-means Window Size
u	-	Number of K-means Clusters
θ	-	SWF Position
m	-	SWF Side Parameter Filter Kernel
n_r	-	SWF Iteration
ρ	-	SWF Target Pixel Position

LIST OF PUBLICATIONS

Journal Articles

Kadmin, A. F., Hamzah, R. A., Manap, N. A. and Hamid, M. S., 2021. A New Pre-Processing Technique for Computational of Stereo Matching Algorithm. *Advances in Mathematics: Scientific Journal*, 10(2), pp. 743-758. (Scopus)

Kadmin, A. F., Hamzah, R. A., Manap, N. A., Hamid, M. S. and Tg. Wook, T. F., 2021. Local Stereo Matching Algorithm using Modified Dynamic Cost Computation. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS),* 22(3), pp. 1312–1319. (Scopus)

Kadmin, A. F., Hamzah, R. A., Manap, N. A., Hamid, M. S. and Abd Gani S. F., 2021. Improved Stereo Matching Algorithm based on Census Transform and Dynamic Histogram Cost Computation. *International Journal of Emerging Technology and Advanced Engineering (IJETAE),* 11(8), pp. 48–57. (Scopus)

Ahmad Fauzan, Rostam Affendi, Nurulfajar, Mohd Saad, Nadzrie, and Tg. Mohd Faisal, 2021. New Stereo Vision Algorithm Composition Using Weighted Adaptive Histogram Equalization and Gamma Correction. *Journal of ICT Research and Applications (JICTRA),* 15(3), pp. 239–250. (Scopus)

Kadmin, A. F., Hamzah, R. A., Manap, N. A., Hamid, M. S., Mohamood, N. and Tengku Wook T. M. F., 2022. Enhancement of Digital Stereo Vision Images based on Histogram and Gamma Correction Strategy. *Iraqi Journal of Science*, 63(1), pp. 313–323. (Scopus)

Conference Articles

Kadmin, A. F., Hamzah, R. A., Manap, N. A., Hamid, M. S., Mohamood, N. and Aziz, K. A. A., 2021. Improvement of Disparity Measurement of Stereo Vision Algorithm using Modified Dynamic Cost Volume. *The 6th International Conference on Electrical, Control and Computer Engineering (InECCE2021)*. Pahang, Malaysia, 23 August 2021. UMP.

Kadmin, A. F., Hamzah, R. A., Manap, N. A., Hamid, M. S., Abd Gani S. F. and Aziz, K. A. A., 2021. Adapted Semiglobal Dynamic Cost Volume Solution for Stereo Matching Framework. *Proceedings of Malaysian Technical Universities Conference on Engineering and Technology (MUCET2021)*. Malacca, Malaysia, 16 – 18 November 2021. MTUN.

AWARDS AND SCHOLARSHIPS

Awards:

2021

Best Paper Award – Improvement of Disparity Measurement of Stereo Vision Algorithm using Modified Dynamic Cost Volume. The 6th International Conference on Electrical, Control and Computer Engineering (InECCE2021). Pahang, Malaysia, 23 August 2021.

Scholarships:

2019-2022

Skim Latihan Bumiputera (SLAB), Ministry of Higher Education, Malaysia.

CHAPTER 1

INTRODUCTION

This chapter provides a brief overview of the stereo matching algorithm and focuses on the appealing qualities that may be used to develop new or improved algorithms for depth map evaluation in computer vision applications. It also describes the goals and reasons for the study that is being presented. Furthermore, the thesis' original contributions are emphasised, followed by the thesis' structure and organisation.

1.1 Background

Computational stereo vision is a topic of pivotal importance in computer vision that is now rapidly becoming an attention and potential area for further research by academics all around the world. One of the recent new areas for investigation has been the field of the stereo vision system, which is conceptually similar to human's two fontal-parallel eyes, allowing them to perceive the world from different perspectives. The motivation of this research is to better understand the information from the environment by interpreting the digital images captured from the outside or within the camera. These images or views are then combined to recover the 3D information about the environment. There is a growing field of research being undertaken using this stereo vision system that contributes fundamentally to the computer vision technology development. There are two directions in the stereo vision approaches. The first is the neurophysiology approach, which focuses on the biological function of the visual cortical cells and the neural pathways. The second approach uses computer hardware and software to create the various model stereo vision

system in order to investigate the role of stereoscopic vision (Y. Liu and Aggarwal, 2005). Therefore, many recent advances have been focused on the computational stereo vision of estimating depth from digitized images of the world using computer programs. Huge advances in software and computer technology have made this feasible in practice to reconstruct the three-dimensional scenes from the depth estimation of the stereo vision.

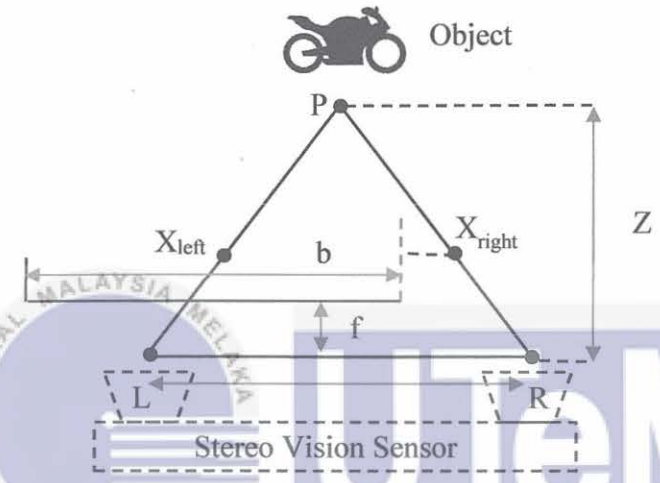


Figure 1.1: Stereo Vision Depth Perception for Object Detection Based on Triangulation Principle

The fundamental of stereoscopic vision is an epipolar geometry mechanism to produce a disparity map based on the establishment of the correspondence between two images. The disparity map accuracy level is a significant issue that has been the focus of stereo matching research for more than a decade. This map contains depth information obtained from the stereo matching framework that is used for the reconstruction of three-dimensional coordinates of point and object detection. The stereo vision system's architecture consists of a stereo camera and a minimum of two cameras that are placed horizontally, left and right, next to each other. The visual disparity information obtained

from these two camera-captured images represents the impression of a scene's perceived depth, similar to human eyes and binocular vision views. The basic architecture of the stereo vision is shown in Figure 1.1.

The distance to the object P can be computed based on the triangulation principle or 2D perception to estimate the depth Z, as implemented by Singh, Kumar and Nongmeikapam (2020).

$$\frac{b}{Z} = \frac{(b+X_{right})-X_{left}}{Z-f}, \tag{1.1}$$

where X_{right} is the right plane coordinate, X_{left} is the left plane coordinate, Z represents the depth value, f for the focal length while b is the baseline. The depth value Z can be obtained using Eq. (1.2) and Eq. (1.3).

$$d' = X_{left} - X_{right}, \tag{1.2}$$

$$Z = \frac{bf}{d'}, \tag{1.3}$$

where d' presents the pixels distance between coordinates at the left and right positions.

Stereo Matching Algorithm (SMA) has emerged over the past several years as a promising and versatile method in computer vision system for acquiring an accurate disparity map for depth measurement. Equation (1.3) demonstrates that this map contains the depth information and has currently become a standard practice by many stereo vision systems to use this information to establish the disparity maps. However, the development of the stereo matching algorithm requires several stages of taxonomy formulation. For example, Scharstein and Szeliski (2002) introduced a fundamental four-stages stereo matching algorithm taxonomy to acquire the disparity map:

- Stage 1: Initial matching cost computation - Determine the stereo image's correspondence points.
- Stage 2: Cost aggregation - Noise reduction and aggregated the cost volume.

- Stage 3: Disparity optimisation and selection - Select the finest disparity value from the cost function.
- Stage 4: Post-processing and disparity refinement – Final disparity map refinement.

In stereo matching algorithm, researchers have extensively developed two types of significant optimization using the local and global methods. The local method employs correlation measurement in the local windows of stereo images to compute disparity or correspondence. Meanwhile, the global method generally utilises an energy minimization function with various constraints to obtain the disparity value. In addition, the Semiglobal Matching (SGM) proposed by Hirschmüller (2008) is the method of combining the trade-off between the local and global methods for better balance disparity accuracy with acceptable computational complexity.

Finding the best technique to acquire the differences between pixels of the image in stereo pairs, generally referred to as the stereo correspondence, is quite challenging but is an essential step in the generation of an in-depth map. The disparity map acquired by the triangulation principle can be applied for extensive estimation in various applications such as navigation for autonomous driving and robotic guidance used by the Mars Exploration Rover (MER) for autonomous passive stereo vision to detect terrain hazards before driving into them. Then, the disparity map also can be used in-depth estimation from realistic motion of static or dynamic environment for object detection (Ross et al., 2014) for underwater tube object detection by AUV, medical diagnosis (Suenaga et al., 2015) for neurosurgery, selection of an accurate location for an object created from real life video (Zhan et al., 2016), depth range image generation for 3D video content (G. S. Hong and Kim, 2017), virtual reality (Diaz et al., 2017) and 3D reconstruction to determine the status and conditions of an object or environment (Rostam Affendi Hamzah et al., 2018).

1.2 Problem Statement

The stereo matching algorithm constraint is to acquire precise correspondence estimation to interpret the information from stereo images. Figure 1.2 shows the diverse challenges to obtain an accurate disparity map.

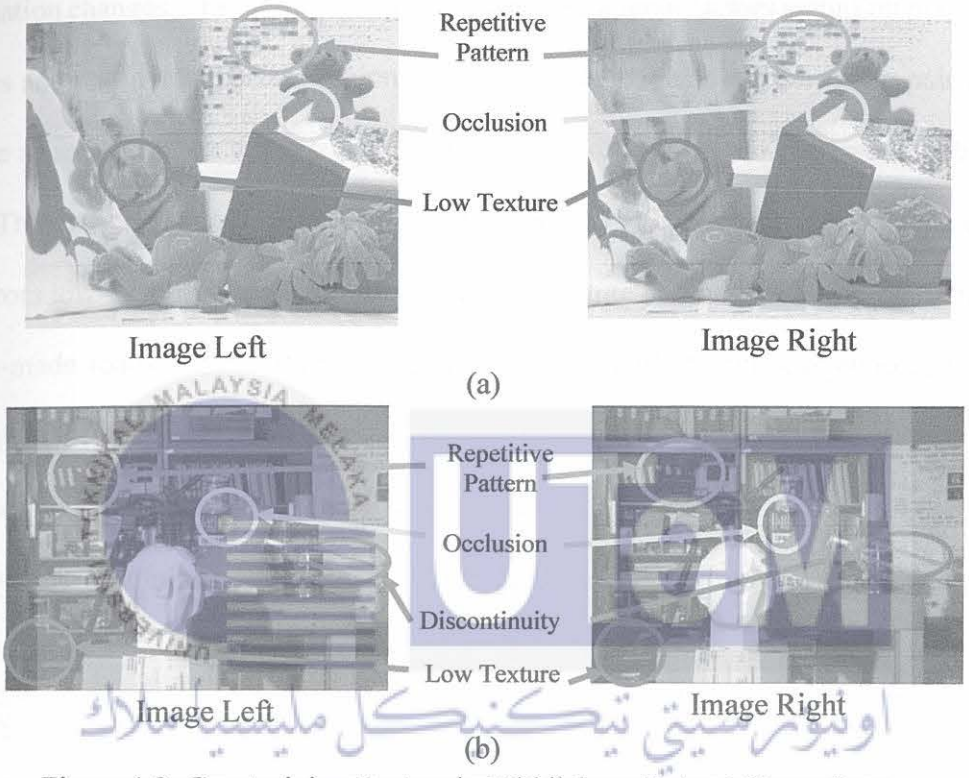


Figure 1.2: Constraining Factors in Middlebury Dataset Stereo Images

a) Teddy b) Tsukuba

Most stereo vision systems consist of a simple matching cost taxonomy, such as block matching to measure the stereo corresponding points taken from two or multiple perspectives. The intensity level between these corresponding points is assumed to be identical to each other. It is difficult to establish an accurate corresponding point between the pixels that lie inside low-textured image areas (Hirschmüller and Scharstein, 2007). The constraint is caused by different illumination conditions, amongst other factors, which need more complex cost functions to account for the radiometric differences. This issue was

explored by Navarro and Buades (2019) using weight distribution in the adaptive support weights method that favours pixels with the same displacement as the reference pixel. The researcher claimed that the framework reduced the fattening effect, and its multi-scale strategy highly decreased errors due to match ambiguities and was robust to additive illumination changes. This problem was caused by a number of factors including plain colour surfaces and textureless surface regions. The low texture regions in the stereo pair images produce almost similar matching costs, thus makes it challenging to compute the disparity value. The low texture or very low contrast stereo pairs occur due to poor image quality taken from low-resolution cameras or caused by the nature of the scene being captured, such as man-made roads and weather covered regions. In highly textureless regions, the final disparity value consists of inaccurate discontinuities, wrong disparity, and blurry boundaries reflected due to poor matching corresponding point. Liu et al. (2021) proposed using two-phase adaptive optimisation of AD-Census and gradient fusion to reduce this problem, which performed better in low texture regions and was robust against radiometric changes and noise.

The second prevalent constraint in the stereo vision system is the occluded areas that significantly contributes to invisible matching. Based on Figure 1.2, the matching corresponding point within the right image of the object is visible but the object itself is not visible in the left image. The constraint is caused by the geometric displacement of the perspective of cameras or sensors, and it gets worse if the area increases due to the expansion of the stereo sensor's baseline. In addition, obtaining an accurate disparity estimation is a challenge since the occluded region is unable to be matched due to the disappearance of the structures, objects or shapes. Bapat and Frahm (2019) discussed the possible algorithm to solve this issue using general MC-CNN optimisation framework that was performed in the pixel space with an edge sensitive regulariser. The method delivered a significant disparity

improvement for non-occluded pixels and outperformed a number of well-established algorithms with additional computational cost.

Further constraint to be taken into consideration is the repetitive or periodic surface areas. This constraint, which has repeatedly been encountered in this research area, occurs normally due to the fact that both space objects and man-made objects have many repetitive textures. Therefore, it is generally challenging to obtain the accurate value of disparity caused by wrong matching coordinates. This is a common situation when dealing with this type of constraint and is caused by the algorithm trying to find the exact match of pixels between the pair images which comprises of many possible intensity values allocations. The work of Kong et al. (2021) proposed a theoretical algorithm for a much more robust disparity value when dealing with illumination or exposure changes which was aimed to improve the repetitive regions based on multi-cost matching cost and adaptive cross-region guided image filtering with orthogonal weights.

A final constraint is the depth discontinuities that lead to the distortion across the depth boundaries. The discontinuity region sizes are important since a bigger difference in size between the two regions contributes to the challenge in the stereo correspondence measurement. This is a typical constraint when designing a stereo vision system and is caused by the algorithm using a predetermined mask size from the reference image to localise within the target image. One way to resolve this is to assign several correct disparity values if the mask contains information from the front-most surface and the rear-most surface across a depth discontinuity. For instance, Mozerov and Van De Weijer (2015) developed a two-step energy-minimisation algorithm using two Markov random field models to enhance the disparity map accuracy in the occluded regions and the depth discontinuities.

The key idea of this thesis is to formulate a novel matching algorithm for stereo correspondence measurement to acquire accurate disparity results. This new stereo matching algorithm formulas will considerably advance the field of depth measurement. Although work on stereo matching algorithm has been ongoing for several years, the constraints in low texture regions, discontinuity and occluded regions continue to be an open problem and a long-standing challenge for the research community that affects the stereo vision system development (Bleyer and Breiteneder, 2013). The difference between stereo vision and other image processing methods is that this method addresses the stereo correspondence issues in the displacement between points corresponding to the same object produced by the perspective differences between the reference and target, which can be translated to depth calculation, while other image processing methods only work on a single surface in 2D. The basic stereo vision system used a simple matching cost function, such as block matching, to measure the corresponding points in stereo images taken from two or multiple perspectives (Kok and Rajendran, 2019). The intensity level between these corresponding points is assumed to be identical to each other. It is challenging to establish an accurate correspondences between pixels that lie around the boundaries due to different illumination conditions, amongst other factors, which require more complex cost functions to account for radiometric differences (Chang and Ho, 2019).

Census Transform (CT) matching cost computation uses window-based approach and pixel-based intensity comparison between central pixel with neighboring pixels values. Kordelas et al. (2015), Song et al. (2015), Lee et al. (2016), Bae and Moon (2017) proposed a support window which is sensitive to noise. Lee et al. (2016) used an alternative CT called Star-Census Transform (SCT) that had similar operation with CT but in a comparison window which resulted in poor disparity accuracy. Inaccurate object disparities at the edges and low texture boundaries may be caused by unsuitable or incorrect window size selections.

If the window is too large and contains plenty of object boundaries, the algorithm will calculate the intensity value incorrectly since it will assume that the intensity values are similar. Nevertheless, selection of smaller window size will contribute to loss of important information at the depth discontinuities region. There is indeed a demand for a cost-effective and reliable technique with a robust function and minimal noise cost value to measure the preliminary performance of the stereo matching algorithm.

The low texture regions in the stereo pair images produce almost similar matching costs, thus making it challenging to compute the disparity value. It cannot be solved directly by increasing the size of aggregation windows or using the global optimisation methods, e.g., dynamic programming. Hamzah et al. (2017) proposed an algorithm employing per-pixel difference adjustment and iterative guided filter aimed to increase the accuracy of the aggregated cost value. This approach was able to smooth the depth of discontinued boundaries in the regions (Haibin Li et al., 2019). Zhu and Chang (2019) proposed hierarchical guided filter to improve the matching accuracy in the uneven texture distribution on the image pairs while an algorithm proposed by Xue et al. (2019) applied the multi-frame and edge matching technique to deal with occlusions boundaries as untextured regions. The proposed algorithm was not able to precisely determine the low texture regions. However, the algorithm was able to achieve a smooth and sharp edge disparity map. Conversely, the low texture and occluded regions had caused a great deal of inaccuracies. In order to address these challenges, this thesis proposes a new method for stereo matching algorithm which will have edge-preserving properties and be effective in the low texture, depth discontinuities and occluded regions.

1.3 The Objectives of This Thesis

The main objectives of this thesis are as follows:

i. To formulate a new computational matching cost function using a combination structure of window-based matching and pyramid cost functions that is able to increase accuracy at the boundaries and low texture regions.

ii. To derive an improved stereo matching algorithm which is robust against the low texture and occlusion regions with enhanced edge-preserving properties based on hybrid random, iterative aggregation and clustering to recover the value at the occlusion, depth discontinuities and to preserve the edges.

iii. To validate the proposed stereo matching algorithm performance using standard benchmarking datasets and real stereo images.

1.4 Scope of Work

The scopes of the research are limited to the following important notes:

1. This research is primarily focused on the formulation of a new matching algorithm to increase the accuracy and the algorithm in the spatial domain.
2. The proposed stereo matching algorithm is based on the standard four stages taxonomy which was developed by Scharstein and Szeliski (2002).
3. This research centres on the evaluation of stereo pair images with fixed-baselines and static sceneries which is executed only on Central Processing Unit (CPU) platforms.
4. A combination software of MATLAB and C++ programming languages are used to construct and validate the effectiveness of the proposed algorithm. All the experiments are executed using a personal computer with the features of Intel(R) Xeon(R) CPU E5-2650 @ 2.60GHz and 64G RAM.
5. The analysis of the research's algorithm performance is based on the standard benchmarking datasets of Middlebury (Sharctein et. al. 2014) and KITTI

(Meinze and Geiger, 2015). These datasets are widely used by researchers to evaluate the quantitative measurement in pixel error percentage. The quantified matching errors will be obtained by comparing the experimental results with the ground truth provided by the datasets, thus enabling the algorithm's accuracy to be evaluated objectively.

6. The additional evaluation in algorithm adaptability is tested using real stereo images from the Universiti Teknikal Malaysia's indoor scenes and implemented in 3D surface reconstruction.
7. The algorithm's performance is evaluated on the basis of qualitative and quantitative accuracy, without consideration for computational time, which can be enhanced with Random Access Memory (RAM) and Graphics Processing Unit (GPU) capacity upgrades (Chang et al., 2022).

1.5 Contribution of the Thesis

This thesis provides several noteworthy contributions to the area of stereo matching algorithm development. These contributions can be spesifically classified into two categories within the stereo matching algorithm. The first part of this thesis is focused on the computation of the cost value from the matching process between the left and right images. This is a well-recognised issue which necessitates a better approach to determine the best cost value to be aggregated. In addition, there is a demand for a cost-effective and robust approach to increase the accuracy of disparity maps, especially at the boundaries and low texture regions. In this thesis, a new unified approach is explored for a matching cost computation based on the pyramid cost function to improve the accuracy at the boundaries and low texture, especially with the introduction of the novel Modified Census Edge (MCE) and Planar Pyramid Fusion (PPF) approaches in multi-cost which differ from conventional

methods, which significantly increased the cost accuracy. Radiometric errors which occur during the matching process can be minimised by implementing this method and improve the sharpness of the texture edges.

The second part of this thesis focuses on a generic method that has been developed to solve a variety of challenges, depth discontinuities and occluded areas. A hybrid random aggregation of clustering, filtering, segmentation, and random walk is proposed in the cost aggregation stage while an iterative refinement and clustering technique is aimed to refine the final value at the last stage. The contribution for this stage includes the new implementation of novel cost aggregation in Hybrid Random Aggregation (HRA), which consists of the new Iterative Non-Local Guided Filter (iNLGF) and the Extended Random Walk Restart (eRWR). A Winner-Takes-All is applied to determine the optimum disparity considering the occluded and discontinuity regions. In the disparity refinement aspect, an improved hierarchical-edge refinement approach is introduced with the K-means clustering and iterative filter method based on the side window applied to preserve the disparity edges and the confidence filling approach to recover the occluded and unreliable disparities, thus improving the disparity accuracy.

1.6 Review of Thesis Organization

This thesis is divided into five (5) chapters based on the objectives and proposed approach previously stated. Hence, each of the chapter's description is presented as follows:

- Chapter 1. Introduction: This chapter presents a brief introduction explaining the background of the study, research problems, objectives, scopes, and research contributions.
- Chapter 2. Literature review: This chapter starts with a concise overview of the computational stereo vision fundamentals, explaining the related review papers, the

conceptual taxonomy, the various methodologies, and the stereo correspondence constraints in recovering accurate disparity map value with current stereo matching algorithms. Specifically, a 3D surface reconstruction method will also be discussed.

- Chapter 3. Methodology: This chapter focuses on the detail explanation of this research's new proposed stereo matching algorithm based on the standard taxonomy's framework. Consequently, the new algorithm will contribute to produce an accurate disparity pixel value which can be used for 3D surface reconstruction.
- Chapter 4. Case studies. This chapter presents the improved design of the stereo matching algorithm which is tested and verified through the assessments performed using the Middlebury Stereo Datasets, The Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) Vision Benchmark Suite and real stereo images. In addition, the experimental results and the comparison tables which have been thoroughly investigated are also discussed in detail.
- Chapter 5. Conclusion and future works: This chapter summarises the finding results as well as the achievements of this research. Last but not least, this chapter also highlights the potential industry applications and provide suggestions for possible future research work and improvements.

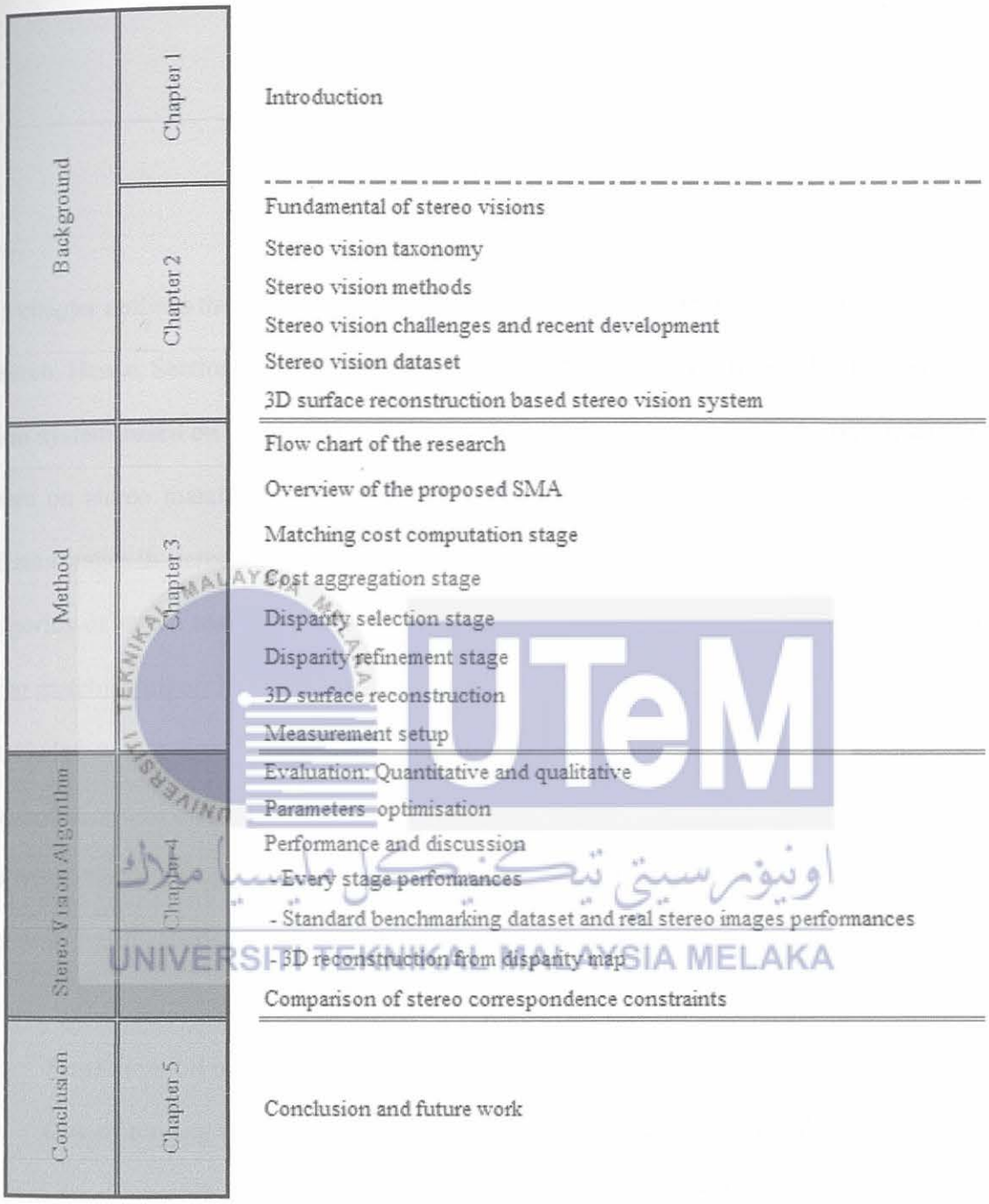


Figure 1.3: The 5 Chapters of the Thesis: Outline and Original Contributions

CHAPTER 2

LITERATURE REVIEW

This chapter outlines the existing methods available in the literature for stereo vision system research. Hence, Section 2.1 of the chapter begins by introducing the basics theory of stereo vision system based on the mathematical models including the summary of recent research papers on stereo matching algorithm. Then, Section 2.2 provides a brief review of the different stages in stereo matching algorithm taxonomy and Section 2.3 explains the various categories of stereo matching methods. Next, Section 2.4 reviews the prior works of the stereo matching algorithm based on major constraint in stereo correspondence. This section is included with a summary of stereo matching algorithm implemented based on method categorization and constraints. Then, Section 2,5 explains the stereo vision datasets used in this work. Finally, Section 2.6 explains the 3D surface reconstruction development while Section 2.7 provides the summary of this chapter.

2.1 Fundamental of Stereo Vision System

One of the key topics in computer vision is stereo vision system, which refers to a method for estimating depth from 3D information extracted from a pair of digital stereo images. The framework to obtain the disparity map is called stereo matching algorithm, widely studied by researchers concentrating to improve the disparity accuracy. A disparity map visually represents the adjacent pixels that have been horizontally shifted between the left and right images. One of the earliest design was by Marr and Poggio (1976) which employed human stereopsis as the main focus for obtaining the stereo image pair's disparity

cost using a cooperative algorithm, whereas G. Yang et al. (2019) recent work was based on the deep learning techniques.

The algorithm developed by Scharstein and Szeliski (2002) consisted of a four-stage framework; 1. Matching cost, 2. Cost aggregation, 3. Disparity selection and optimization, and 4. Disparity refinement. This framework was aimed to acquire a disparity map and as a result, it can be translated into depth assessment for depth-based processing and communications. The disparity estimation accuracy evaluation is quite crucial since small inaccuracies will impact the result of the 3D application. Therefore, the disparity measurement accuracy will need to be the quantitative assessment which can be compared with other algorithms (Cabezas, Padilla and Trujillo, 2011).

A notable research challenge in the field is the accuracy level of the disparity map produced by the stereo matching framework, which is used for reconstruction of three-dimensional coordinates of point and object detection. The architecture of a stereo vision system consists of a stereo camera which is a minimum of two cameras that are placed horizontally left and right of one another. The visual depth information produced from the two images captured from these cameras is the impression of a perceived depth when both human eyes and binocular vision view a scene. The basic structure of the stereo vision system is shown in Figure 2.1. The distance to the object P can be computed based on the triangulation principle or 2D perception to estimate depth Z as expressed in equation (2.1):

$$\frac{b}{Z} = \frac{(b+x_r)- x_i}{Z-f} \tag{2.1}$$

where x_r is the right plane coordinate, x_i is the left plane coordinate, Z represents the depth value, f for the focal length while b is the baseline. The depth value Z can be obtained using equation (2.2) and equation (2.3):

$$d = x_i - x_r \tag{2.2}$$

$$Z = \frac{bf}{d} \tag{2.3}$$

where d presents the pixels distance between coordinates at the left and right position.

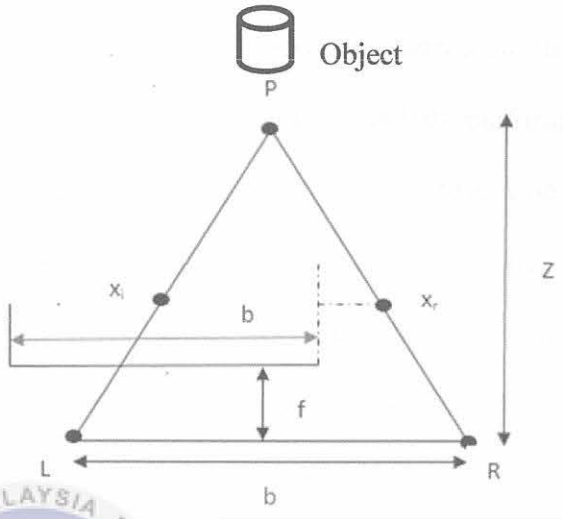


Figure 2.1: Depth Perception from 2D Stereo Geometry

New methods and techniques have been developed over the years to solve the fundamental problem of pixel differences leading to a significant improvement in disparity accuracy (R. Szeliski, 2009). The challenge is to find the best techniques for solving the correspondence problem to acquire the differences between pixels of the image in stereo pairs , which is essential for in-depth map generation (Kavitha and Balakrishnan, 2020). The triangulation principle's correspondence, or difference in location of the object's image between the reference and target images can be used for in-depth estimation for various applications, including 3D reconstruction, virtual reality, navigation for autonomous driving, and robotic guidance, medical diagnosis and object detection (Yousif et al., 2022). The most widely used disparity accuracy for quantitative assessment is the Bad Matched Pixels (BMP) methods (Cabezas, Padilla and Trujillo, 2011). Several academic research centers providing the quantitative and qualitative evaluation for disparity accuracy include Scene Flow (Wedel et al., 2011), Middlebury (Szeliski, 2020) and KITTI (Geiger et al., 2020).

In recent years, researchers have conducted numerous research to implement stereo matching for depth estimation application. This data can be observed from the latest past six years' review articles from 2017 to 2022, which are listed in Table 2.1, and each of which also include a summarised review. Each review investigated the stereo matching algorithm methods, the platform used, the execution time, and its performance towards disparity accuracy level on the cited algorithm. However, none of these review papers went into great detail regarding the stereo vision algorithm in response to the stereo correspondence's constraints. Based on these reviews, this work focuses on the formulation of SMA, which considers the stereo correspondence constraints to increase disparity accuracy using the framework developed by Scharstein and Szeliski (2002).

2.2 Stereo Matching Taxonomy

As shown in Figure 2.2, the majority of the most recent algorithms developed by several researchers were built upon a framework that consists of four stages. This framework produces a smooth, dense, or sparse disparity map depending on the input, which is a stereo image pair (Scharstein et al., 2014). Each block in the framework is the area of focus by many researchers, consisting of one or several algorithms to process the image and enhance the overall performance of the disparity accuracy level or the time execution (Kordelas et al., 2016; Qi and Liu, 2022).

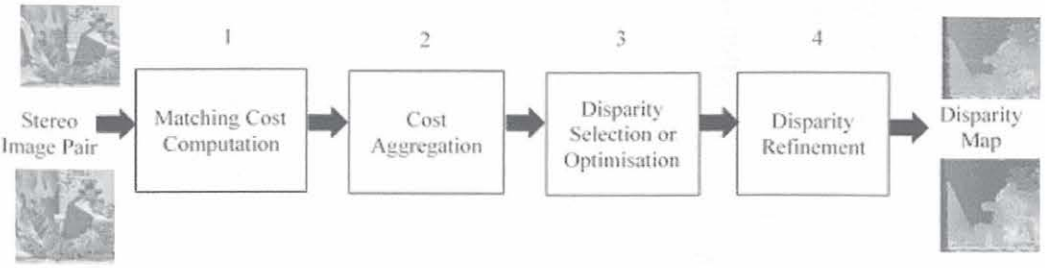


Figure 2.2: Traditional Stereo Matching Framework (Scharstein et al., 2014)

Table 2.1 : Recent Review Papers on Stereo Matching Algorithms

No.	Year	Author	Focus
1	2017	E. Bebeselea-Sterp, R. Brad and R. Brad (Bebeselea-Sterp et al., 2017)	This article compared stereo vision and matching algorithms for correspondence issues. The algorithm performance was compared to Middlebury Benchmarks. Local and global methods used include pixel-wise correspondence, shiftable window, dynamic programming, and graph cut. Several quality tests were performed.
2	2019	K. Y. Kok and P. Rajendran (Kok and Rajendran, 2019)	The research explored stereo vision depth estimation. Findings showed computational complexity, occlusion, radiometric distortion, depth discontinuity, and textureless areas influence disparity map accuracy and computational load. Stereo matching algorithms were developed by comparing existing stereo matching methods for design, performance, and enhancements.
3	2020	K. Hou, Z. Meng and B. Cheng (K. Zhou et al., 2020)	This review covered deep learning-based on stereo matching algorithms. Authors categorised deep learning into non-end-to-end, end-to-end, and unsupervised categories. Authors examined each algorithm's challenges, approaches, benefits, and downsides. Speed, precision, and time utilisation determine algorithm effectiveness.
4	2020	H. Chen, L. Wang and G. Liu (H. Chen et al., 2020)	The stereo matching algorithm, widely described as a binocular algorithm was discussed in this review for binocular stereo vision. In terms of their features, the global matching algorithm, and the local matching algorithm were the two methods that were addressed.

5	2021	Lin Cao and Weiwei Yu (Lin Cao and Weiwei Yu, 2021)	This work examined binocular stereo vision before describing stereo matching. Then it discussed stereo matching's history and research significance. Finally, recent passive stereo matching algorithm research and challenges were summarised.
6	2021	Guobiao Yao, Alper Yilmaz , Fei Meng, and Li Zhang (Yao et al., 2021)	The author investigated learning-based feature recognition, description, and end-to-end picture matching. The author provided a stage-by-stage inspection and deconstruction of the latest representative study. By using the author's extensive experiment results on actual wide-baseline stereo images, researchers analysed and compared numerous deep-learning methods.
7	2022	Mingyu Jang, Hyunse Yoon, Seongmin Lee, Jiwoo Kang, and Sanghoon Lee (Jang et al., 2022)	This work examined the active stereo method and pattern texture. Experiments with pattern intensity, contrast, number of pattern dots, and global gain were performed to evaluate the active stereo matching method. The discovery contributed to building active stereo systems.
8	2022	Ali N. Yousif*, Hassan M. Ibrahim, Safaa J. Alwan, and Mohammed Sh. Majid (Yousif et al., 2022)	This study examined stereo matching algorithms, systems, and image processing. These systems and methods could be easily integrated with multiple applications to overcome stereo vision problems like low accuracy, mismatched algorithms, and correspondence concerns. Various authors used different types of frameworks aimed to solve problems in this review.

Computation of matching cost. The process of determining whether the values of two pixels correspond to the same point in a scene is known as matching cost computation, which obtains the cost value. Wang et al. (2015) proposed an essential matching cost absolute difference (AD) cost initialisation for the high-quality real-time system which was based on pixel matching. An enhanced AD of Sum of Absolute Differences (SAD) was used by Mannan Mondal et al. (2017) to calculate each pixel matching and is presented in equation (2.4):

$$\mathrm{SAD}(p,d) = \sum |I_r(p) - I_t(p-d)|, \tag{2.4}$$

Despite being sensitive to radiometric differences, TAD has proven to be an excellent method in the presence of multilayer color image and in the region of flexible aggregation (Ma et al., 2016). Another familiar pixel matching cost computation is NCC, applied by Yousif et al. (2022) and as expressed in equation (2.5):

$$\mathrm{NCC}(p,d) = \frac{\sum_{p\in w} I_r(p).I_t(p-d)}{\sqrt{\sum_{p\in w} I_r^2(p).\sum_{p\in w} I_t^2(p-d)}}, \tag{2.5}$$

Zhu and Yan (2017) adopted a block-based matching using an improvement of Census Transform consisting of local texture metric used to calculate the initial cost as given in equation (2.6) and (2.7):

$$\mathrm{CT}(p,d) = \sum_{p\in w} \mathrm{Hamming}\big(\mathrm{Census}_l(p) - \mathrm{Census}_r(p-d)\big), \tag{2.6}$$

$$\mathrm{Census}(x,y) = \mathrm{Bitstring}_{(i,j)\in w}\big(I(i,j) \geq I(p)\big), \tag{2.7}$$

L. Li et al. (2018) introduced a new matching framework based on feature matching using Patch Match-based superpixel slice and image 3D labels. Zhao et al. (2019) proposed an algorithm with combination of feature matching and pyramid network to allow both direct and indirect based pose optimisation for achieving coarse-to-fine calculation and getting more accurate results.. A more advanced algorithms used CNN-based matching cost function, which exploited block and feature matching employed by Mozerov and Van De

Weijer (2019). Wei et al. (2021) incorporated improved block Census transform and adaptive weighted pixel bidirectional gradient information to estimate initial matching cost and to increase cost calculation accuracy. The brief characteristics of each respective method are presented in Table 2.2. This table shows the simple computation for matching cost for pixel and block matching; however, this type of matching cost produced low disparity accuracy. The feature matching offered acceptable disparity accuracy, but focusing only on certain features, such as edges and boundaries, resulted in hidden details and was interrupted by noise at similar characteristics. The implementation of multi-cost matching will increase the robustness of the initial matching cost calculation and compensate for weaknesses in matching singularity; however, the computational complexity and execution time will increase.

Cost Aggregation. Traditionally, a square-shaped fixed-size window was utilised at this stage due to low computational costs and simple to implement (Matsuo et al., 2015). Ma et al. (2017) performed cost aggregation using a cross-scale technique by incorporating weighted least square with intrascale of smoothness constraint with slanted surfaces and discontinuities (Yibo Li et al., 2018). Various approaches have been applied, such as shiftable window by Zeglazi et al. (2018) and support window by Qi and Liu (2022). The shiftable window function are expressed in equation (2.8) and (2.9):

$$T^r(x, y, d) = \min_{-r \le m < r} \left(A^r_{square}(x + m, v, d) \right), \tag{2.8}$$

$$A^r_{shiftable}(x, y, d) = \min_{-r \le n < r} \left(T^r(x, y + n, d) \right), \tag{2.9}$$

While a support window function based on the optimal window from a set of windows of various shapes or sizes at each pixel point is defined in equation (2.10):

$$A^r_{adaptive-avg}(x, y, d) = \frac{1}{r} \sum_{k=1}^{r} A^r_{shiftable}(x, y, d), \tag{2.10}$$

Table 2.2: Characteristics of Matching Cost Computation Methods

Method	Description	Advantages	Disadvantages
Pixel matching (Mannan Mondal et al., 2017)	The method is performed by comparing one to one pixel between left image pixel and right image pixel. The difference is aggregated in luminance. Example: Absolute differences and square differences.	• Simpler computation. • Fast execution runtime. • Only the corresponding single pixel is involved between the left and right images, respectively.	• Repetitive patterns and low texture regions are noise sensitive. • Contribute radiometric errors. • Sensitive to change in illumination.
Block matching (Zhu and Yan, 2017)	The process involves the aggregation of pixels over a small region. Those collectives of pixels are often referred to as "blocks" or "windows." The matching is determined based on the magnitude differences between the windows of the left and right images. Example: SAD, SSD, NCC, Census Transform (CT), and Gradient Magnitude (GM).	• The window size affects computation time. • The correct window size selection is crucial to the accuracy of the disparity map.	• The edge of an object will become wider or blurrier because of poor window selection. • Depth discontinuity could be caused by improper window selection.
Feature matching (L. Li et al.,	This process is performed using feature-based techniques. This approach attempts to establish correspondence only for similar feature points that can be unambiguously matched such as visual	• Focus entirely on the image's features (such as object boundaries, edges, and corners).	• Obtain a disparity map that includes a "sparse" feature. • Not entirely reliable

2018), (Zhao et al., 2019)	features, statistical characteristics, and transformation features. Example: Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), and Histogram of Gradients (HOG).	• Fast execution runtime with optimal accuracy.	• Certain details in one image might be hidden in the other. • Noise could interrupt the similarity between two characteristics. • Poor performance in low texture regions and sensitive to occlusion
Multiple cost matching (Mozerov and Van De Weijer, 2019), (Wei et al., 2021)	This process is performed by combining multiple cost functions during matching cost computation. The reason is to compensate for their respective limitations. This process also can include other functions such as pyramid network, iterative, and multi-scale method to smooth the cost volume and remove noises as a coarse representation. Example: CT + TAD, TAD+ GM, Gradient-based Census Transform, and CNN.	• Compensate weakness in matching singularity. • Reduce errors in most regions, especially those containing repeated color and shapes. • Increase the robustness of initial matching cost calculation.	• Complex computational • Increase execution runtime

A better solution is to implement Adaptive Support Weight (ASW) by adjusting each point's cost weight, such as similarity and proximity between adjacent pixels and target points in a fixed size window by Liu et al. (2019), as presented in equation (2.11) and (2.12):

$$T^{r}(x, y, d) = \frac{\sum_{m=-r}^{r}(w(x,y,m,0).C(x+m,y,d))}{\sum_{m=-r}^{r}=w(x,y,m,0)}, \quad (2.11)$$

$$A^{r}_{weight}(x, y, d) = \frac{\sum_{m=-r}^{r}(w(x,y,m,0,n).T^{r}(x,y+n,d))}{\sum_{m=-r}^{r}=w(x,y,m,0,n)}, \quad (2.12)$$

Y. Zhang et al. (2018) combined ASW with CNN to end network as cost aggregation to get more support for thin structures. Huo and Luo (2019) adopted a combination of a minimum spanning tree based on support region and joining object flow in cost aggregation rather than using fixed-sized windows. Liu et al. (2020) also produced an improvement in disparity accuracy for a large area of textureless regions by implementing an Adaptive Guided Filter (AGF) over the support region. An enhanced Minimum Spanning Tree (MST) called 3DMST-CM was proposed by Xiao et al. (2020) handling cases based on image pixel ambiguity to achieve a high-level accuracy in the disparity map. Additionally, the latest usage of the encoder-decoder network framework in deep learning by Y. Zhang et al. (2020) included multiple outputs and calculation loss. Numerous methods exist and the most popular is the averaging of matching cost in a local window based on the filter type aggregation. A quiet popular approach is using the Guided Filter (GF) developed by He et al. (2013) and employed with the adaptive support weight (ASW) by Liu et al. (2020), and Yuan et al. (2021). Although a well-developed method, there are still issues with gradient reversal artifacts, generating halo artifacts at sharp edges and susceptible to the lack of texture. Hence, W.J. Yang et al. (2019) and Hamzah et al. (2020) employed a slightly different technique using the Bilateral Filter (BF) in the cost aggregation.

Table 2.3: Characteristics of Cost Aggregation Methods

Method	Description	Advantages	Disadvantages
Fixed window (Matsuo et al., 2015)	A square-shaped fixed-size window.	• Low computational complexity. • Fast execution runtime.	• Low accuracy. • Poor performance on low texture regions and boundaries. • Sensitive to noise especially illumination variation. • Does not yield reliable disparity estimates for all the pixels.
Multiple window (Qi and Liu, 2022)	A coarse-to-fine multi-window algorithm can be implemented by forming multiple smaller windows	• Eliminating the blurring of boundaries. • Perform well in low texture regions. • Focuses on the dissimilarity calculation	• High computational cost. • Need to address the number of windows and the criteria for selecting reliable disparities. • Poor performance in depth discontinuity reshaped window models may be inappropriate for depth discontinuities regions.

Adaptive window (Zeglazi et al., 2018)	Adaptive window construction method which can alter the window's shape and size adaptively based on threshold of features or disparity variation.	• Perform well in-depth discontinuity regions. • Fast execution runtime. • The degree of disparity varies smoothly in each region	• Final output depended on the choice of the initial disparity estimation. • Perform poor in occluded regions. • Difficult when dealing with highly textured images.
Adaptive support weight (Liu et al., 2019), (Y. Zhang et al., 2018)	Changing the cost weight of each point according to factors like the similarity and closeness of target points and adjacent pixels in a fixed-size window.	• Perform well in low texture regions, object boundaries, or depth discontinuities.	• Computational complexity directly depends on the size of the support windows.
Filter aggregation (Liu et al., 2020)	Cost aggregation based on filter function. It computes a weighted local of the cost volume.	• Moderate computation • Fast execution runtime • Focus on surface smoothness easily overlaps object boundaries and depth discontinuities. • Smooth the matching cost while preserving the disparity boundaries efficiently.	• Perform poorly at the edge if poor selection of filter. • Execution runtime depends on kernel size. • Cannot perform well for low texture regions especially large textureless regions.
Cross-scale (Ma et al., 2017), (Yibo Li et al., 2018)	A cross-scale technique by incorporating several different functions in the left and right aggregation. Stereo correspondence search across multi-scales and aggregated costs across multi-	• Robust cost volume and accurate disparity map.	• High computational cost. • High execution runtime.

	scales. builds the cost volume in the scale space and makes sure the consistency across multi-scales.	• Focus on smoothness constraint on neighborhood cost.	
Segment tree (Huo and Luo, 2019), (Xiao et al., 2020)	Cost aggregation based on tree structure of reference color or intensity image. Graph connectivity is determined with local edge weights.	• Preserve depth edges with high efficiency. • Better handle low texture regions • Smooth out high-contrast details while preserving major edges. • Low complexity • Fast execution runtime	• Suffer from edge blurring effect. • Assumes that disparity smoothes at every point. • Poor performance in highly textured regions.
Deep learning (Y. Zhang et al., 2020)	A hierarchical learning network of multiple layers involves aggregating the cost volume.	• High accuracy • Powerful representation ability. • Perform well on low texture and boundaries.	• High computational burden. • Limited receptive field. • Lack of context information. • Still more or less uses postprocessing functions. • Speed and memory usage problem.

Table 2.3 summarises the characteristics of each cost aggregation method. Based on this table, deep learning, cross-scale, and ASW methods produce the highest disparity map accuracy, but the trade-off is the high computational complexity, which requires an extra processing upgrade. The simplest and fastest execution for cost aggregation methods is the fixed window; however, this method produces poor performance in disparity accuracy and is sensitive to noise. The methods that moderate disparity map accuracy and execution runtime are the filter and segment-tree aggregations.

Disparity Optimisation. This is a fundamental disparity optimisation utilising the Winner-Takes-All (WTA) strategy by Zhu and Dai (2017) and as given in equation (2.13):

$$d_p = \arg\min_{d \in d_r} C(p, d), \tag{2.13}$$

The final disparity value of that particular pixel point is selected from the disparity range with the minimum cost value after the cost aggregation. Chang, Lu and Yang (2017) adopted a cross-based window voting in WTA called Range-WTA to estimate the disparity and to fill the white holes. In contrast, global optimisation commonly produces higher accuracy of disparity maps compared to local methods at a computational cost disadvantage. Yao et al. (2019) applied adaptive smoothness components by incorporating the energy function using a model that considered all four images traversal directions and the correlation between neighbouring pixels. Meanwhile, Hallek et al. (2022) introduced a global approach to WTA optimisation with dynamic programming.

Based on data from the matching cost and the smoothness energy, which is stated in equation (2.14), this global technique resulted in fewer errors caused by textureless, occlusion, and discontinuity regions.

$$E(d) = E_{data}(d) + E_{smooth}(d), \tag{2.14}$$

Ni et al. (2018) implemented a second-order smoothness constraint based on angle direction with matching cost and smoothness constraint for semi-global optimisation to

increase the matching accuracy in the weak-textured regions. Yao and Feng, (2021) utilised semiglobal optimisation disparity maps for 8 directions to complement one-dimensional scanline optimisation for semi-global matching.

Disparity Refinement. Post-processing in the stereo matching framework is called disparity refinement to remove any outliers, uncertainties, and noise from the map for the disparity to achieve a greater accuracy level. Chang et al. (2017) proposed a technique using the cross-voting of image-based to improve the occlusion. The remaining noises and outliers only can be detected by performing an occlusion constraint by Zhang et al. (2018), a left-right consistency checking by Huo and Luo (2019), a bi-modality and a match goodness jumps. Additionally, a median filtering also can be applied for an unflawed depth measurement cost using the triple image approach to identify textureless regions and false matches. One of the most widely used image filters, the Bilateral Filter (BF), which preserves and provides a smooth the object boundaries were applied by Jia et al. (2020). The BF equation is given in equation (2.15):

$$B(p,q) = \exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right)\exp\left(-\frac{|d(p)-d(q)|^2}{\sigma_c^2}\right), \quad (2.15)$$

Various approaches have been proposed in the disparity refinement, such as the probabilistic approach by Jia et al. (2016), the Order-Based and the segment-tree structure. Li et al. (2019) introduced a coarse-to-fine (CTF) image segmentation, including labelling and histogram to correct the occlusion and error pixels.

2.3 Stereo Matching Methods

Despite the challenges, there have been various development on algorithms for SMA contributed by researchers in recent years. The algorithms can basically be classified into three main methods: local, global and semiglobal. Nonetheless, this thesis applies the stereo

matching algorithm four-stages framework from Scharstein and Szeliski (2002) with additional use of the three main methods for machine learning.

Local Method. This method, which typically yields fast and high-quality results, allocates disparities of the map based on the information produced by the neighboring pixels. Both Lee et al. (2013) and Jafari Malekabadi et al. (2019) developed a variant of this popular algorithm which achieved high speed and reasonable quality. For the purpose of addressing the issue of window size dependence and computational complexity, the recursive edge-aware filters (REAF) introduced by Cigla (2015) provided an in-depth examination and categorised the recursion according to their characteristics. Historically, Aboali et al. (2017) was among the first to develop the box filtering for stereo block matching that was aimed to replace each image's pixel respectively to the average inbox size. In addition, Chang and Maruyama (2018) introduced a multi-block matching (MBM) technique based on normalised cross-correlation (NCC), which was generated using a 3 x 3 window centred on the pixel and adding the NCCs in blocks of varying sizes and shapes centred on the target pixel. Meanwhile, Chang and Ho (2019) proposed a practical algorithm for producing disparity using combination pixel-wise and block-wise matching.

This method includes gradient magnitude and optical flow pioneered by Zhang et al. (2018) to investigate the application of minimal spanning tree based on support region and joining object flow in cost aggregation instead of fixed-size windows approach that leveraged on temporal information from consecutive frames similar to motion flow. Furthermore, Li et al. (2018) introduced another method called the Patch Match algorithm that incorporated strategy using efficient random search and additional coarse-to-fine scheme for dense stereo correspondence. Similar work were pursued by Lim and Lee (2019) using Patch Match technique for randomised-based search with an efficient image filtering for the correspondence's edge awareness. To achieve an effective weighted aggregation

calculation, Kong et al. (2021) offered an orthogonal weight based approach on the structural feature of the ACR and filters using ACR-GIF with orthogonal weights (ACR-GIF-OW).

Conceptually, a similar work was carried out by Zhan et al. (2016) by combining raw stereo images and double gradient model to acquire two directional gradients. Other techniques applied the similar concepts which comprised of stereo images segmentation to compute the adaptive support window corresponding to the range of the respective segmentation region (Shi et al., 2016). More recent work in this area was from Zhou et al. (2019) which extended the local method by applying colour segmentation and calculate the colour's matching cost by fusing weighted absolute difference and gradient. Furthermore, Liu et al. (2021) proposed an enhanced AD-Census algorithm based on gradient fusion and two-phase adaptive optimisation. In this study, an adaptive nonlinear constraint of the intensity difference between pixels was employed during the cross-arm creation phase to obtain the ideal arm length.

The preceding studies by Vieira et al. (2018) and Zeglazi et al. (2018) addressed these local methods through the support window approach. The algorithm was developed using the ASW approach by integrating both disparity continuity constraint and segmentation information. Later, Wu et al. (2019) extended the method using a generic fusing ASW framework to determine the dual support windows for each pixel to specify the local window and the entire window. Another method described in Peng et al. (2018) utilised a combination of per-column cost matrix with a feature-vector-based weighting strategy to accomplish the efficiency in both computational and matching accuracy. A novel similarity measure known as characteristic correspondence measure was explored by Han at al. (2019), which used ridge regression of guided filter to extend multiple linear regression. Moreover, Navarro and Buades (2019) applied the adaptive support weights approach to implement a local method, with the weight distribution preferring pixels that shared the same

displacement as the reference pixel. More current methods apply the weight function which is solely dependent on image attributes. For instance, Zhang et al. (2020) used weighting process based on the gravity theory which identified the weighting coefficient for cost based on the gravitational model.

Another type of local method is the feature-based matching algorithm that can be categorized into two forms: segmentation and hierarchical. ELAS (Efficient Large-scale Stereo), a generative probabilistic model introduced by Geiger, Roser and Urtasun (2011) for stereo matching, enables dense matching with small aggregation windows by minimising the correspondence ambiguities. Jung et al. (2015) introduced boundary-preserving stereo matching using adaptive disparity classification and particular region detection for a hierarchical form of matching. In contrast, Jellal et al. (2017) proposed an upgraded of a line segment extending to the efficient large scale stereo matching algorithm. In the work of Hu et al. (2018), the segmentation form were addressed using the parallax plane and Delaunay triangulation algorithm to divide the whole image into a series of linked triangular facets. Furthermore, Fu et al. (2019) proposed a method that was more stable and robust using the cost function and feature extraction based on Spearman Rank (SR) correlation suitable for image matching under different photographing condition.

Global Method. Despite the local method's success, some drawbacks to this technique also exist due to the sensitivity to occlusions and uniform texture. Men et al. (2015) also concurred that this method could increase the risk of local errors being produced along a scan line which could also affect the suitable matches. However, using global methods as opposed to local methods often yields more accurate results. Unfortunately, the main difficulty with this approach is the increased in computational complexity. Zhou et al. (2016) described one of global approach to solve the mismatching of disparity in depth-discontinuous regions. Nevertheless, the utmost popular global methodology is the Dynamic

Programming (DP), proposed by Zhou et al. (2016), Bhalerao at al. (2017), Jeong and Jay Kuo (2019) and Zhu et al. (2019) to categorise the optimisation constraints into reduced and simpler sub-problems. Xiao et al. (2020) and Xu et al. (2020) performed a comparison of different methods and asserted that a cost-aggregation method can be integrated with minimum spanning tree and 3D label search instead of using fixed-size patches as the aggregation.

Applying the Graph Cut Programming (GCP) method was observed to produce optimal segmentation solution globally. Kim et al. (2015) proposed a progressive unit of Ground Control Surfaces (GCS) and the Probabilistic Laplacian Surface Propagation (PLSP) framework to improve the constraints of advanced conventional and superpixel-based approach. This method enhances the efficiency and accuracy of disparity values for stereo matching and joint object segmentation (Xu et al., 2015). Wang and Liu (2015) and Taniai et al. (2018) proposed an alternative method, Belief Propagation (BP) from, which provides a belief aggregation in the SMA. Wang and Liu (2015) also introduced a technique that uses Initial-Value Belief Propagation (IVBP) paired with a Self-Adapting Dissimilarity Data Term (SDDT) to reduce iterations and to increase energy and smoothness terms that ignore the data term. The BP algorithm shows several limitations including textureless regions, ambiguous edges, and slow speed of convergence. For top-down cues, Hadfield et al. (2017) presented a unified framework, which reduces baseline issues with bottom-up reconstruction. The framework offers clues for identifying surface normal, collinear edge or coplanar structures like walls, and semantic edges (such as concave, convex, and occlusion borders). Pan, Liu and Huang (2019) developed a stereo matching algorithm that is robust due to the extraction refinement feature using a smoothed BP for belief volume, a stack auto-encoder, and a guided filter to BP performance on edge regions.

Another type of global method is Nonlinear Diffusion (ND) that focuses on difficulties contributed from foreshortening and occlusion. In order to solve the energy-minimisation problem for the fully connected model, Mozerov and Van De Weijer (2015) used a two-step approach based on two Markov Random Field (MRF) models: 1) a completely connected model defined on all of the pixels in an image, and 2) a locally connected model. Li et al. (2016) presented a unified MRF model coordinating method and a 3D surface to generate multiple disparity proposals and mediate multiple disparities. The relative disparity proposals reflect the 3D structures, whereas the absolute disparity proposals are accurate point-wise predictions. Moreover, Geng and Luo (2017) proposed a different type of stereo matching algorithm based on multiple neighbors' nonlinear diffusion of costs aggregation to improve the global costs function. In addition to the above explained global approaches, many alternative solutions have been developed such as Simulated Annealing (SA) and Wavelet Transform (WT). Yao et al. (2018) advanced an adaptive meshing technique using several modifications such as the addition of sparse stereo correspondence and view synthesis, the application of the module of block-saliency design, the fine-grained optimisation and the novel parallelisation strategies.

Semi-Global Methods. This is a slightly different approach which is quite prevalent that incorporated both local and global optimisation. Since Hirschmüller et al. (2008) developed the first SGM model , many other algorithms or frameworks have been proposed such as from Bethmann and Luhmann (2015), Chang et al. (2017), Yin et al. (2017) and Yue et al. (2018). Hirschmüller et al. was able to quickly approximate radiometric differences, using a pixelwise, mutual information (ML) based matching cost, supported by a SGM smoothness requirement that is often stated as a global cost function. Similarly, Bethmann and Luhmann (2015) suggested a modified SGM approach in which a path-wise minimisation is allocated to object space and a semi-global optimisation guides to the index

maps instead of the disparity maps that successively indicate 3D coordinates of best matches. Another algorithm developed by Kordelas et al. (2016) exploited the weighted semi-global optimisation that aims to enhance the accuracy of the disparity estimation. Hamzah and Ibrahim (2018) proposed using 16 different directions of a 2D path to minimise the errors and increases the disparity map accuracy level. In addition, Lee et al. (2018) introduced a combination of a Gaussian Mixture Model (GMM) function with SGM.

Scharstein et al. (2018) evaluated the SGM approach for coarser resolution plane orientation priors produced via stereo matching to improve the performance for challenging weakly textured scenes. However, Loghman et al. (2018) proposed a more efficient implementation on semi-global matching using adaptive stripe-based optimisation and fast depth estimation. Recent work by Schönberger et al. (2018) extended the method by applying a learning-based approach called SGM Forest algorithm to handle the combination costs of optimising various 1D scanlines, which could undermine disparity accuracy under challenging circumstances using per-pixel classification. Jafari Malekabadi et al. (2019) compared the various SGM algorithms available for geometric tree characteristics used to predict water consumption, biomass, yield and fertilizer application in citrus crops. An improvement over SGM method was developed by Zhu et al. (2019) by aggregating one-dimensional dynamic programming and extending to the multi-scan line optimisation. Most recently, Rathnayaka and Park (2020) employed dense stereo matching algorithm method for a pool of multi-baseline stereo images called Iterative Guided-Gaussian Multi-Baseline (IGG-MSB) stereo matching.

Machine Learning. Currently, stereo matching algorithms have become an exciting topic in Machine Learning (ML) field with the deployment of several Artificial Intelligence (AI) and Deep Learning (DL) as reviewed by Zhou et al. (2020) and Hamid et al. (2020). Zbontar and Lecun (2016) introduced a supervised network training method by constructing

a binary classification data set with samples of similar and dissimilar pairs of patches used to obtain depth information from a rectified image pairing under supervised learning category. The combination of multiple matching functions and confidence estimations in both matching directions resulted in the method by Batsos et al. (2018) describing the calculation for the stereo matching volume. In order to address colorisation and depth super resolution issues, Bapat and Frahm (2019) proposed an universal optimisation framework called Domain Transform Solver that directly operates in the pixel space while maintaining distances in the combined colour and pixel space. Another method was proposed by Nguyen and Ahn (2020), which is an evolutionary algorithm for parameter selection framework for stereo correspondence and trained in a supervised manner.

Reinforcement Learning (RL) also belongs in the ML category. Yang et al. (2017) introduced a technique using matching cost of Euclidean learning to include a CNN with a triple-based loss function. Ye, Li, et al. (2017) employed a method using two patch-based network architectures in matching cost computation that manipulate multi-layer and multi-size pooling unit with no strides to learn cross-scale feature representations. In order to estimate an accurate starting disparity, Lu et al. (2018a) presented a multi-dimension aggregation sub-network with 2D and 3D convolution operations. This network is able to provide rich context and semantic information. Zhang and Wah (2018) proposed using two vital philosophies based on the distinctiveness of features and consistency. Moreover, Mahato et al. (2019) developed an algorithm for dense stereo matching approach using Genetic Algorithms (GAs) in addition to multi-objective fitness function-based. Recently, the most popular method in ML for stereo matching applications is the Deep Learning method, which is divided into three groups: Non-End-to-End, End-to-End and Unsupervised Learning. Du et al. (2019) and Xu and Zhang (2020) studied the Non-End-to-End approach of the CNN, which aims to swap one or more steps in the traditional stereo framework.

The End-to-End disparity map algorithm networks aims to incorporate all stages in the stereo matching framework under joint optimisation. Wang et al. (2016) examined a deep conditional random field based stereo matching algorithm that extracts information and a connection between CNN and CRF. Huang et al. (2017) pioneered the research on conditional adversarial networks to a stereo matching method that performs a conditional adversarial training process on two networks. A discriminator recognises if the disparity map is from the generator or ground truth, and an RGB generator gathers RGB images mapping to a dense disparity map. Later, Knöbelreiter et al. (2017) proposed a hybrid CNN+CRF model for stereo that uses CNNs to compute good unary and pairwise costs and the CRF to effectively integrate long-range interactions with efficient training. Likewise, Zhang et al. (2018) obtained an accurate dense disparity maps from stereo images directly using modified joint optimisation.

Several unsupervised learning methods have been employed with the view synthesis and spatial transformation over the past few years. Kim et al. (2019) introduced a deep model which measures stereo correspondence confidence consisting of a pooling module and a residual networking cost and a deep, unified network. The recent study developed by Hong and Ahn (2020) in Single-View Videos (SMV) applied an unsupervised method to build a deep learning-based of the algorithm for more accurate depth estimation. Zeng and Tian (2022) developed stereo matching strategies by improving network structure, inspired by network compression, decomposition, and sparsification. The networks are sparsified and fragmented into smaller networks, which are cascaded to attain a bigger network's performance. Although the ML method provides a mixed result between moderate and high accuracy, it involves high computational complexity and run time due to too many functions and subsequent processing in the algorithm, which require high processing capability unsuitable to be used in real-time applications.

Based on the discussed stereo matching methods above, it can be summarised that although the local methods provide fast execution, simple computation, and are suitable for real-time application compared with global, semiglobal, and ML, this method only offers low to moderate disparity accuracy, which will be the focus of this work. The local method is also sensitive to noise, especially at illumination variations, low textures, and boundaries. This is challenging to determine accurate correspondence between pixels that lie around the boundaries due to different illumination conditions, based on work by Kok and Rajendran (2019) and Zhou et al. (2019). The intensity level between these corresponding points is assumed to be identical to each other, especially in the untextured regions. It is difficult to establish accurate correspondences between pixels, especially at the boundaries, due to various lighting conditions and other factors, so it requires the derivation of more complex cost functions at the matching cost for this problem.

Then, the local method is usually based on support window or pixel-based intensity, such as work by Lee et al. (2013), Jafari Malekabadi et al. (2019), and Chang and Ho (2019) at the matching cost computation, which contributed to the comparison between the central pixel and neighbouring pixels that may cause problems at incorrect disparities. These types of approaches are sensitive to noise and contribute to inappropriate or wrong window size selections, which produce inaccurate object disparities at the edges and low texture boundaries. Next, the local method also has poor performance in low-texture regions at the cost aggregation stage due to the cost volume producing almost identical matching costs. This problem of similar matching costs at low texture cannot be solved efficiently with increasing window aggregation size or the implementation of global optimisation, which require an increase in computational complexity. The usage of edge-preserving filters such as iGF by Hamzah et al. (2017), Haibin Li et al. (2019), and HGF by Zhu and Chang (2019) produces good accuracy in the discontinuity regions but high error in the plain colour regions.

While the work of Xue et al. (2019) applied the multi-frame and edge matching technique, they were not able to precisely determine the low texture regions but achieved a smooth and sharp edge disparity map. Based on this problem, this work also focuses on the derivation of new SMA that have robust window matching cost computation, window filter cost aggregation, and refinement to solve the accuracy issue in low texture, edges, and discontinuity regions.

2.4 Stereo correspondence constraints and recent development

The primary constraint in the stereo vision systems is the calculation of stereo matching algorithm known as the stereo correspondence. The constraints are focused on the spatial displacement determination between two corresponding pixels in the stereo image pair. The local method allocates the map's disparities based on the information produced by the neighboring pixels to deliver a generally fast and good quality result. However, despite the local method's success, limitations to this method include sensitivity to occlusions and uniform texture. Alternatively, the global approach utilises non-local limitations which are more accurate than the local method, but this strategy requires intensive computation. Therefore, this section summarises the significant stereo correspondence constraints and the different algorithms developed to address the constraints.

(a) Left image (b) Right image

Figure 2.3: Radiometric Differences

Radiometric Differences. As shown in Figure 2.3, one of the common challenges in stereo correspondences is the inconsistency in two points of stereo images that need to match due to the different intensity and colour depending on the perspective of the images. Possible reasons for this include camera sensor characteristics or image noise, vignetting and slightly different settings (Lee et al., 2013). Another challenge is from the non-Lambertian surfaces of the reflected light from the camera's different viewing angle. Zhan et al. (2016) carried out a study to improve the differences by combining the raw stereo images and using double gradient model based on the gradient operator to acquire two-directional gradients. Han et al. (2019) employed ridge regression of guided filter and extended stereo matching framework of multiple linear regression to improve the image halo and noise in the edge area for local guided image filtering.

To improve matching accuracy and to lessen the impact of fattening, Chang and Maruyama (2018) introduced the Multi-Block Matching (MBM) algorithm employing NCC. It is quite challenging to determine an accurate stereo correspondence with stereo image pair's viewpoint that consists of exposure to poor conditions and illumination variations (Chang and Ho, 2019). The weight distribution in the adaptive support weights method used by Navarro and Buades (2019) favours pixels with the same displacement as the reference pixel. This method significantly lowers errors caused by match ambiguities, reduces the fattening effect, and is resistant to additive illumination. Several researchers found that a matching cost computation using an adaptive pixel-wise with a block-wise technique that incorporated illumination conditions achieved good results in radiometric differences based on the Middlebury dataset. Census Transform (CT) is another technique which can be used due to the brightness or intensity comparison between central pixel with neighbouring pixels' values within a support window and is sensitive to noise (Ji et al., 2020). Kong et al. (2021)

proposed a much more robust approach when illumination or exposure changes for a pair of images using multi-cost matching cost and ACR-GIF-OW filtering.

The first model in SGM was developed by Hirschmüller et al. (2008) and since then, many other SGM algorithms or frameworks have been proposed, including Chang et al. (2017). For example, the Graph Cut Programming (GCP) method implementation minimises image noise, illumination source changes, non-Lambertian surfaces, vignetting, and device features. Kim et al. (2015) designed a progressive unit of Ground Control Surfaces (GCSs) and The Probabilistic Laplacian Surface Propagation (PLSP) framework which concurrently improves the constraints of advanced conventional and superpixel based techniques. Another research by Li et al. (2019) presented an image fragments type of stereo matching that deals with a segment to produce high matching accuracy and better images with illumination. Recently, an improvement over SGM method was developed by Zhu et al. (2019) by aggregating one-dimensional dynamic programming and extending to the multi-scan line optimisation.

Zhang and Wah (2018) applied the SMA with the Reinforcement Learning (RL) such as stereo matching algorithm using two vital philosophies based on the distinctiveness of features and consistency to solve radiometric changes, noises, over-exposure and textureless regions. Joung et al. (2020) implemented unsupervised CNN with combined domain learning and stereo epipolar constraints. The result was promising in radiometric differences and low texture regions but the matching in the repetitive areas was very poor. The End-to-End algorithm proposed by Song et al. (2020) aimed to incorporate all stages in the stereo matching framework under joint optimisation, obtaining directly the dense disparity maps from the stereo images invariant to illumination changes. Comparably, Liang et al. (2021) employed a ML method using multi-scale feature constancy and multi-level

cost volume which was successful to solve illumination variations for indoor scenes under controlled lighting conditions.

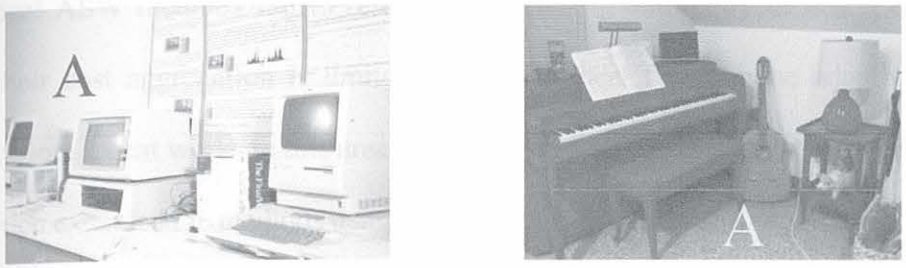


Figure 2.4: Low Texture Regions

Untextured and low texture region. Referring to Figure 2.4, the area that is denoted by the letter A contributes to the mismatching process caused by the plain colour and textureless surface regions. Thus, this can also occasionally produce a constant luminance in large areas. Consequently, it is more difficult and extremely tricky to develop the algorithm in the larger low texture regions due to the similarity of pixel intensities. A new technique, similar in principle to Lu et al. (2017) but using different monocular sensors and a hybrid Patch Match and Hash Match algorithm to estimate depth on mobile phones, was proposed by Valentin et al. (2018). The approach is more successful in the areas with no texture and is more robust to temporal changes. Chang and Maruyama (2018) employed the MBM algorithm with NCC which emphasised on improving the low texture regions and reducing the computational complexity when executed in real-time system. The matching algorithm by Liu et al. (2019) using adaptive support weight with two refinement steps achieved better result in low texture regions compare with the algorithm applying aggregation scheme called Pervasive Guided Image Filtering (PGIF) by Zhu and Chang (2019a). Research by Liu et al. (2021) described an interesting methodology using Two-Phase Adaptive Optimization of AD-Census and Gradient Fusion. This method performs

better in regions with disparity discontinuity and textureless regions and demonstrates sufficient robustness for radiometric changes and noise.

Local ASW methods fail to resolve the matching ambiguity in low texture areas because their cost aggregation is limited within the local fixed or the adaptive support windows. More recent works in this area were proposed by Wu et al. (2019) and Zhou et al. (2019) which extended the methods using a generic fusing Adaptive Support Weight (ASW) framework to determine dual support windows for each pixel including to specify the local window and the entire window.

Bethmann and Luhmann (2015) proposed a modified SGM method known as the path-wise minimisation method to assign into object space for the best matches of 3D positions. Another algorithm developed by Kordelas et al. (2016) exploited weighted semi-global optimisation to enhance the disparity estimation accuracy. The algorithm uses pixel-wise and a second order SGM on slanted plane iterative optimisation. This stereo algorithm is accurate when applied near the textured areas but produces bad matching result in the occlusion area due to the shortage of neighbouring pixels data (Ni et al.,2018). Xu et al. (2020) introduced a matching algorithm based on state measurement system with structure-driven cost volume fusion and data that contribute to good matching in the untextured regions.

The most widely used global methodology for low texture smoothing is the Dynamic Programming (DP) categorising the optimisation constraints into reduced and simpler subproblems. However, this method has the possible risk of local mistakes produced along a scan line, affecting the suitable matches (Men et al., 2015) and (Zhou, Wu and Zhu, 2016). Huang (2015) introduced an iterative optimisation algorithm, comprising of surface fitting, accelerated region BP and bi-cubic B-spline to solve ambiguous edges, textureless regions, and slow convergence speed. Further, Wang and Liu (2015) presented an algorithm

employing Initial-Value Belief Propagation (IVBP) combined with a Self-Adapting Dissimilarity Data Term (SDDT) to reduce the number of iterations and improve the energy terms and smoothness terms to ignore the significance of the data term. Taniai et al. (2018) proposed an alternative method, the Belief Propagation (BP) method which includes a belief aggregation in the algorithm. Nevertheless, Li et al. (2019) succeeded in solving the mismatch in high and low texture regions for disparity accuracy with the global SMA using fragment matching and tree structure.

Wang et al. (2016) proposed a deep conditional random field that extracts information and a connection between CNN and CRF. The first explored research in conditional adversarial networks was by Huang et al. (2017) that accomplished two conditional adversarial training processes to improve the high mismatches in textureless regions. Next, Ye et al. (2017), Williem and Park (2018) discovered that the disparity values can be extensively rectified using deep learning frameworks to improve the weak texture, discontinuities, illumination difference and occlusions. A more recent study by Hong and Ahn (2020) in a single-view videos (SMV) use an unsupervised method to build a deep learning-based algorithm for a more accurate depth estimation. The result showed excellent disparity and mismatching accuracy at the surface of low texture regions and depth boundaries.

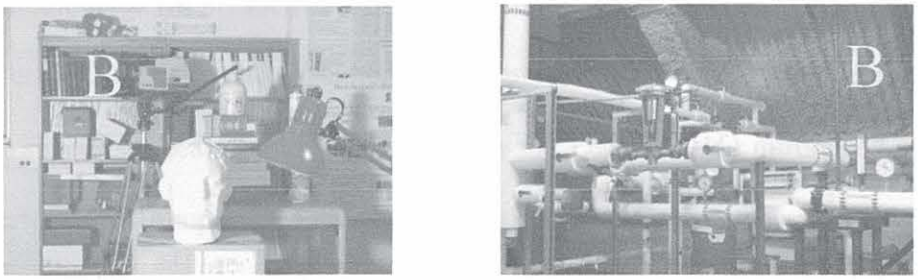


Figure 2.5: Repetitive Regions

Repetitive region. The areas denoted with B in Figure 2.5 represent the next constraint in the stereo matching algorithm. One of the problems that must be considered with the stereo matching algorithm is the area that contains periodic and repetitive surface texture. Human-made objects and space objects will typically consist of repetitive textures. This creates a challenge for the algorithm to process the object since wrong matching coordinates contribute to mismatching issues since these regions contain many possible intensity values.

Li et al. (2016) introduced a conceptual study on global method using DP and MRF to coordinate several disparity proposals to enhance the repetitive textures and surface fragments. Scharstein et al. (2018) also addressed this issue using the SGM and surface orientation priors, which operate as soft constraints throughout the matching process and only impose a low computation cost. The method illustrates that even a simple surface prior able to produce substantial gains in challenging indoor scenes with slanted surfaces and poor repetitive texture. The SGM method such as Khan et al. (2018) employed IGCM combined explicitly with a colour formation model handled the matching process quite effectively at the repetitive region. Furthermore, Wu et al. (2019) applied oriented linear tree structure for each pixel to accomplish a non-local cost aggregation technique. The study of Jafari Malekabadi et al. (2019) compared available algorithms for the geometric tree characteristics employed to forecast citrus crop water consumption, biomass, yield and fertiliser distribution. The local method algorithm, such as those produced by Wu et al. (2019), Du and Jia (2019) and Kong et al. (2021), generally resulted in poor matching outcome at the repetitive regions.

For supervised learning methods, Zbontar and Lecun (2016) introduced a network training performed in a supervised method by creating a set of binary classification data with dissimilar and similar patches pair samples that are used for obtaining depth information from a rectified image pair. To deal with the repetitive region, Dong et al. (2018) proposed a two-fold fusion structure that constructs wide-ranging cost volumes while Liang et al.

(2021) employed a CNN network-based algorithm with Multi-scale Feature Constancy and Multi-level Cost Volume in a three sub-modules, i.e., shared feature extraction, initial disparity estimation, and disparity refinement. .

For unsupervised learning methods, several approaches have been employed with the view synthesis and spatial transformation over the past few years for reliable and unreliable pixels in repetitive regions. Kim et al. (2019) introduced a deep model which measures stereo correspondence confidence consisting of a pooling module and residual networks in matching cost and, a deep, unified network.

Figure 2.6: Depth discontinuities regions

Depth discontinuity. The depth discontinuities problem is denoted as C in Figure 2.6. Valentin et al. (2018) and Li et al. (2019) studied the stereo matching algorithm's capability to determine the mask's size, including the stereo reference image to localise in the stereo target image. Further, Valentin et al. (2018) developed a hybrid Patch Match and Hash Match approach to estimate depth on mobile devices, which compromised edge errors but improved depth discontinuities. Usually, this will contribute to distortion across the depth boundaries and can be more challenging to find valid corresponding points if the size of the regions of stereo pair images has a massive difference between them. The support regions near the discontinuity consist of attributes from more than one depth. Thus, Xu et al. (2019) developed a disparity optimisation technique employing an improvement for a guided filter.

Lim and Lee (2019) and Zhang et al. (2019) established the minimum spanning tree and Patch Match technique while maintaining acceptable radiometric differences level with the aim of maintaining continuity at the non-edge and discontinuity at the edge.

Lin and Liu (2015) and Zeglazi et al. (2018) addressed the discontinuities problem through the use of a support window approach. In preserving depth discontinuity, Jung et al. (2015) introduced boundary-preserving stereo matching using adaptive disparity classification and particular region detection for a hierarchical form of matching. Buades and Facciolo (2015) dealt with depth discontinuities and slated surface in the image by employing oriented windows technique where the windows assist in determining disparities accurately on non-fronto parallel surfaces compared to the application of adaptive support windows. In contrast, Jellal et al. (2017) proposed an upgraded line segment extending to the efficient large scale stereo matching algorithm. Navarro and Buades (2019) employed non-dense local estimation that was filtered and interpolated using a novel variational formulation that made use of intermediate scale estimates of the local method. Due to the use of estimations at coarser scales, the small details can be approximated at full resolution while remaining robust to noise, depth discontinuity, and untextured areas.

The energy-minimisation problem for the fully connected model was solved by Mozerov and Van De Weijer (2015) using a two-step energy-minimisation algorithm with two MRFs, which was enhanced in occluded regions and depth discontinuities. Another global method to address this issue is described in Zhou et al. (2016) by solving the mismatching of disparity in depth-discontinuous regions. However, the researchers proposed to use different methods at the cost-aggregation stage which can be integrated with minimum spanning tree and 3D label search instead of using fixed-size patches as aggregation. Ye et al. (2017) proposed a more advanced algorithm to process the refinement with outliers classification into the left most occlusions, treatment of non-border occlusions, and a filling

order to prevent surface decision and error propagation. Moreover, Loghman et al. (2018) developed a modification on cost aggregation using the cross method in SGM to adaptively construct the cross's shape aimed specifically at each pixel.

Occlusion. In Figure 2.7, the area which is denoted with the letter D is the occluded regions, a standard type of constraint in stereo matching technology. Due to geometric displacement, one of the scenes is not visible to both cameras, making the image match pattern at the target image less visible compared to the reference image. Consequently, the stereo images cannot be matched if both cameras cannot see something. The occlusion regions contain unknown objects, shapes or structures that are problematic to estimate, producing low accurate disparity values. Zhang et al. (2020) presented a new development in local method to improve occlusion and discontinuity consisting of weighting process based on the gravity theory, which appropriately identifies the cost-based weighting coefficient for using the gravitational model.

Figure 2.7: Occluded Regions

Hu et al. (2018) developed an algorithm based on Efficient Large-scale Stereo (ELAS) for processing the parallax plan aimed at high mismatching amounts in image's occluded areas. Using a global approach, J. Li et al. (2018) proposed a relationship between long baseline and short baseline optimisation. Another global method used to overcome occlusion and foreshortening is Nonlinear Diffusion (ND). A study on multiple neighbours' nonlinear diffusion of costs aggregation was conducted with the aim to improve the global

cost function (D. D. Geng and N. Luo, 2017). Another method employed by (X. Ye et al., 2017) used two patch-based network architectures in matching cost computation that manipulates multi-layer and multi-size pooling unit with no strides to improve the occlusion recovery. Rathnayaka and Park (2020) proposed an optimisation algorithm using SGM, which improves the quality of occluded regions. These authors claimed that the Iterative Guided-Gaussian Multi-baseline method is effective and efficient in disparity values compared with original SGM algorithm.

Ye et al. (2017) used machine learning strategy to resolve a significant area of occlusions whereas the multistep post-processing leaves them unhandled. An algorithm of dense stereo matching approach using a novel CNN+CRF model for stereo estimation is presented by Knöbelreiter et al. (2017). For matching and distinguishing colour edges, CNNs produce expressive features that are used to determine the unary and binary CRF costs. The method is successful to enhance the occluded area's performance and recovers object boundaries. Lu et al. (2018b) undertook similar research and proposed a revolutionary architecture called a cascaded multi-scale and multi-dimension network (MSMD) of CNN. The result is a receptive field expansion capability of the algorithm and the loss of features in the area of occluded regions. Recent work by Bapat and Frahm (2019) applied a general MC-CNN optimisation framework that directly acts in the pixel space with an edge sensitive regulariser to improve MC-CNN approaches. The method offers significant disparity improvement for non-occluded pixels and outperforms conventional MC-CNN with little increased computational cost. Similarly, Mahato et al. (2019) developed the Genetic Algorithm (GA) in addition to multi-objective fitness function-based to solve occlusions, discontinuities, geometric, and radiometric distortions.

Computational complexity. Many studies in the stereo matching algorithm focus on improving computational complexity in terms of a software-based or hardware-based

algorithm. The challenge is to develop an algorithm or a design that is completely adaptable and fully utilises the capabilities of powerful computation. Further research should allow for the comparison of the output quality and the latency of dense stereo matching algorithm designs with several trade-offs (i.e., a reasonable assessment of the disparity map). Generally, local and SGM method such as Zhan et al. (2016), Khan et al. (2018), Loghman et al. (2018), Chang and Ho (2019) and Liu et al. (2019) provided low computational complexity due to the straightforward WTA optimisation implementation which compromises between time and quality, achieved output errors average between 4% and 20% while the execution times between 40ms and 1,400ms.

Historically, box filtering was amongst the first to be developed for stereo block matching that aims to replace each pixel respectively of an image to the average inbox size (Lee et al., 2013). A variant of this popular algorithm was proposed by Ttofis et al. (2016) that achieves high speed and reasonable quality. Zhang et al. (2015) addressed the segmentation form using the parallax plane and Delaunay triangulation principle to split the entire image in need of compensation between running time and matching accuracy. These researchers introduced an adaptive meshing to achieve real-time performance using several modifications such as sparse stereo correspondence and view synthesis.

In contrast, the global method's high computational complexity is a result of the numerous redundancies, which directly affects the execution time. Hence, algorithm from Cheng et al. (2015) , Zha et al. (2016), and Taniai et al. (2018) produced high execution time between 432ms and 34,290ms and quality average between 3.4% and 9.5%. Zhang et al. (2019) designed a better algorithm for stereo matching real-time application consisting of global-based optimisation on a single FPGA. The algorithm results showed a good cost of 30 frames per second (fps) for window 1920 x 1680.

Zbontar and Lecun (2016), Ye et al. (2017), Williem and Park (2018), Zhu et al. (2019) and Mahato et al. (2019) applied the ML strategy algorithm involving high computational complexity and run time but with a mixed results quality. The algorithms achieved an average quality of around 3.5% to 17.5% and execution times between 1300ms to 43700ms. Yang et al. (2017) introduced a matching Euclidean learning costs technique to include a triple-based loss function due to the CNN matching cost being computationally expensive and time-consuming. Further, Yang et al. (2020) applied a popular Non-End-to-End approach for the CNN and proposed to swap one or more steps in the traditional stereo framework to reduce the number of image primitives for subsequent processing. Another method proposed by Nguyen and Ahn (2020) using an evolutionary algorithm for parameter selection framework for stereo correspondence and trained in a supervised manner presented in a 3-D plane fitting to reduce the disparity search range.

This section summarises the most recent stereo correspondence constraints and the current algorithms developed from 2017 until 2022, as shown in Table 2.4. This table shows that most local methods produce poor or moderate performance in the areas of radiometric differences, low texture, repetitive patterns, and depth discontinuity compared with other methods. The ML and global methods performed greatly and accurately for radiometric differences, low texture, occlusion, and depth discontinuities, but performed poorly in the repetitive region. The ML method also contributed high computational complexity with almost similar performance as the global method. Meanwhile, the SGM method produces moderate performance for every stereo correspondence constraint, and most of the SGM algorithms have moderate to high computational complexity.

Table 2.4 : Summary of Constraint Level for The Recent Algorithm (2017-2022) Based on Middlebury and KITTI Results

No.	Author	Year	Method	Algorithm	Radiometric Differences	Low Texture	Repetitive	Discontinuity	Occlusion	Computational Complexity
1	(Ma et al., 2017)	2017	Local	Intrascale Cross-Scale CA + WLS	Moderate	Good	Good	Poor	Good	Moderate
2	(Hu et al., 2018)	2018	Local	LS - ELAS	Moderate	Poor	Moderate	Poor	Good	Low
3	(Ma et al., 2018)	2018	Local	CI-ELAS	Moderate	Moderate	Poor	Moderate	Good	Low
4	(Peng et al., 2018)	2018	Local	DAISY CA	Moderate	Moderate	Moderate	Poor	Poor	Moderate
5	(Chang and Maruyama, 2018)	2018	Local	MBM	Moderate	Moderate	Poor	Good	Good	Low
6	(Valentin et al., 2018)	2018	Local	Hybrid Patch Match and Hash Match	Poor	Moderate	Good	Moderate	Moderate	Low
7	(Fu et al., 2019)	2019	Local	Rank Encoding	Moderate	Moderate	Poor	Poor	Moderate	Moderate
8	(Liu et al., 2019)	2019	Local	ASW + Two Refinement	Moderate	Moderate	Poor	Good	Good	Medium
9	(Xu at al., 2019)	2019	Local	Improved GF	Moderate	Poor	Poor	Good	Moderate	-
10	(Zhu and Chang, 2019a)	2019	Local	PGIF	Moderate	Poor	Good	Moderate	Good	Medium
11	(Chang and Ho, 2019)	2019	Local	Modified ANCC	Good	Poor	Poor	Moderate	Poor	High
12	(Z. Zhang et al., 2020)	2019	Local	Gravitational	Good	Moderate	Poor	Moderate	Poor	High
13	(Du and Jia, 2019)	2019	Local	Neighbourhood Correlation	Good	Poor	Poor	Poor	Moderate	-
14	(Wu et al., 2019)	2019	Local	Fusing ASW	Good	Good	Poor	Moderate	Moderate	Medium

No.	Author	Year	Method	Algorithm	Radiometric Differences	Low Texture	Repetitive	Discontinuity	Occlusion	Computational Complexity
15	(Navarro and Buades, 2019)	2019	Local	AW Aggregation	Moderate	Moderate	Good	Poor	Moderate	Low
16	(Lim and Lee, 2019)	2019	Local	Patch Match	Good	Good	Moderate	Good	Good	Medium
17	(Ji et al., 2020)	2020	Local	QCT	Good	Moderate	Moderate	Poor	Poor	-
18	(Liu et al., 2020)	2020	Local	Gradient CT + Adaptive GF	Good	Moderate	Moderate	Good	Good	-
19	(Zhou et al., 2020)	2020	Local	CT and GF	Good	Moderate	Good	Moderate	Moderate	-
20	(Yuan et al., 2021)	2021	Local	Fast Domain GF	Good	Moderate	Poor	Good	Moderate	Medium
21	(Wei et al., 2021)	2021	Local	Multi-cost + Adaptive Cross Window	Good	Moderate	Poor	Poor	Good	-
22	(Liu et al., 2021)	2021	Local	AD-Census + Gradient Fusion	Moderate	Good	Moderate	Good	Good	-
23	(Kong et al., 2021)	2021	Local	ACR-GIF-OW	Moderate	Good	Moderate	Good	Good	High
24	(Liu et al., 2021)	2021	Local	AD-Census and Gradient Fusion	Moderate	Good	Moderate	Good	Good	High
25	(Hou et al., 2022)	2022	Local	iCT + Texture Filtering	Moderate	Good	Poor	Good	Poor	-
26	(Huang and Yang, 2022)	2022	Local	Quadruple Sparse CT + Adaptive Multi-Shape	Moderate	Moderate	Poor	Moderate	Moderate	-
27	(Qi and Liu, 2022)	2022	Local	CT + ASW	Good	Moderate	Poor	Poor	Good	-

No.	Author	Year	Method	Algorithm	Radiometric Differences	Low Texture	Repetitive	Discontinuity	Occlusion	Computational Complexity
28	(Loghman et al., 2018)	2018	SGM	Multi-scan Line	Moderate	Moderate	Moderate	Good	Moderate	Low
29	(Ni et al., 2018)	2018	SGM	Iterative Slanted Plane	Good	Good	Poor	Good	Moderate	-
30	(Hamzah and Ibrahim, 2018)	2018	SGM	GMD	Good	Good	Good	Moderate	Moderate	-
31	(Schönberger et al., 2018)	2018	SGM	SGM-Forest	Moderate	Good	Good	Good	Good	High
32	(Scharstein et al., 2018)	2018	SGM	SGM Surface Orientation Priors	Good	Good	Moderate	Good	Poor	High
33	(Khan et al., 2018)	2019	SGM	IGCM	Good	Poor	Moderate	Poor	Poor	Medium
34	(Wu et al., 2019)	2019	SGM	OLT	Moderate	Good	Good	Good	Moderate	Medium
35	(Rathnayaka and Park, 2020)	2020	SGM	Iterative MBS	Good	Good	Poor	Moderate	Good	High
36	(J. Xu et al., 2020)	2020	SGM	Cost Fusion + Smoothness	Moderate	Poor	Poor	Poor	Good	-
37	(Yao and Feng, 2021)	2021	SGM	Ensemble learning	Good	Good	Good	Moderate	Good	Medium
38	(Zhang and Huang, 2021)	2021	SGM	SGM-Edge	Good	Moderate	Good	Good	Good	Low
39	(Zhou et al., 2016)	2016	Global	Differential DP	Moderate	Good	Poor	Moderate	Poor	High
40	(J. Li et al., 2018)	2018	Global	Long Baseline	Good	Good	Moderate	Good	Good	Medium
41	(Taniai et al., 2018)	2018	Global	3D Local Exp	Good	Good	Moderate	Good	Good	High
42	(Y. Li et al., 2019)	2019	Global	Fragment Matching	Good	Moderate	Moderate	Good	Good	Medium

No.	Author	Year	Method	Algorithm	Radiometric Differences	Low Texture	Repetitive	Discontinuity	Occlusion	Computational Complexity
43	(Zhang et al., 2019)	2019	Global	MST	Moderate	Good	Moderate	Good	Good	-
44	(Cheng et al., 2015)	2020	Global	Cross Tree + Edge + Superpixel	Moderate	Good	Good	Moderate	Good	High
45	(Kerkaou et al., 2021)	2021	Global	Multi-cost + DP	Good	Moderate	Moderate	Poor	Good	Low
46	(Lu et al., 2021)	2022	Global	Improved Graph Cut	Moderate	Good	Good	Moderate	Poor	-
47	(Ye et al., 2017)	2017	ML	Order-Based	Good	Moderate	Moderate	Good	Good	-
48	(Ye et al., 2017)	2017	ML	Deep Local	Moderate	Good	Moderate	Good	Good	Moderate
49	(Knöbelreiter et al., 2017)	2017	ML	Hybrid CNN-CRF	Good	Good	Good	Good	Moderate	Moderate
50	(Williem and Park, 2018)	2018	ML	Deep Self-Guided	Good	Good	Good	Moderate	Poor	High
51	(Dong et al., 2018)	2018	ML	Dual Fusion	Moderate	Good	Good	Good	Good	High
52	(Zhu et al., 2019)	2019	ML	CNN + WTADP	Good	Poor	Moderate	Good	Poor	Medium
53	(Liang et al., 2021)	2019	ML	CNN	Good	Good	Moderate	Good	Good	High
54	(Bapat and Frahm, 2019)	2019	ML	DTS	Good	Good	Good	Good	Moderate	Medium
55	(Lu et al., 2018b)	2019	ML	Cascaded Multi-Scale CNN	Moderate	Poor	Poor	Moderate	Good	Low
56	(Joung et al., 2020)	2020	ML	Unsupervised CC	Good	Good	Poor	Poor	Good	High
57	(Liang et al., 2021)	2021	ML	Trainable CNN	Good	Poor	Poor	Good	Good	Medium

No.	Author	Year	Method	Algorithm	Radiometric Differences	Low Texture	Repetitive	Discontinuity	Occlusion	Computational Complexity
58	(Jia et al., 2021)	2021	ML	MSCVNet CNN	Good	Good	Good	Moderate	Good	Low
59	(Yang et al., 2022)	2022	ML	CNN RDNet	Good	Moderate	Moderate	Good	Good	Low

Based on the literature review in Section 2.3, Section 2.4, and Table 2.4, there are opportunities in several areas that can be explored and improved for this work. It shows that to improve the issues of different illumination at boundaries and identical intensity levels, the work from Wei et al. (2021) and Liu et al. (2021) could be the baseline to explore a new unified approach for a matching cost computation based on multi-cost matching, especially the introduction of enhanced CT and pyramid cost functions, which are new to the current local method. This new approach also aims to solve the matching cost window selection problem for low texture and depth discontinuity regions. Then, the work from Kong et al. (2021) will be the study for a new proposed method for cost aggregation and refinement stages by introducing a generic method that has been developed to solve a variety of challenges, including depth discontinuities and occluded areas. This generic method could be a combination of random walk and filter methods at the cost aggregation, eliminating the problem selection of window aggregation, while for the refinement, a clustering and smooth filter can be used to remove the remaining noise in the algorithm.

2.5 Stereo Vision Dataset

A stereo vision dataset is a collection of images captured from two or more cameras placed at different viewpoints, simulating the human binocular vision. These datasets are commonly used in computer vision research and applications to develop and evaluate algorithms related to depth perception, 3D reconstruction, object detection, and scene understanding. In a stereo vision setup, the cameras are typically calibrated such that their relative positions and orientations are known. This calibration allows for the calculation of the disparity map, which is a per-pixel measure of the horizontal shift between corresponding points in the two images. Disparity maps can be used to estimate the depth information of

objects in the scene. Stereo vision datasets are crucial for training and testing algorithms designed for tasks such as:

a. Depth Estimation: Using the disparity information, algorithms can estimate the depth of each pixel in the scene, creating a depth map.
b. 3D Reconstruction: By triangulating corresponding points in the stereo images, it's possible to reconstruct a 3D representation of the scene.
c. Object Detection and Segmentation: Depth information from stereo images can enhance object detection and segmentation algorithms, improving their accuracy and robustness.
d. Scene Understanding: Stereo vision helps in understanding the 3D structure of the environment, aiding in tasks like scene understanding, navigation, and obstacle avoidance.
e. Augmented Reality: Stereo vision datasets are used to improve the accuracy of augmented reality applications, aligning virtual objects more accurately with the real world.

There are several stereo vision datasets that have been widely used by researchers for developing and evaluating stereo matching algorithms such as Scene Flow (Wedel et al., 2011), Middlebury (Szeliski, 2020) and KITTI (Geiger et al., 2020). These datasets provide ground truth disparity maps and serve as benchmarks for assessing the performance of different stereo algorithms. In this work, there are two standard datasets used to formulate and evaluate the SMA. The Middlebury Stereo Evaluation dataset is one of the most widely used benchmarks for the development and evaluation of stereo matching algorithms. It has played a crucial role in advancing the field of stereo vision by providing standardized test cases and ground truth data for evaluating the performance of various stereo algorithms. The

dataset consists of multiple versions, each covering different challenges and characteristics commonly encountered in stereo vision problems. Key features of the Middlebury Stereo Evaluation dataset include:

a. Image Pairs: The dataset includes a collection of stereo image pairs captured from different scenes and viewpoints. These images are specifically selected to represent various difficulties such as occlusions, repetitive patterns, textureless regions, and depth discontinuities.

b. Ground Truth Disparity Maps: One of the most important aspects of the Middlebury dataset is that it provides pixel-level ground truth disparity maps for each stereo pair. These maps are generated using accurate depth measurement techniques, and they serve as the reference for evaluating the accuracy of stereo matching algorithms.

c. Different Versions: The dataset has undergone several versions, each introducing new challenges and scenes. The dataset versions include Middlebury 2001, Middlebury 2003, Middlebury 2005, Middlebury 2006, and Middlebury 2014. Each version adds more complexity to the scenes, making it progressively more challenging for algorithms to achieve accurate disparity estimation.

d. Evaluated Metrics: The Middlebury dataset comes with standardized evaluation metrics that allow researchers to quantitatively compare the performance of stereo algorithms. Common metrics include bad pixel error, percentage of bad pixels, and root mean squared error between the algorithm's output disparity map and the ground truth.

e. Benchmarks and Competitions: The Middlebury dataset has been used in various stereo matching algorithm competitions and benchmarking challenges. These

events provide a platform for researchers to showcase their algorithms and assess their performance against others in the field.

f. Impact on Research: The Middlebury dataset has significantly influenced the development of stereo matching algorithms. Researchers often use this dataset to test the robustness and accuracy of their algorithms, as well as to identify areas for improvement.

g. Challenging Scenarios: The dataset covers a range of challenging scenarios that real-world stereo systems might encounter. This includes cases where objects are at varying depths, have different textures, and exhibit occlusions.

Due to its well-defined challenges, comprehensive ground truth data, and established evaluation metrics, the Middlebury Stereo Evaluation dataset has been instrumental in driving advancements in stereo vision research. It has become a standard reference for evaluating stereo matching algorithms and has contributed to the development of more accurate and robust techniques for depth estimation and 3D reconstruction. Then, The KITTI Vision Benchmark Suite is also a widely used dataset for the development and evaluation of various computer vision algorithms, including stereo matching, in the context of autonomous driving. It's known for its realistic and challenging scenes, making it a valuable resource for testing algorithms under real-world conditions. The dataset provides a comprehensive set of stereo image pairs, along with ground truth data, enabling researchers to develop and benchmark stereo matching algorithms effectively. Key features of the KITTI Vision Benchmark Suite include:

a. Stereo Image Pairs: The KITTI dataset includes a collection of stereo image pairs captured from a car-mounted rig while driving in urban and highway

environments. These images cover a wide range of scenarios, including different lighting conditions, weather, traffic situations, and road types.

b. High-Resolution Images: The images in the dataset are of high resolution, capturing fine details in the scene. This helps researchers address challenges related to small objects, textureless regions, and complex scene structures.

c. Calibration Data: The dataset provides accurate camera calibration parameters, including intrinsic and extrinsic parameters, which are essential for accurate stereo matching and 3D reconstruction.

d. Ground Truth Disparity Maps: The KITTI dataset includes pixel-level ground truth disparity maps for a subset of the stereo image pairs. These disparity maps are generated using a LiDAR sensor and serve as reference data for evaluating the accuracy of stereo matching algorithms.

e. Object Annotations: In addition to stereo images and disparity maps, the dataset includes object annotations for tasks like object detection and tracking. This makes the dataset suitable for evaluating multi-modal algorithms that combine stereo and object detection.

f. Odometry and Mapping Data: The dataset also provides odometry and mapping data, enabling researchers to evaluate algorithms related to localization and mapping.

g. Benchmarks and Challenges: The KITTI dataset has been used for benchmarking and evaluating stereo matching algorithms in various challenges and competitions. This provides researchers with an opportunity to compare their algorithms against others in the field.

h. Realistic Driving Scenarios: The dataset captures real-world driving scenarios, making it particularly relevant for algorithms designed for autonomous vehicles.

It includes scenarios like driving in urban environments, on highways, and dealing with challenging lighting and weather conditions.

These datasets have been extensively used by researchers to develop and compare stereo matching algorithms, evaluate their performance under different conditions, and push the boundaries of stereo vision technology. They provide standardized evaluation metrics and ground truth data that enable fair comparisons between different algorithms, making it easier to track progress in the field of stereo vision research.

2.6 3D Surface Reconstruction based on Stereo Vision System

The method presented in this study is sufficiently general to be applied to a stereo vision sensor that will act as a passive optical system. This system incorporates the coordinates of an object-based only on the information obtained by the input images. The stereo vision system uses the triangulation principle to define the depth and 3D coordinates system in space based on the stereo image projections. Therefore, this process requires stereo camera parameters.

Figure 2.8 shows a flowchart of the 3D surface reconstruction based on the work by Fan et al. (2018) employing fast bilateral stereo for road surface reconstruction. The perspective view of the target image is transformed into the reference view, which significantly enhances the road surface similarity between the reference and target images. Normalised Cross-Correlation (NCC) is used to measure the similarity between each pair of the selected blocks. The computed correlation costs are stored in two 3D cost volumes. To adaptively aggregate the neighborhood systems' correlation costs, bilateral filtering is performed on the two cost volumes. Finally, the estimated disparity map is post-processed. The 3D surface is reconstructed, whereas each 3-D point $p^W = [x^W , y^W , z^W]^T$ can be

computed from its projections $p_L = [u_l, v_l]^T$ and $p_r = [u_r, v_r]$ using the intrinsic and extrinsic parameters of the stereo system, where v_r is equivalent to v_l, and u_r is associated with u_l by d.

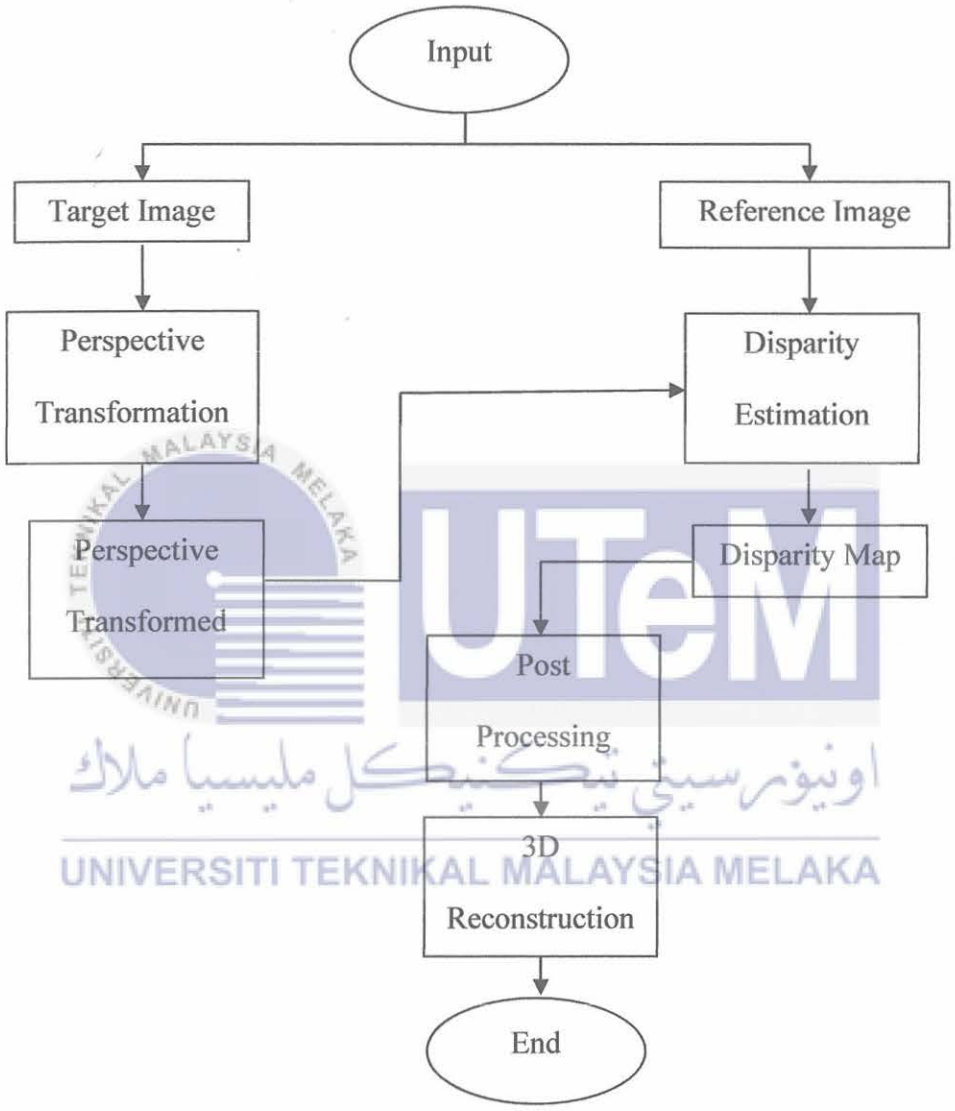


Figure 2.8: A Flowchart of 3D Surface Reconstruction Based on Fast Bilateral Stereo

2.7 Summary

The development of stereo matching algorithms remains a challenge for researchers. Hence, a growing number of new approaches are developed to achieve a good disparity map

or increase processing speed. Acquiring familiarity with the state-of-the-art algorithms is challenging and time-consuming. This chapter discussed the significant constraints of stereo vision and their characteristics. This information is useful as a reference when designing and developing a new algorithm with better performance in terms of the disparity map quality and computation.

Generally, research on local approaches mainly focuses on increasing the disparity map's quality by improving the algorithms in terms of cost computation, cost aggregation, or the refinement processes. Conversely, the global approach aims to increase the accuracy with more calculations and complexity or to decrease the computational load by simplifying algorithms without affecting the disparity map quality or, if possible, improving it further. Appropriate hardware design can greatly accelerate the algorithm runtimes. Although the global approach has received more attention than the local method recently due to higher accuracy, current existing local approaches have yet to reach their potential limits. With appropriate occlusion, homogenous areas, and discontinuity handling, a local approach has the potential to determine the comparable disparity maps with limited information within constrained search space areas.

CHAPTER 3

RESEARCH METHODOLOGY

This chapter presents the methods used in this research for a new Stereo Matching Algorithm (SMA) and the detail reasoning of the methodology. This chapter is divided into seven sections. The chapter begins with the flow chart of this work and a brief overview of the proposed SMA, are explained in Section 3.1 and Section 3.2. Then, Section 3.3 describes the architecture of the new Multi-Cost Pyramid Fusion (MPF) in the matching cost computation. Further, the technique to aggregate the cost function based on Hybrid Random Aggregation (HRA) is discussed in section 3.4. This is followed by Section 3.5 presenting the strategy to optimise the algorithm and the disparity selection. Next, Section 3.6 provides the disparity refinement stage of hierarchical cluster-edge based on occlusion filling and Side Window Filter (SWF) to produce an accurate disparity value. Section 3.7 discusses the reconstruction of 3D surface while Section 3.8 describes the experimental evaluation used in this work. Finally, section 3.9 provides the summary of this chapter.

3.1 Flow Chart of the Research

This section explains all the procedures of the research as presented in Figure 3.1 to produce an accurate disparity map. The procedures heavily rely on software for algorithm development and, an experimental measurement is used to validate them against a standard online benchmarking database. To achieve the research objectives, a new functional SMA is developed which mark the starting point for this research with the aim to produce the disparity map. In essence, the algorithm processes the stereo images of reference and target

to produce the disparity map. To evaluate the accuracy, the experimental images used is primarily a standard online benchmarking dataset from the Middlebury. This dataset consists of 15 training images that are used to produce the final disparity map. The accuracy is measured according to error pixel percentage of non-occluded pixel (*nonocc*) and all pixels (*all*) through the quantitative measurement from the disparity map that is uploaded to the online benchmarking evaluation system. The parameters are unknown, and thus need to be estimated through manipulation, modification of parameters in equations and with continuation of experimental validation until the optimal accuracy is accomplished. This process has improved the quality of the results significantly.

The parameter is varied between stages depending upon the disparity map accuracy result. In this research, there are three stages in the equations that are improved based on the SMA stages: the matching cost computation, the cost aggregation and the disparity refinement from the original four stages of SMA. The process begins with the first parameter from the first stage of the algorithm where the parameter is adjusted and tuned repetitively until the best accuracy is achieved. When the first parameter is completed, the process is repeated with the second and third parameters until the final and the most accurate disparity map is obtained. Then, the finalised SMA is processed with the Middlebury test dataset. During the parameter's adjustments, additional qualitative data is gathered to address any gaps based on the complex scene's scenario set up based on internal and external environments. Next, the finalised SMA is processed and corresponded into disparity map using the KITTI Stereo datasets. These results must be recorded and analysed. Furthermore, the 3D surface reconstruction is selected as an application for the SMA which is based on the aid of Open CV library software. Finally, the results of the reconstruction are then documented and presented in this thesis.

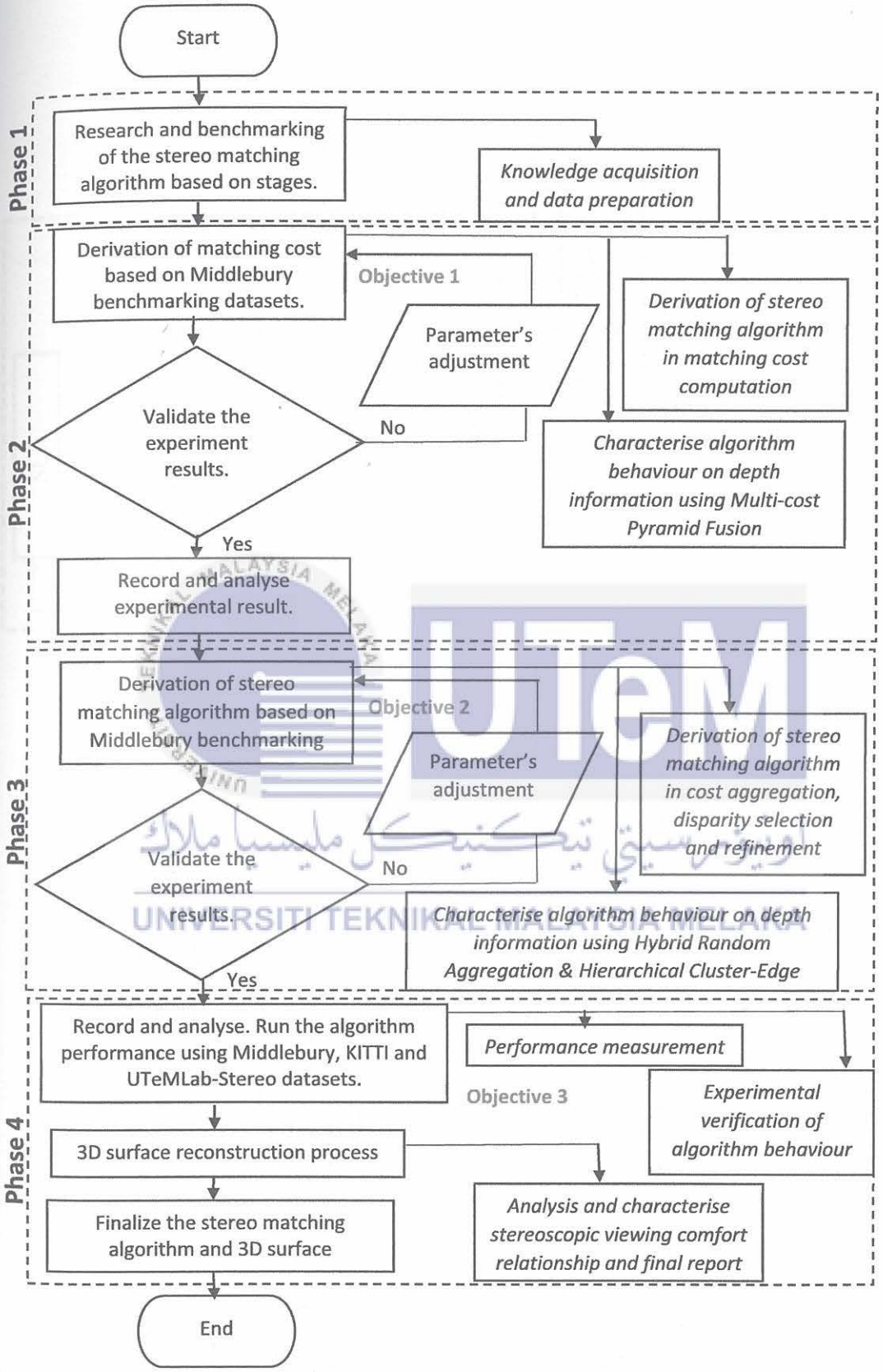


Figure 3.1: The Flowchart for The Research

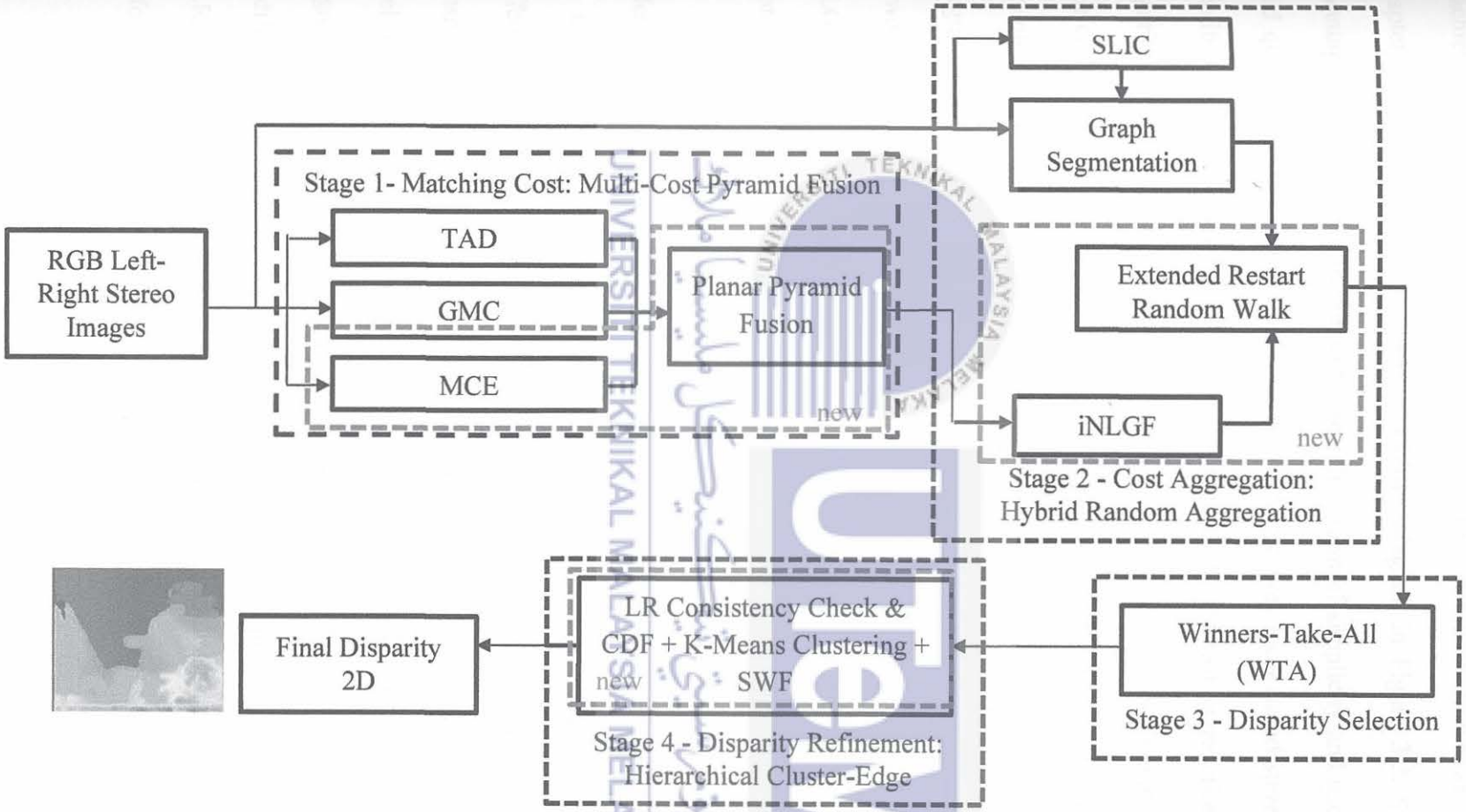


Figure 3.2: Proposed SMA Block Diagram

3.2 Overview of the Proposed SMA

A new SMA based on local method is proposed to produce an accurate disparity map to achieve the research objectives and the stereo vision constraint factors as explained in the Chapter 1 and Chapter 2 as shown in a block diagram in Figure 3.2. Such method is advantageous in a variety of areas of research in term of simplicity and usually delivers a good quality result. The taxonomy of the SMA involves local based stereo matching as described in Chapter 2 from the matching cost (Stage 1), the cost aggregation (Stage 2), the disparity selection (Stage 3) and the disparity refinement (Stage 4). Figure 3.2 shows the main contributions to this work in red boxes. These include the implementation of multi-cost pyramid fusion, especially the derivation of the new modified census edge method, and the integration of planar pyramid fusion, which is the integration of costs along the pyramid network at Stage 1. Then, the derivation of a hybrid technique that combined pixel cost from iNLGF and segment cost from eRWR at Stage 2 Lastly, the new implementation of K-mean clustering and SWF as hierarchical cluster-edge refinement at Stage 4.

The new matching cost computation involves multi-cost combination of per pixel differences measurement along with a new derivation of Planar Pyramid Fusion (PPF). There are three differences in matching cost measurements using the Truncated Absolute Differences (TAD), Gradient Magnitude CLAHE (GMC) and a new formulation of Modified Census Edge (MCE) followed with the combination of these costs using the PPF to produce pixel cost volume. Then, a new proposed of cost aggregation based on the Hybrid Random Aggregation (HRA) is implemented. This facilitates a hybrid type approach that utilizes modified Iterative Non-Local Guided Filter (iNLGF), Simple Linear Iterative Clustering (SLIC), Graph Segmentation (GS) and Extended Restart Random Walk (eRWR) for optimal performance.

After that, a Winner-Take-All (WTA) approach is implemented to select the location of minimum aggregated value corresponding to the disparity value for each pixel. Then, a refinement stage is explored to detect and to replace the outlier pixels (i.e., invalid pixel) with a valid selected disparity values at the occlusion and low texture regions by Left-Right (LR) consistency checking process and Confidence Disparity Filling (CDF). During the refinement stage, also known as the hierarchical cluster-edge refinement, the K-means clustering is used to recover the low texture regions in the disparity map and the Side Window Filter (SWF) is used to remove the remaining noises.

3.3 Matching Cost Computation Stage

The computation of matching cost is performed as the initial stage of the SMA algorithm development. These well-known parametric-based cost methods, including Absolute Differences (AD), Sum of Absolute Differences (SAD), Normalized Cross Correlation (NCC) have a significant risk to radiometric distortions. As demonstrated by Ji et al. (2020), this issue is typically overcomed by the standard non-parametric-based cost approach, such as the Census Transform (CT). However, due to the CT's sensitivity to intensity issue, the local window determination could possibly introduce salt-and-pepper noise in the matching cost. Similarly, Rathnayaka and Park (2020), and Jang et al. (2022) reported low accuracy issue in the disparity map, particularly in the regions with non-occluded and low texture. In this research, an effective approach to address these issues is proposed by employing a Multi-Cost Pyramid Fusion (MPF) which is the pyramid combination from the three components of the parametric and the non-parametric cost methods. The term "fusion" refers to the technique of combining the cost volume in each layer along the pyramid network. This MPF aims to minimise and reduce the algorithm's inaccuracy.

The three cost components include the Truncated Absolute Differences (TAD), Gradient Magnitude CLAHE (GMC), and a new Modified Census Edge (MCE). Later, the Planar Pyramid Fusion (PPF) is applied to integrate these costs to develop the pixel cost volume. Originally, the traditional GM solution could handle stereo images with radiometric changes (Li, Zhang and Gao, 2020). However an enhanced image gradient-based matching cost GMC is applied due the strong robustness against radiometric distortion and noises (Liu et al., 2020). The third component in the matching cost is the MCE, a modified version of the traditional CT that was previously prone to being disturbed by illumination distortion and noise in complex environments (Ji et al., 2020) and (Qi and Liu, 2022).

The TAD is based on the intensity differences between two RGB pixels at the left image I_l and right image I_r as demonstrated by Ma et al. (2016) and Cao et al. (2019). The initial function can be expressed as a combination of $C_{RGB}(p, d)$ and $C_{GRAD}(p, d)$ which are the RGB cost channels from pixel intensity differences and gradient differences. This is presented in equation (3.1) and (3.2) as follows:

$$C_{RGB}(p, d) = \frac{1}{3}\sum_{p\in(r,g,b)} \min|I_l(p) - I_r(p - d)|, \quad (3.1)$$

$$C_{GRAD}(p, d) = \min|(\nabla I_l(p) + 0.5) - (\nabla I_r(p - d) + 0.5)|, \quad (3.2)$$

where the pixel of interest's coordinates (x, y) is represented by the p and the d of the disparity value. The gradient cost computation in ∇ horizontal direction between -0.5 to 0.5 in the range of [0,1] is based on the parameter setting by Ma et al. (2016). The absolute difference function $AD(p, d)$ is given in equation (3.4) as follows:

$$AD(p, d) = (1 - \sigma_{AD})C_{RGB}(p, d) + \sigma_{AD}C_{GRAD}(p, d), \quad (3.3)$$

where σ_{AD} is a constant value to modify the pixel-by-pixel intensity differences applied by Wang and Xie (2018). This AD weighting aims to balance the traits of absolute color and

gradient, as well as to minimise the sensitivity to radiometric differences. The final value of the TAD is expressed as follows:

$$\mathrm{TAD}(\mathrm{p},\mathrm{d}) = \begin{cases} \tau_{\mathrm{TAD}}, & \text{if } \mathrm{AD}(\mathrm{p},\mathrm{d}) > \tau_{\mathrm{TAD}}, \\ \mathrm{AD}(\mathrm{p},\mathrm{d}), & \text{otherwise}, \end{cases} \tag{3.4}$$

where τ_{TAD} indicates the truncated value for the TAD cost as employed by Cao et al. (2019) to improve the resilience towards outliers of the stereo matching. The truncated value in equation (3.4) is the limiting parameter which constraints the cost to a certain maximum value.

The GM differences methods are well established and are described by Ende et al. (2018) and Dong and Feng (2018). The proposed GMC is the extended version of GM by enhancing the input stereo images using Contrast Limited Adaptive Histogram Equalization (CLAHE) (Liu et al., 2020). This is the standard technique in histogram equalisation $\mathrm{HE}(\mathrm{p})$ to acquire a better edge information as presented in equation (3.5):

$$\mathrm{HE}(\mathrm{p}) = \mathrm{CLAHE} \sum_{\mathrm{j}=0}^{\mathrm{k}} \frac{\mathrm{n}(\mathrm{i}(\mathrm{p}))_{\mathrm{j}}}{\mathrm{n}(\mathrm{I}(\mathrm{p}))} . (\mathrm{L}-1), \tag{3.5}$$

where p is the coordinate (x, y) pixel of interest. k is the number of gray levels from 0 until L-1 and L indicates the maximum gray levels in the image. $\mathrm{n}(\mathrm{i}(\mathrm{p}))_{\mathrm{j}}$ denotes the number of times j-th gray level exists while $\mathrm{n}(\mathrm{I}(\mathrm{p}))$ is total number of pixels in the image. The magnitude value of the left grayscale CLAHE images $\mathrm{HE_l}(\mathrm{p})$ and right grayscale CLAHE image $\mathrm{HE_r}(\mathrm{p}-\mathrm{d})$ is used by the GM component as shown in equation (3.6).

$$\begin{aligned} \mathrm{Gr}(\mathrm{p},\mathrm{d}) = {} & (|\nabla_{\mathrm{x}} \mathrm{W_{GM}} * \mathrm{HE_l}(\mathrm{p})| - |\nabla_{\mathrm{x}} \mathrm{W_{GM}} * \mathrm{HE_r}(\mathrm{p}-\mathrm{d})|) \\ & + (|\nabla_{\mathrm{y}} \mathrm{W'_{GM}} * \mathrm{HE_l}(\mathrm{p})| - |\nabla_{\mathrm{y}} \mathrm{W'_{GM}} * \mathrm{HE_r}(\mathrm{p}-\mathrm{d})|), \end{aligned} \tag{3.6}$$

where d denotes the disparity value and $*$ operator represents the convolution between gradient magnitude and Sobel operator. ∇_x and ∇_y indicate the directional gradient magnitude for horizontal and vertical. The new proposed Gr is modified with the

implementation of W_{GM} indicating the 5 x 5 window of Sobel operator to emphasise the image edges while W'_{GM} is the transpose of Sobel operator. τ_{GM} is the truncated value for the GMC which acts as the maximum value of the final GMC as applied by Hamzah et al. (2020) and expressed as equation (3.7) as follows:

$$\mathrm{GMC(p,d)} = \begin{cases} \tau_{\mathrm{GM}}, & \text{if } |\,\mathrm{Gr(p,d)}| > \tau_{\mathrm{GM}}, \\ |\,\mathrm{Gr(p,d)}|, & \text{otherwise}, \end{cases} \tag{3.7}$$

As demonstrated by Tabssum et al. (2017) and Hou et al. (2022), the CT pattern does not take into account the traditional CT relationship between adjacent pixels; instead, it simply considers the single-pixel brightness variation between its neighbours. As a result, the CT pattern will lose some image information during conversion. For instance, by comparing the middle point 54 with the neighbour points 42 and 22, respectively, two CT values can be determined: (54 (center), 42(neighbour) 1), and (54 (center), 22(neighbour) 1). The computation process shows that these two CT values do not precisely reflect the variations in brightness between 42 and 22. For a brightness change between the center point and the other point, only one point from these two CT values can be identified, and this point can be reflected throughout the entire image. Hence, this research provides a new MCE approach to address this issue especially at the boundary and low texture regions.

Initially, the input stereo images are first applied with the edge enhancement to instantly increase the image contrast in the area especially around the edge and in the region of low texture as employed by Jin and Wei (2022). The edge enhancement method for the horizontal and vertical direction are presented in equation (3.8), (3.9) and (3.10) as follows:

$$\mathrm{H(i,j)} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{3.8}$$

$$\mathrm{EG_x(p)} = \mathrm{EG_x^0} + \mathrm{H(i,j)} * \mathrm{I(p)}, \tag{3.9}$$

$$\mathrm{EG_y(p)} = \mathrm{EG_y^0} + \mathrm{H^T(i,j)} * \mathrm{I(p)}, \tag{3.10}$$

where * operator denotes the modulus function and p is the (x, y) coordinate pixel of interest. The pixel intensity is presented as I(p). EG_x^0 and EG_y^0 are designated as the initial edge value directions. H(i, j) is the Gaussian kernel applied to the edge enhancement horizontal direction to enhance and preserve edges in noisy images. Meanwhile, the vertical direction is presented by transpose Gaussian kernel $H^T(i,j)$. Equation (3.11) provides the gradient magnitude using the gradient components $EG_x(p)$ and $EG_y(p)$ as follows:

$$g(p) = \sqrt[2]{EG_x(p)^2 + EG_y(p)^2}, \tag{3.11}$$

The final value of the image edge enhancement method is determined by threshold parameter, and this can be described through an equation (3.14) as follows:

$$m(p) = \begin{cases} 255, & \text{if } g(p) \geq \tau_{edge} \\ g(p), & \text{if } g(p) < \tau_{edge} \end{cases}, \tag{3.14}$$

where τ_{edge} signifies the threshold parameter. In order to address the limitation of the traditional CT, a modification is done to compensate the CT sensitivity towards the intensity due to illumination changes, which allows equation (3.15) and (3.16) to be modified as follows:

$$CN(p) = \otimes_{q \in w_{CE}} \xi(I_c(p), I(q)), \tag{3.15}$$

$$ED(p) = \otimes_{q \in w_{CE}} \xi(m_c(p), m(q)), \tag{3.16}$$

where $\otimes$ operator refers to a bit-wise catenation and w_{CE} is the support window of the census block. CN(p) and ED(p) denote the census cost original image and edge image. Both these CN and ED consist of local information which allows computing between the two bit-string under Hamming distance. The coordinates of the pixel of interest and the neighbours are indicated by the letters p and q, respectively. $I_c(p)$, $m_c(p)$ represent the intensity of center pixel while I(q), m(q) are the intensity at the neighbour pixels. This thesis proposed a new approach to establish the value of the center pixel for the census code by employing

the error threshold and the adaptive center pixel. The determination of the new middle point intensity value for the overall MCE is represented in equation (3.17) and (3.18) as follows:

$$I_m(q) = \frac{1}{k}\left[\sum_{i=1}^{j}(I_j(q))\right], \tag{3.17}$$

$$E_m(q) = \frac{1}{k}\left[\sum_{i=1}^{j}(m_j(q))\right], \tag{3.18}$$

$$k = \frac{N-1}{2}, \tag{3.19}$$

where $I_m(q)$ with $E_m(q)$ represent the census texture and census edge neighbour intensity determination. The value of k indicates the number of odd neighbour pixels from total number of neighbour pixels, while N in the census window w x w. The $I_j(q)$ and $m_j(q)$ present the odd neighbour pixel intensity value at j-th location. The j denotes maximum number of q-th location of the pixels based on $j = 2k + 1$ calculation. The standard CT's sensitivity to intensity is compensated by finding a new middle point intensity value that is adaptively adjusted based on the odd neighbour pixel intensity values. Then, the differences of the neighbour intensity value I_{diff} and m_{diff} are calculated based on equation (3.20) and (3.21) for both channels as follows:

$$I_{diff}(p) = |I(p) - I_m(q)|, \tag{3.20}$$

$$m_{diff}(p) = |m(p) - E_m(q)|, \tag{3.21}$$

where I(p) and m(p) are the intensity of center pixel for both census texture and edge. Then, an error threshold τ_{diff} is proposed to determine the new value of center pixel intensity which would be used in the binary function. Some minor modifications are made to the method described in Lv et al. (2021) which the value of difference between the I(p), m(p) with $I_{diff}(p)$ and $m_{diff}(p)$ is greater than τ_{diff}, $I_{diff}(p)$ or $m_{diff}(p)$ is employed for the central pixel instead of I(p) or m(p). Else, the central pixel for census encoding; I(p) or m(p) as shown in the equation (3.22) and (3.23) as follows:

$$I_{new}(p) = \begin{cases} I(p), & \text{if } I_{diff} < \tau_{diff} \\ I_{diff}(p), & \text{otherwise} \end{cases}, \quad (3.22)$$

$$E_{new}(p) = \begin{cases} m(p), & \text{if } m_{diff} < \tau_{diff} \\ m_{diff}(p), & \text{otherwise} \end{cases}, \quad (3.23)$$

The census encoding later determines the binary function for the census texture and edge where the binary function corresponding to the new intensity value of center pixel $I_{new}(p)$ and $E_{new}(p)$ with the neighbour pixel I(q) and m(q), respectively, in the range of 3 x 3 support window, w_{CE}. The binary function is expressed in equation (3.24) and (3.25) as follows:

$$\xi(I_c(p), I(q) = \begin{cases} 0, & \text{if } I_{new}(p) < I(q) \\ 1, & \text{otherwise} \end{cases}, \quad (3.24)$$

$$\xi(m_c(p), m(q) = \begin{cases} 0, & \text{if } E_{new}(p) < m(q) \\ 1, & \text{if otherwise} \end{cases}, \quad (3.25)$$

where ξ is the mapping function of the MCE. Determining the difference between the two-bit strings, the left and right images for census texture and census edge, is accomplished through Hamming distance, which is represented by equation (3.26) as follows:

$$MCN(p, d) = \text{Hamming}(\sigma_{CN} |CN(p) - CN(p - d| + \sigma_{ED} |ED(p) - ED(p - d)|), \quad (3.26)$$

where σ_{CN} and σ_{ED} are the weightage for census texture and edge to balance the texture and edge term. The weightage technique presented is capable of producing realistic controls for the sensitive census function and amplitude distortion. τ_{CN} is the truncated value for the MCE, which serves as the maximum value of the final MCE as implemented by Ma et al. (2016). This is written inequation (3.27) as follows:

$$MCE(p, d) = \begin{cases} \tau_{CN}, & \text{if } |MCN(p, d)| > \tau_{CN}, \\ |MCN(p, d)|, & \text{otherwise}, \end{cases} \quad (3.27)$$

Three similarity measures in the matching cost for the combined TAD, GMC and MCE provide different aspects of the raw cost data. Therefore, integrating these elements is essential for an accurate evaluation and to enhance matching implementation. Ende et al.

(2018) and Liu et al. (2020) proposed a standard normalised cost function to combine with the final matching cost function. However, an alternative method is proposed in this work using the PPF as the combination of cost function based on the study by Zhang et al. (2022). As shown in Figure 3.3, cost volume decomposition is generated from the Gaussian pyramid (GP), that is a multiresolution cost reconstruction achieved through recursive reduction of the cost raw data from decimation and lowpass filtering.

The matching layer of the Gaussian pyramid is determined by the difference between one layer and the expanded cost volume generated from the immediate top layer. The expanded Gaussian Pyramid is a structure composed out of the expanded cost. The pyramid fusion equation initially starts with the Gaussian pyramid which the sequence of reduce cost volume generated from the original cost volume by a reduction function. The REDUCE function is given by equation (3.28), (3.29) and (3.30) as follows:

$$\mathrm{MCE}_{k+1}(p,d) = \mathrm{REDUCE}\big(\omega \mathrm{MCE}_k(p,d)\big), \tag{3.28}$$

$$\mathrm{GMC}_{k+1}(p,d) = \mathrm{REDUCE}\big(\omega \mathrm{GMC}_k(p,d)\big), \tag{3.29}$$

$$\mathrm{TAD}_{k+1}(p,d) = \mathrm{REDUCE}\big(\omega \mathrm{TAD}_k(p,d)\big), \tag{3.30}$$

where k indicates the number of k-th layers in the pyramid while ω presents the weighting function which has similar setting with work by Shengxy (2015). Then, the pyramid is extended to a sequence of expanded cost volume generated from the REDUCE function. This can be estimated as described in equation (3.31), (3.32) and (3.33) as follows:

$$\mathrm{MCE}'_k(p,d) = \mathrm{EXPAND}\big(\omega \mathrm{MCE}_{k+1}(p,d)\big), \tag{3.31}$$

$$\mathrm{GMC}'_k(p,d) = \mathrm{EXPAND}\big(\omega \mathrm{GMC}_{k+1}(p,d)\big), \tag{3.32}$$

$$\mathrm{TAD}'_k(p,d) = \mathrm{EXPAND}\big(\omega \mathrm{TAD}_{k+1}(p,d)\big), \tag{3.33}$$

After that, the difference between the expanded and reduction for the texture synthesis is determined and is expressed in equation (3.34), (3.35) and (3.36) as follows:

$$\mathrm{LM_k(p,d) = MCE_k(p,d) - MCE'_k(p,d),} \tag{3.34}$$

$$\mathrm{LG_k(p,d) = GMC_k(p,d) - GMC'_k(p,d),} \tag{3.35}$$

$$\mathrm{LT_k(p,d) = TAD_k(p,d) - TAD'_k(p,d),} \tag{3.36}$$

The pyramid cost volume differences are combined and reconstructed into final matching cost function $\mathrm{MC(p,d)}$ as described in equation (3.37) and (3.38) as follows:

$$\mathrm{LP_k(p,d) = (\sigma_{LM}LM_k(p,d) + \sigma_{LG}LG_k(p,d) + \sigma_{LT}LT_k(p,d)),} \tag{3.37}$$

$$\mathrm{MC(p,d) = \sum_{i=0}^{k} LP_{k+1}(p,d) + \omega LP_k(p,d),} \tag{3.38}$$

where σ_{LM}, σ_{LG} and σ_{LT} are the balancing parameters between cost volume differences in the pyramid as shown in Figure 3.3. A flowchart of the complete matching cost computation methodology consisting of the three components listed above is shown in Figure 3.4, 3.5, 3.6 and 3.7.

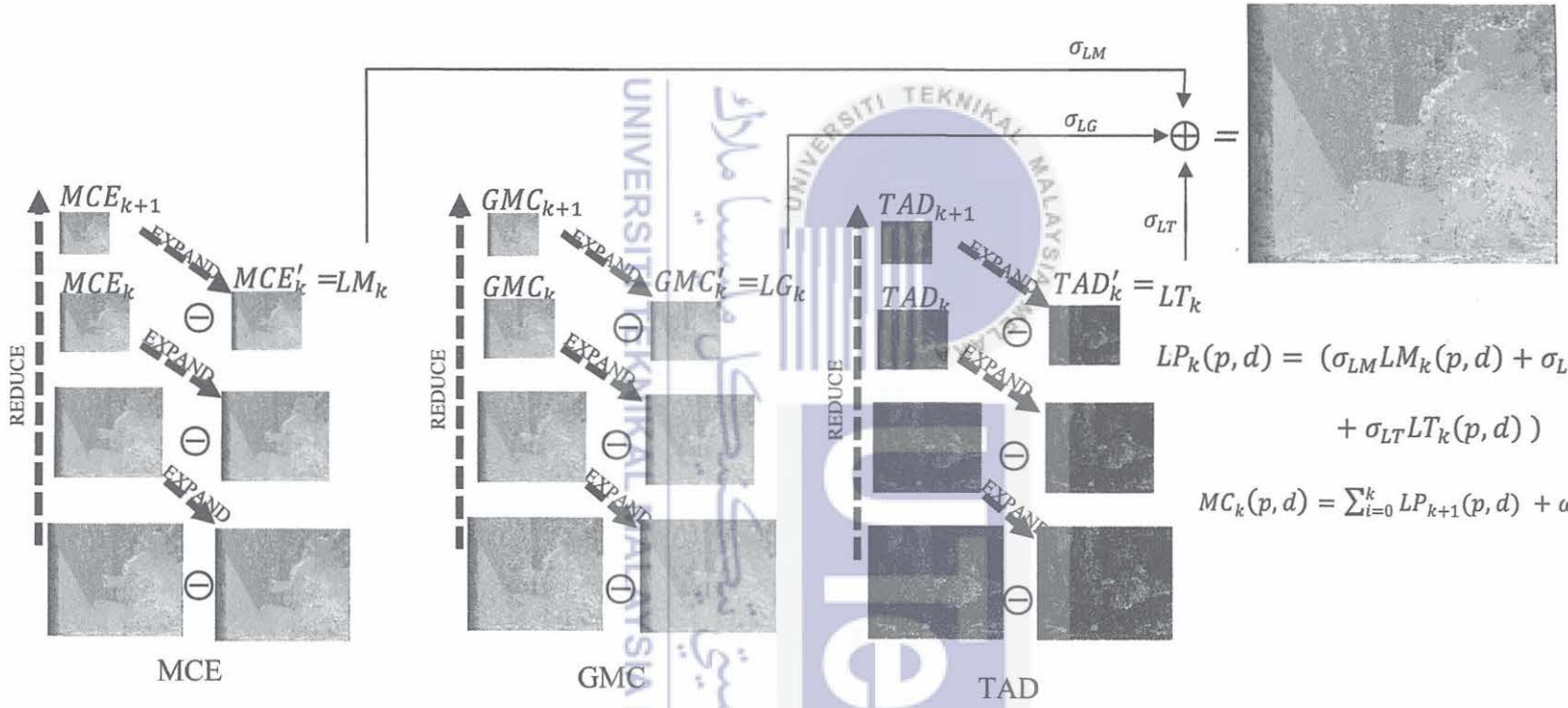


Figure 3.3: Proposed Planar Pyramid Fusion for Cost Combination

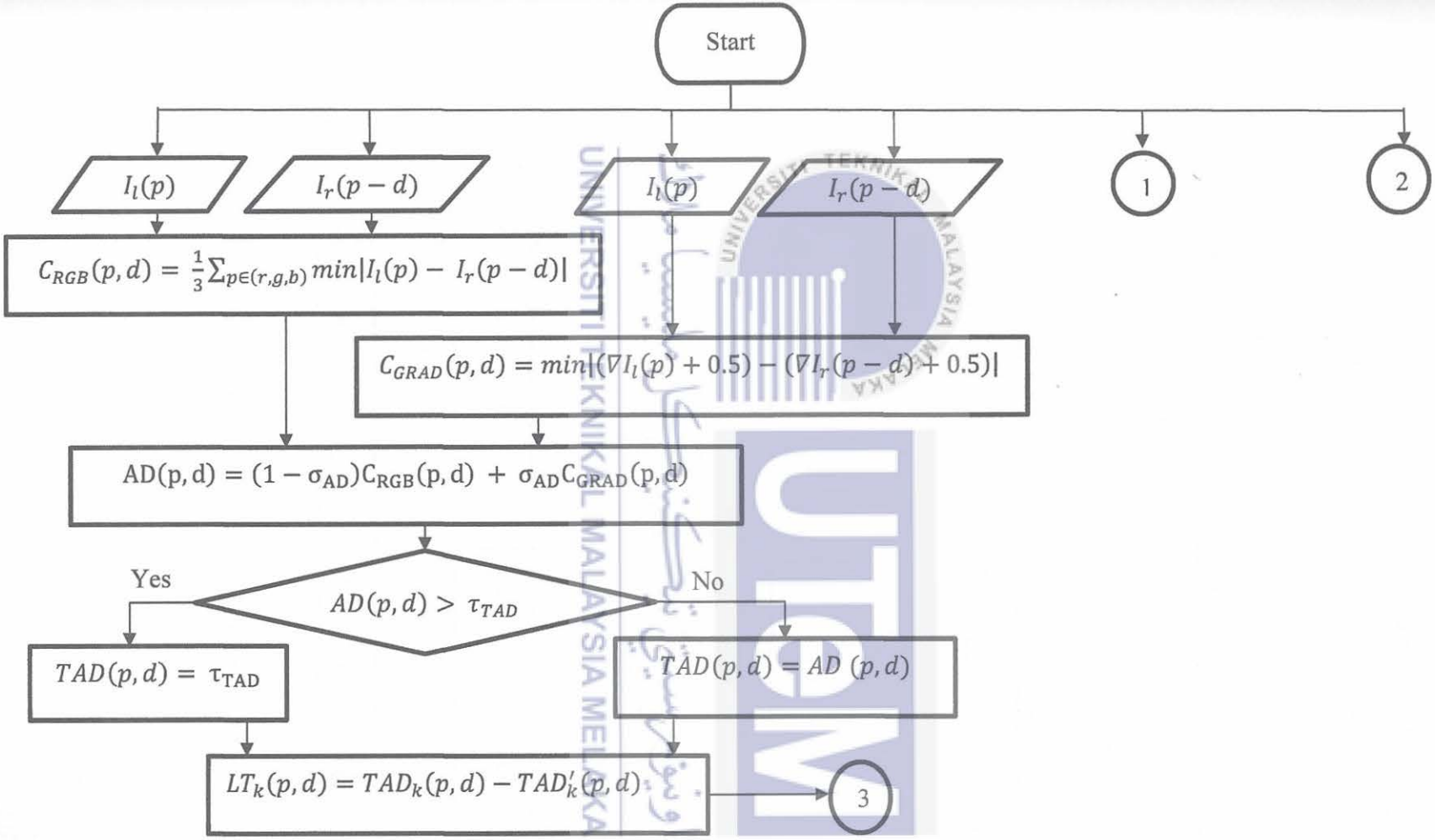


Figure 3.4: Matching Cost Computation Flowchart – TAD

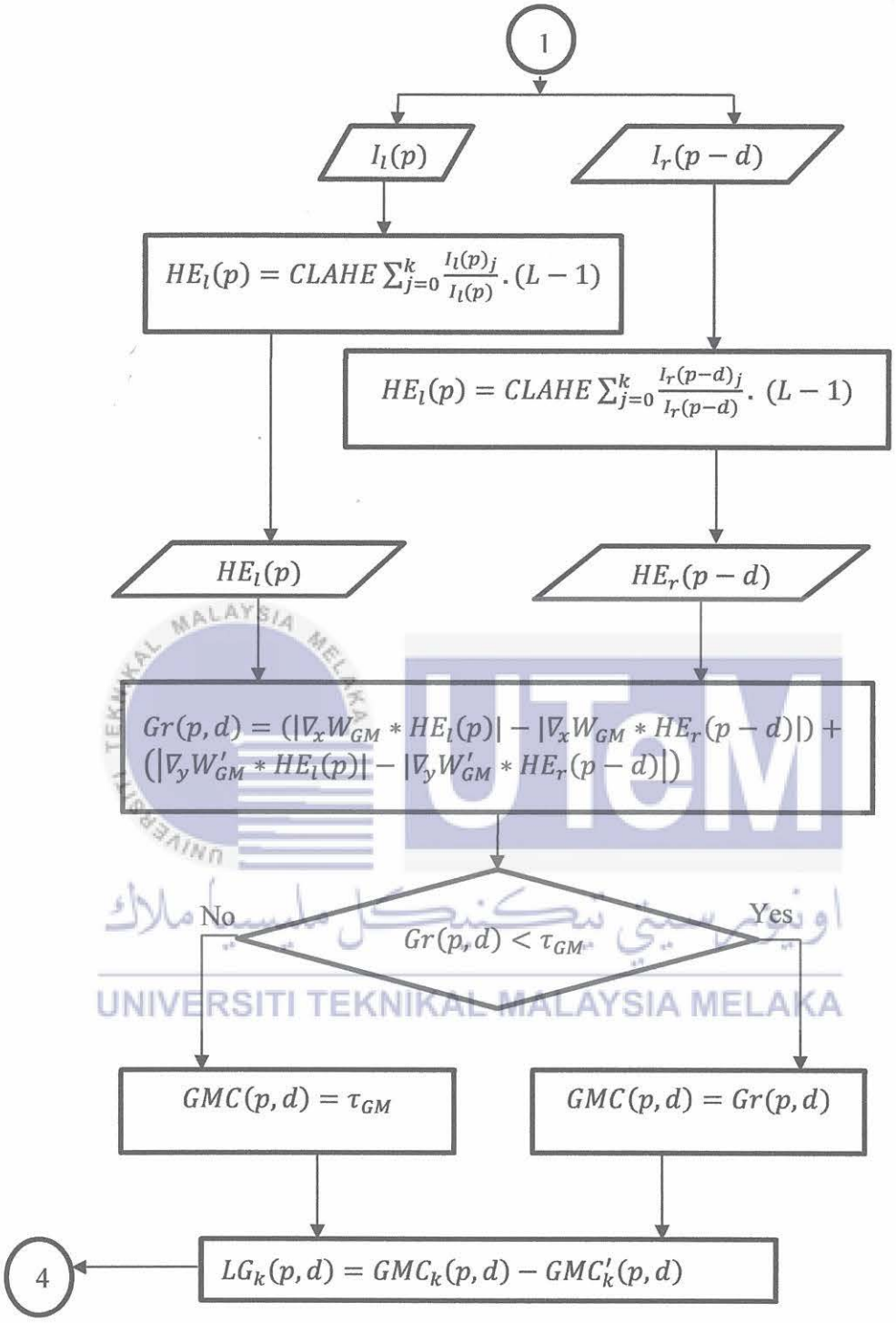


Figure 3.5: Matching Cost Computation Flowchart - GMC

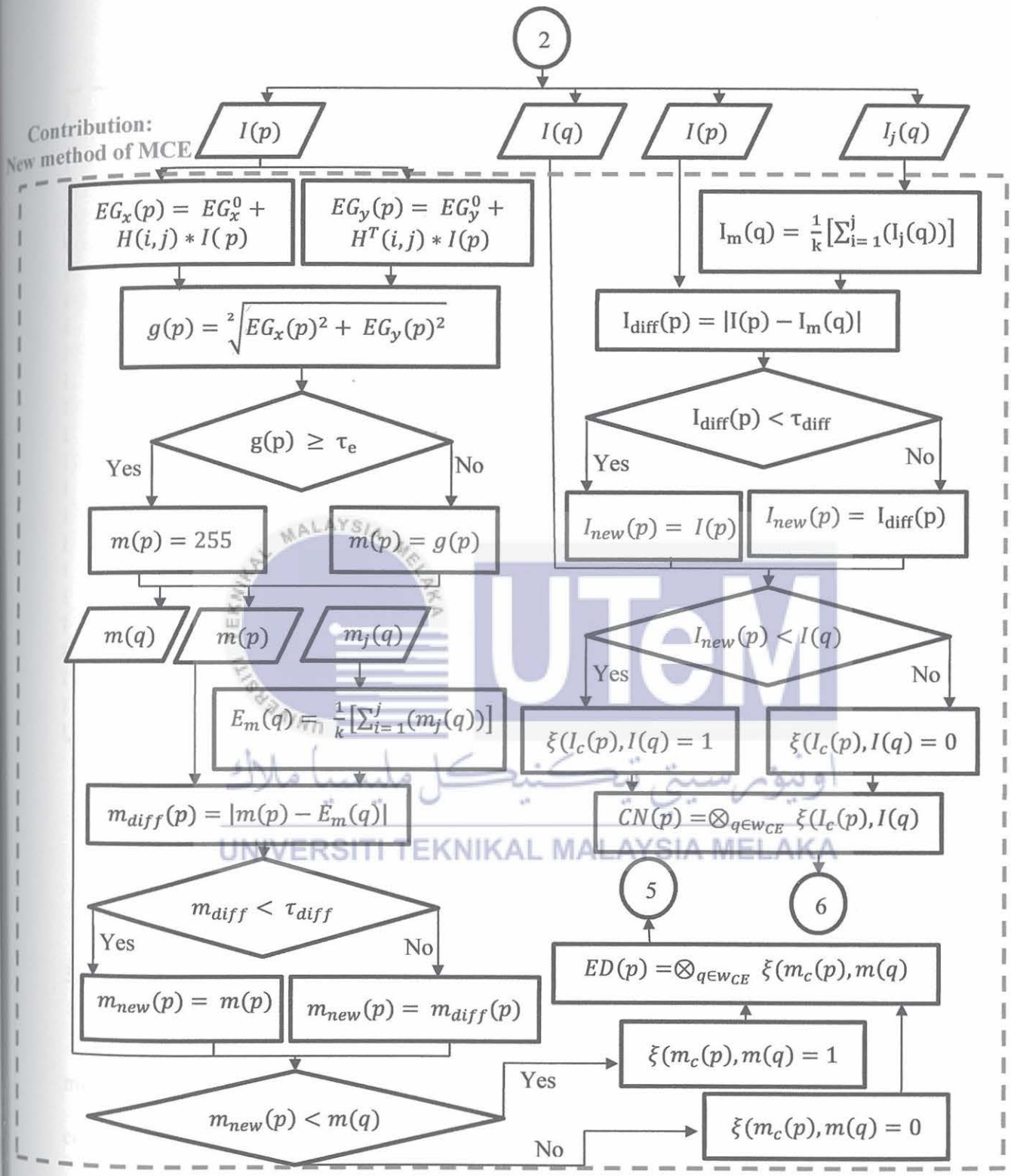


Figure 3.6: Matching Cost Computation Flowchart - MCE

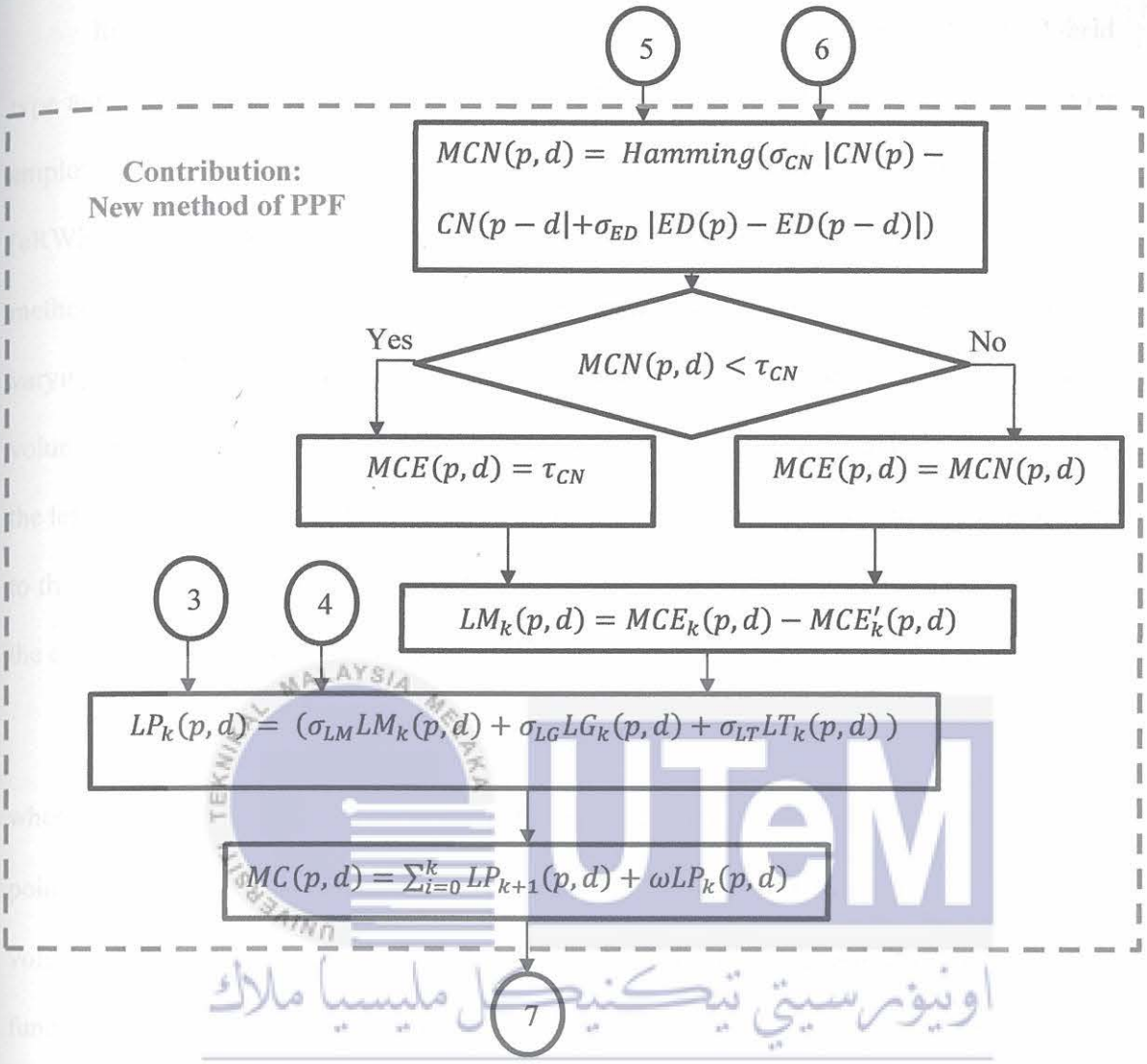


Figure 3.7: Matching Cost Computation Flowchart - PPF

3.4 Cost Aggregation Stage

The second stage of SMA is the cost aggregation which is the important stage to minimise the ambiguities from the matching process in the first stage; matching cost computation. This is considered to be the standard for local and SGM methods to determine the best aggregated cost for disparity range selection. This stage also reflects the accuracy of the disparity map in general.

In this research, the window-based cost aggregation is reconsidered, and a new hybrid type aggregation, the HRA is proposed. This HRA facilitates a hybrid type approach that employ the iterative Non-Local Guided Filter (iNLGF) and extended Random Walk Restart (eRWR) based on the studies by Lee et al. (2015) and Hamzah et al. (2017). The proposed method not only considers the occlusion and depth discontinuities, but also accounts for varying illumination and edge preserving. The iNLGF is using the non-local means cost volume from the matching cost based on neighbouring pixels as the guided imaging either the left or the right cost volume. The left cost volume is selected in this work as the guidance to the iNLGF process. Initially, the filtering starts with determining the non-local value of the cost volume employed by Wang et al. (2018),expressed in equation (3.39) as follows:

$$\mathrm{NL}(\mathrm{p},\mathrm{d})_n = \frac{1}{\mathrm{C(p)}}\int_{\Omega} \mathrm{MC_n}(\mathrm{p},\mathrm{d})\mathrm{f}(\mathrm{p},\mathrm{q})\mathrm{dq}, \tag{3.39}$$

where Ω denotes the area of the cost volume and $\mathrm{NL}(\mathrm{p},\mathrm{d})_n$ represents the filtered value at point p while n_{NL} states the number of iterations for iNLGF. $\mathrm{MC}(\mathrm{p},\mathrm{d})$ is the matching cost volume at point q and $\mathrm{C(p)}$ refers to the normalising factor. $\mathrm{f}(\mathrm{p},\mathrm{q})$ indicates the weighting function to determine how closely related the cost volume at the point p is to the cost volume at the point q. The $\mathrm{f}(\mathrm{p},\mathrm{q})$ of the non-local weighting function is given by equation (3.40) as follows:

$$\mathrm{f}(\mathrm{p},\mathrm{q}) = \mathrm{e}^{\frac{\left|\mathrm{B(q)_{q\in w_q}} - \mathrm{B(p)_{p\in w_p}}\right|^2}{(\mathrm{sd})^2}}, \tag{3.40}$$

where $\mathrm{B(q)}$ and $\mathrm{B(p)}$ are the local mean value of the cost volume surrounding q and p under support window $\mathrm{w_q}$ and $\mathrm{w_p}$ while sd refers to the filtering parameter of standard deviation. Then, the non-local mean values through the iNLGF filter kernel $\mathrm{G_{p,q}(NL_n)}$ as applied by He et al. (2013) and expressed by equation (3.41) as follows:

$$\mathrm{G_{p,q}(NL_n)} = \frac{1}{|\mathrm{w}|^2}\sum_{\mathrm{q}\in \mathrm{w_g}}\left(1 + \frac{(\mathrm{NL(p)}_{n-1} - \mu_{\mathrm{g},n-1})(\mathrm{NL(q)}_{n-1} - \mu_{\mathrm{g},n-1})}{\sigma^2_{\mathrm{g},n-1} + \varepsilon}\right), \tag{3.41}$$

where NL_n refers to the non-local mean cost volume at n-th iteration and p represents the coordinates pixel of interest (x, y). The size of support window, r x r is denoted as w_g and w represents the number of pixels in the support window, w_g. The NL(p) and NL(q) are the cost volumes from the non-local means with q and p representing the neighbouring pixel in the support window and the center pixel. The control element for the smoothness term is represented by the letter ε. The μ_g and σ_g indicate the guidance cost mean and variance of cost values which are given by equation (3.42) and (3.43) as follows:

$$\mu_g = \frac{1}{|w|}\sum_{q \in w_g} NL(q), \tag{3.42}$$

$$\sigma_g = \frac{1}{|w|}\sum_{q \in w_g} NL(q) - \mu_g. \tag{3.43}$$

The aggregation cost volume for the iNLGF at this process is expressed as in equation (3.44) and (3.45) as follows:

$$GF_n(p, d) = G_{p,q}(NL_n)MC_n(p, d), \tag{3.44}$$

$$PC(p, d) = GF_n(p, d), \tag{3.45}$$

where the $G_{p,q}(NL_n)$ is the weight of iNLGF and MC(p, d) refers to the cost from matching cost computation at n-th iteration from equation (3.38). The PC(p, d) can be referred as the pixel cost in the cost aggregation stage.

The matching cost is then aggregated using SLIC, graph segmentation, and eRWR to produce the segment cost as a hybrid combination with pixel cost in order to obtain the final cost volume in the cost aggregation based on studies by Lee et al. (2015) and Li et al. (2020). The SLIC algorithm is applied in the proposed model towards superpixel segmentation for both the left and the right images. Since an entire superpixel is matched to the target image rather than based on the pixel-wise matching, the local results are more resilient to noise variations. The SLIC cost function, FS(p) for both left and right images are expressed by equation (3.46) and (3.47) as follows:

$$FS_l(p) = \frac{1}{n_s}\sum_{p\in s} I(p)\, PC(p,d), \tag{3.46}$$

$$FS_r(p-d) = \frac{1}{n_s}\sum_{p-d\in s} I(p-d) PC(p,d), \tag{3.47}$$

where the n_s represents the number of pixels in the cluster of superpixel s. The SLIC cost function is computed individually for the left and the right images with the pixel cost. Then, the value of the SLIC cost function is used to reconstruct the graph propagated to surrounding nodes with a probability proportional to the edge weights, where the edge weights are determined by the similarity in intensity between surrounding superpixel which is given in equation (3.48), (3.49) and (3.50) as follows:

$$w_{ij} = (1-\tau_e)\exp\left(-\frac{\left(FS(p_i)-FS(p_j)\right)^2}{\sigma_e}\right)+\tau_e, \tag{3.48}$$

$$W = [w_{ij}]_{s\times s}, \tag{3.49}$$

$$\overline{W} = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}, \tag{3.50}$$

where w_{ij} is the graph segmentation and s refers to the number of superpixels which are used to form the weighted graph matrix W with diagonal equal to zero. $FS(p_i)$ and $FS(p_j)$ are the intensities of the i–th and j-th superpixels. σ_e and τ_e denote the weightage and truncation of the graph segmentation acted to control the shape of the graph. According to equation (3.48), superpixels with similar intensities are often more likely to produce a larger influence. The weighted graph matrix W in equation (3.49) is used to determine matrix $\overline{W}$ by normalising the rows of W at which row values of matrix W are added to construct the diagonal matrix D.

Subsequently, the SLIC cost function, weighted matrix and pixel cost are used in the eRWR which is updated until convergence is obtained iteratively. The smoothness constraint in the images prevent the RWR algorithm from often producing a good solution, however it does provide one local minima. Therefore, a modification is made to the standard update

process to adaptively update the matching costs based on the current determination of which superpixels are in the occlusion or the depth discontinuity areas. However, this has a negative impact on the performance because the smoothness term performed poorly at the area of the occlusion and the depth discontinuity since the smoothness assumes the disparities between the neighbouring pixels are similar. This research establishes a visibility texture algorithm within the eRWR that provides for an occluded pixel to have no matches on the target image and a non-occluded pixel to include a minimum of one match to solve for the occlusion issues. The visibility texture algorithm is given in equation (3.51) as follows:

$$VT_t(s) = \begin{cases} 1 \text{ if } |FS_l(p_s) - FS_r(p_s - d)| \leq 1 \\ 0 \text{ if } |FS_l(p_s) - FS_r(p_s - d)| > 1' \end{cases} \quad (3.51)$$

where $FS_l(p_s)$ and $FS_r(p_s - d)$ is represented by the left and right segments from equation (3.45) and (3.46) superpixels, respectively, and p_s indicates the (x, y) coordinates for centroids of the superpixel s. t indicates the number of iterations in the eRWR. The value must be consistent for the left and right check which indicates an occlusion is determined if the relation is not satisfied. The value SLIC is validated and is vectorised with pixel cost as given in equation (3.52) as follows:

$$VT_t(p, d) = PC(p, d) \odot [VT_t(s)]_{s\,x1}, \quad (3.52)$$

where $\odot$ refers to the element-wise product function and $PC(p, d)$ is the pixel cost from equation (3.44). An additional fidelity texture algorithm is utilised to prevent the depth boundaries blurred from the visibility texture. In the depth discontinuity regions, the smoothness condition often fails. When the disparity values between the foreground and background differ significantly, a depth discontinuity is typically established. This can be expressed in equation (3.53) as follows:

$$d_i' = \frac{\sum_{j \in N(i) \cup i} w_{ij} \bar{d}_j VT_t(p_j)}{\sum_{j \in N(i) \cup i} w_{ij} VT_t(p_j)}, \quad (3.53)$$

where w_{ij} is the calculated similarity between two neighbouring superpixels using equation (3.48) and p_j denotes the (x,y) coordinates of neighbouring superpixels of the i-th superpixels. $VT_t(p_j)$ is the result from the left-right consistency check of visibility texture algorithm while $\bar{d}_j$ is the optimal disparity range of the neighboring superpixel. Thus, the value of d_i' is used in the penalty function, $\Upsilon_t(d, d_i')$ as formulated in equation (3.54) and (3.55) as follows:

$$\Upsilon_t(d, d_i') = \begin{cases} \left(\frac{(d_i' - d)}{\sigma_\Upsilon}\right)^2, & \text{if } |d_i' - d| \leq \tau_\Upsilon, \\ \left(\frac{\tau_\Upsilon}{\sigma_\Upsilon}\right)^2, & \text{if } |d_i' - d| > \tau_\Upsilon, \end{cases} \quad (3.54)$$

$$SC_t(p, d) = [\Upsilon_t(d, d_i')]_{s\,x1}, \quad (3.55)$$

where d is the disparity value, τ_Υ represents the truncation parameters and σ_Υ denotes the scaling parameter. These parameters control the penalty function to preserve depth boundaries by conserving the intensity difference between neighboring superpixels and vary within τ_Υ to prevent depth discontinuity. $SC_t(p, d)$ is the final value from robust fidelity and visibility texture. The eRWR algorithm updates the matching cost iteratively based on the visibility texture, fidelity texture and pixel cost known as segment cost given in equation (3.56) as follows:

$$SC_{t+1}(p, d) = (1 - c)\bar{W}SC_t(p, d) + cPC_0(p, d), \quad (3.56)$$

where $SC_t(p, d)$ represents the visibility and fidelity texture updates. The current pixel cost $PC_0(p, d)$ is used to compute the visibility and fidelity terms. For the pixel costs to spread along graph W, the initial pixel costs are added to the pixel costs, which are proportional to the restart probability of c. In the second stage of the proposed SMA, the final aggregated cost is the sum of the costs of each pixel and each segment. This hybrid combination can be translated into equation (3.57) as follows:

$$CA(p,d) = PC(p,d) + \gamma SC_t(p,d), \tag{3.57}$$

where γ refers to the superpixel and segment cost weighting parameters to balance towards pixel cost. Figures 3.8 and 3.9 show the step-by-step cost aggregation that consists of the iNLGF, SLIC, graph segmentation and eRWR.

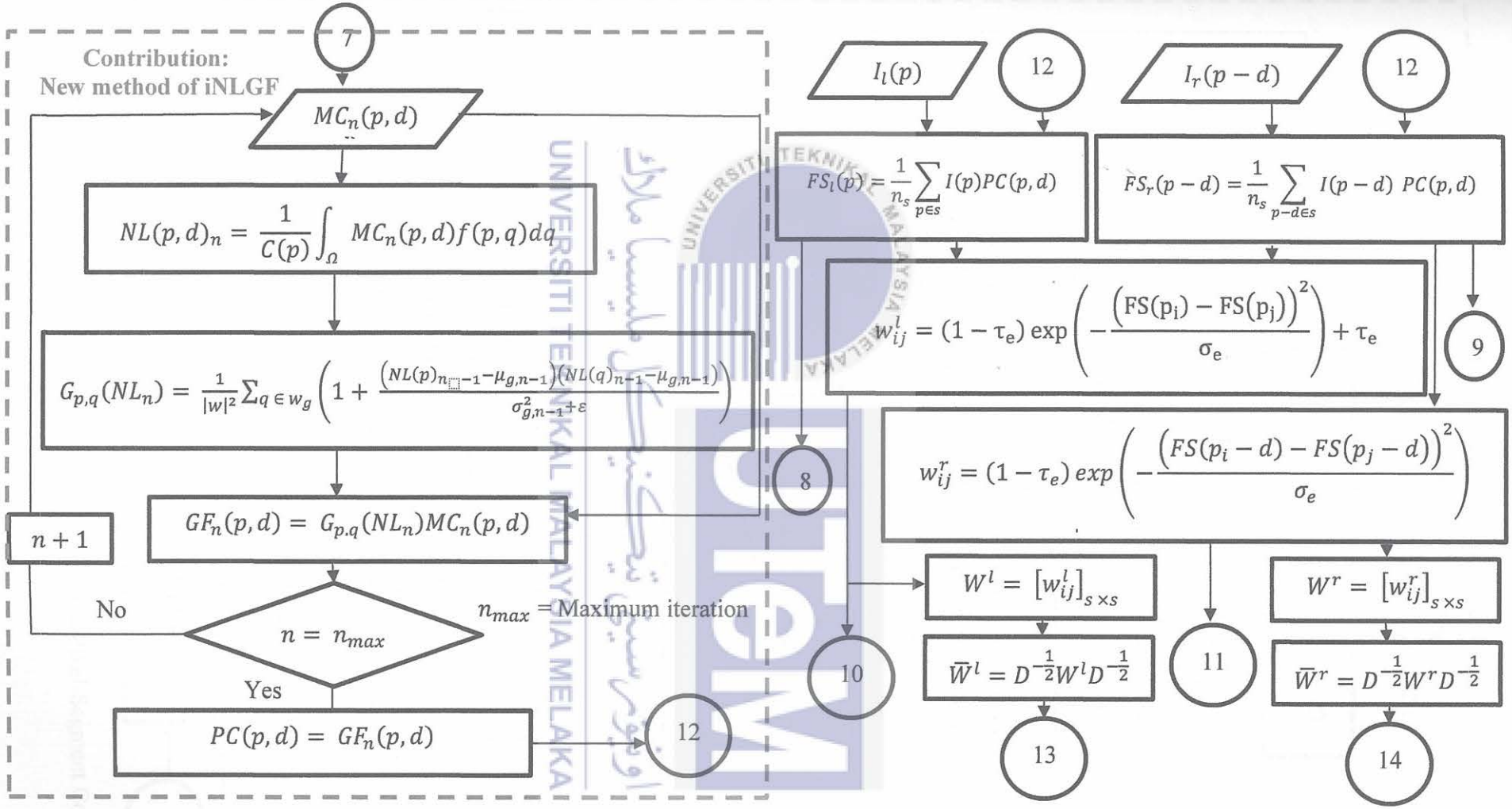


Figure 3.8: Cost Aggregation Flowchart – iNLGF, SLIC and Graph Segmentation

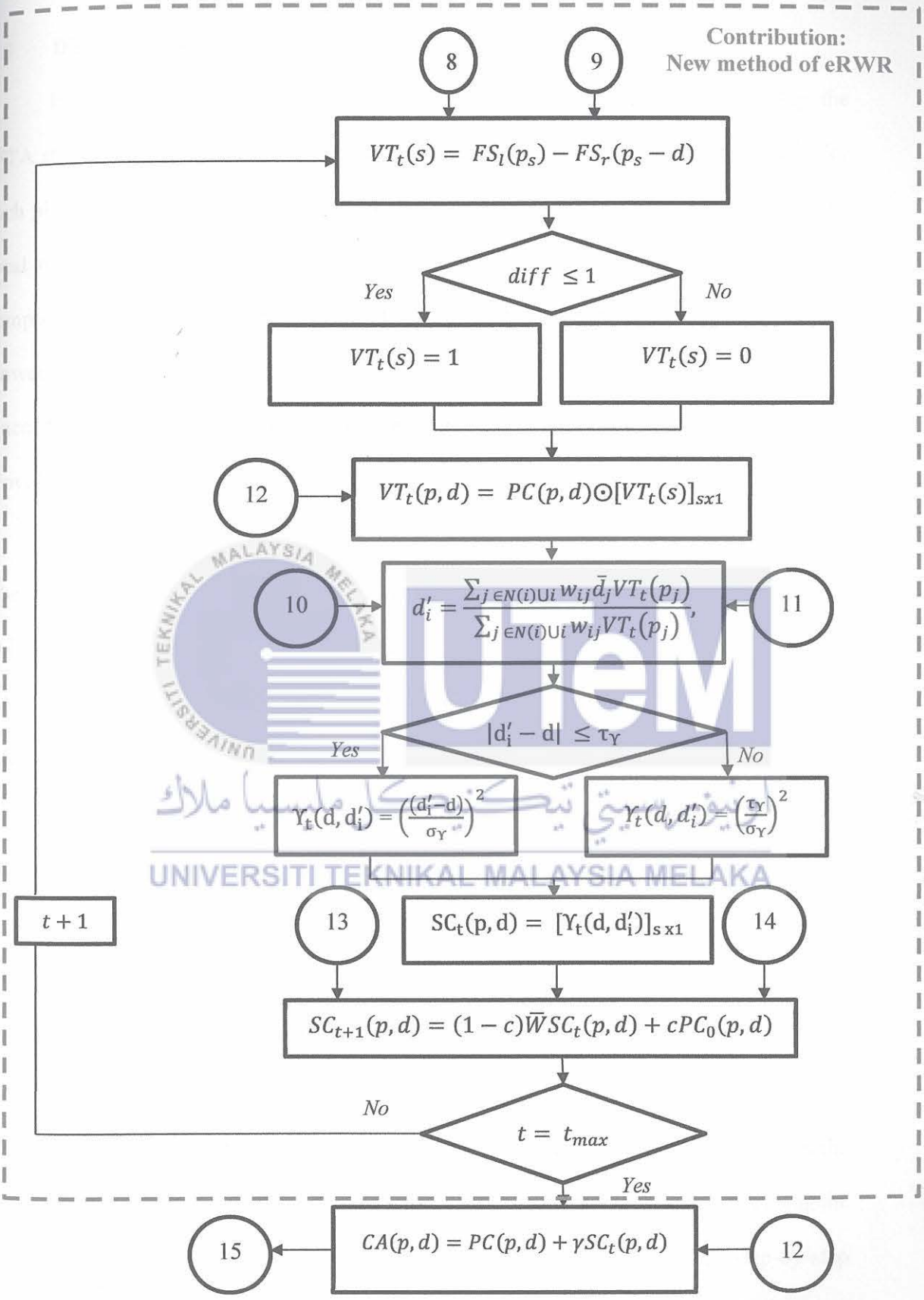


Figure 3.9: Cost Aggregation Flowchart – eRWR and Hybrid Pixel-Segment Cost

3.5 Disparity Selection Stage

In order to determine the final disparity and establish an accurate disparity map, the WTA strategy was employed to determine a minimum aggregated corresponding value for each pixel. This is one of the most common ways to determine the disparity values for the local method. According to Emlek et al. (2018) and Zhu et al. (2019) formulation, the computational complexity can be reduced by using the WTA technique for local algorithms. However, based on their findings, the disparity maps obtained up until this point still contain inaccuracies in the regions, some of which are occluded or have low texture. Equation (3.58) provides the WTA equation as follows:

$$\mathrm{DS}(p) = \arg\min_{d \in d_{r1}} \mathrm{CA}(p, d), \tag{3.58}$$

where the disparity produced by the minimum aggregated cost (d) at each pixel location (x, y) is selected. The range of allowed disparity values in an image is represented by the CA(p,d), which corresponds for the cost aggregation volume from equation (3.57) and d_{r1} refers to the value of disparity range in the cost volume for selection of the first lowest local minima. The lowest value from the disparity range d_{r1} is then selected using the minimal aggregated cost, also referred to as the first local minima. In the disparity refinement, these two local minima are used to compute the pixel confidence, which is employed in the invalid pixel filling process. Equation (3.59) presents the method for determining the second local minima from the WTA process, DO(p) .

$$\widehat{\mathrm{DS}}(p) = \arg\min_{d \in d_{r2}} \mathrm{CA}(p, d), \tag{3.59}$$

where $\widehat{\mathrm{DS}}(p)$ is the disparity value from the second lowest value from disparity range of the cost volume. The second lowest value from the disparity range d_{r2} is then selected using the minimal aggregated cost, also referred to as the second local minima. The step-by-step process in the disparity selection which calculates the minimum aggregated cost, and the two local minima is shown in Figure 3.10.

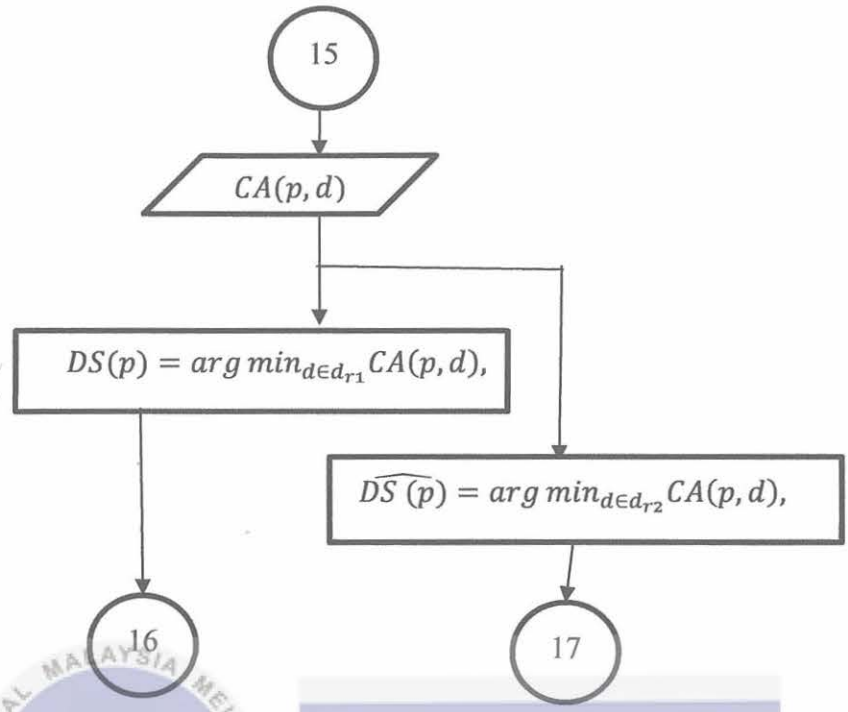


Figure 3.10: Disparity Selection Flowchart – Disparity Selection and Confidence

3.6 Disparity Refinement Stage

The algorithm concludes with a final stage of post-processing known as disparity refinement, which aims to further eliminate mismatches produced by occlusion, low texture, and other causes. In this research, the disparity refinement is performed using hierarchical cluster-edge refinement. The disparity refinement process begins by applying the left-right consistency check process, disparity confidence computation and invalid pixel fill-in based on median interpolation. Subsequently, an enhanced technique is used to determine the occlusion, noise and texture handling based on the K-means clustering along with the Side Window Filter (SWF) to preserve edges and boundaries.

The left-right consistency check is performed for each pixel (x,y) in the left disparity in order to determine the coordinates that correspond to this pixel on the right disparity. If the left disparity value in the current pixel is inconsistent with the right disparity, the map is

invalid (i.e., 0 = outlier) while the consistent pixel is considered valid (i.e., 1 = inlier) disparity locations based on studies by Emlek et al. (2018) and Sung et al. (2019). The location for disparity validation map, $d_0(p)$ at point p, which is the (x,y) coordinate as expressed by equation (3.60).

$$d_0(p) = \begin{cases} 0, & \text{if } |DS_{LR}(p) - DS_{RL}(p)| \leq \tau_{LR} \\ 1, & \text{otherwise} \end{cases}, \tag{3.60}$$

where $DS_{LR}(p)$ refers to the left disparity map and $DS_{RL}(p)$ denotes the right disparity maps. The τ_{LR} is the error threshold. The error threshold is set to 1.0 to minimise the error in the disparity map. In order to improve the performance, an additional stage to detect outlier pixels is introduced using the disparity confidence computation based on a study by Jachalsky et al. (2010). The confidence map computation equation is given in equation (3.61) as follows:

$$d_c(p) = \sum_{p \in d} \left| \frac{\widehat{DS}(p) - DS(p)}{\widehat{DS}(p)} \right|, \tag{3.61}$$

where $DS(p)$ and $\widehat{DS}(p)$ are the first and second local minima of the disparity range value based on equations (3.58) and (3.59) to compute the peak ratio for the confidence map. The invalid pixel fill-in is performed after the outlier is detected using the validation and confidence map. In this work, the invalid pixel is replaced by using median interpolation, as shown by equation (3.62).

$$u(p) = \begin{cases} \text{median}_{p \in w_p}\{|d(p)|\}, & \text{if } d_0(p) = 0 \text{ and } d_c(p) \leq \tau_{CF} \\ d(p), & \text{otherwise} \end{cases}, \tag{3.62}$$

where function median| | is the median interpolation within the median window, w_p which optimum selected using 3 x 3 size (Jachalsky et al., 2010). The $d_0(p)$ and $d_c(p)$ are the validation and confidence maps. $d(p)$ signifies a disparity value of coordinate p while τ_{CF} is the outlier threshold. If the pixel is valid, the original disparity value is copied and moved forward to the next pixel on the same scanning line. If the location has an invalid

pixel and a low confidence map, the algorithm performs a median search around the neighbourhood pixels within the median window. The invalid pixel is excluded and replaced with the final median value. Although a well-developed method, this filling-in and replacing process has the shortcomings of producing unwanted streak artefacts and failing to address the errorness disparity at the leftmost side of the disparity map. However, the approach in this research differs significantly from the approaches used by Da Silva Vieira et al. (2018) and Tatar et al. (2021) to remove the noise. A clustering method using K-means clustering is used in the disparity refinement to segment the interesting area of the disparity map from the background. The purpose of the K-means clustering is to minimise the objective function J, which determines the distance of data points to cluster centers of h. The Euclidean distance, d is calculated between the center with each pixel of the image is provided by equation (3.63) as follows:

$$d = \|I_l(p) - C_j\|, \tag{3.63}$$

where $I_l(p)$ is the left RGB stereo image and C_j represents the centroid of cluster j. Based on the distance d, each pixel is assigned to the nearest centroid, and the new location of the centroid, C_j is recalculated after all the pixels have been assigned. This process is repeated until it fulfils the tolerance or error value. The cluster is performed by the reshape of the image towards the objective function, J to reconstruct the final clustering. These stages are expressed in equation (3.64) and (3.65) as follows:

$$C_j = \frac{1}{h}\sum_{y \in C_j}\sum_{x \in C_j} I_l(p), \tag{3.64}$$

$$J_h = \sum_{j=1}^{h}\sum_{i=1}^{u}\left\|x_i^{(j)} - C_j\right\|^2, \tag{3.65}$$

where u refers to the number of clusters, h indicates the number of cases while $x_i^{(j)}$ denotes the data point case of image. The disparity values are then median filtered using a main condition that segregates them into clusters based on equation (3.65). Every pixel's median

value is computed, but only by using the data from the same colour cluster. The median value is computed after the data from different colour clusters have been excluded. This strategy works well for handling occlusions as given in equation (3.66).

$$d'(p) = \text{median}\{|u(p)_{w_h} \in J_h|\}, \tag{3.66}$$

where w_h denotes the windows size for median filtered of the disparity map.

The final stage of the disparity refinement in this work is performed by the SWF based on the study by Gong et al. (2018) to perform the texture smoothing and the edge preserving tasks. The output of the SWF has the minimum L2 distance to the input pixel as the final output. The SWF generated eight outputs, $B'^{\theta,\rho,r}_i$ by applying a filter kernel, F to the disparity map in each side window as shown in equations (3.67), (3.68), (3.69) and (3.70).

$$F\left(d'_{n_f}(p)_i, \theta, \rho, r\right) = \frac{1}{N_n}\sum_{j\in\omega_i^n} \omega_{ij} d'(p)_j, \tag{3.67}$$

$$N_n = \sum_{j\in\omega_i^n} \omega_{ij}, n \in S \tag{3.68}$$

$$B'^{\theta,\rho,r}_i = F\left(d'_{n_f}(p)_i, \theta, \rho, r\right), \tag{3.69}$$

$$\theta = m \times \frac{\pi}{2}, \tag{3.70}$$

where m refers to the side parameter of the filter kernel in the range of $m \in [0,3]$ that provides the angle between the window and the horizontal line labelled as θ contributes to position S = {L, R, U, D, NW, NE, SW, SE}. ρ denotes the position of the target pixel i at coordinates (x,y). r represents the radius of the window for SWF. ω_{ij} is the weight of pixel j, which is in the neighborhood of the target pixel i. By changing θ and fixing (x, y), the direction of the window can be changed while aligning the side with i.

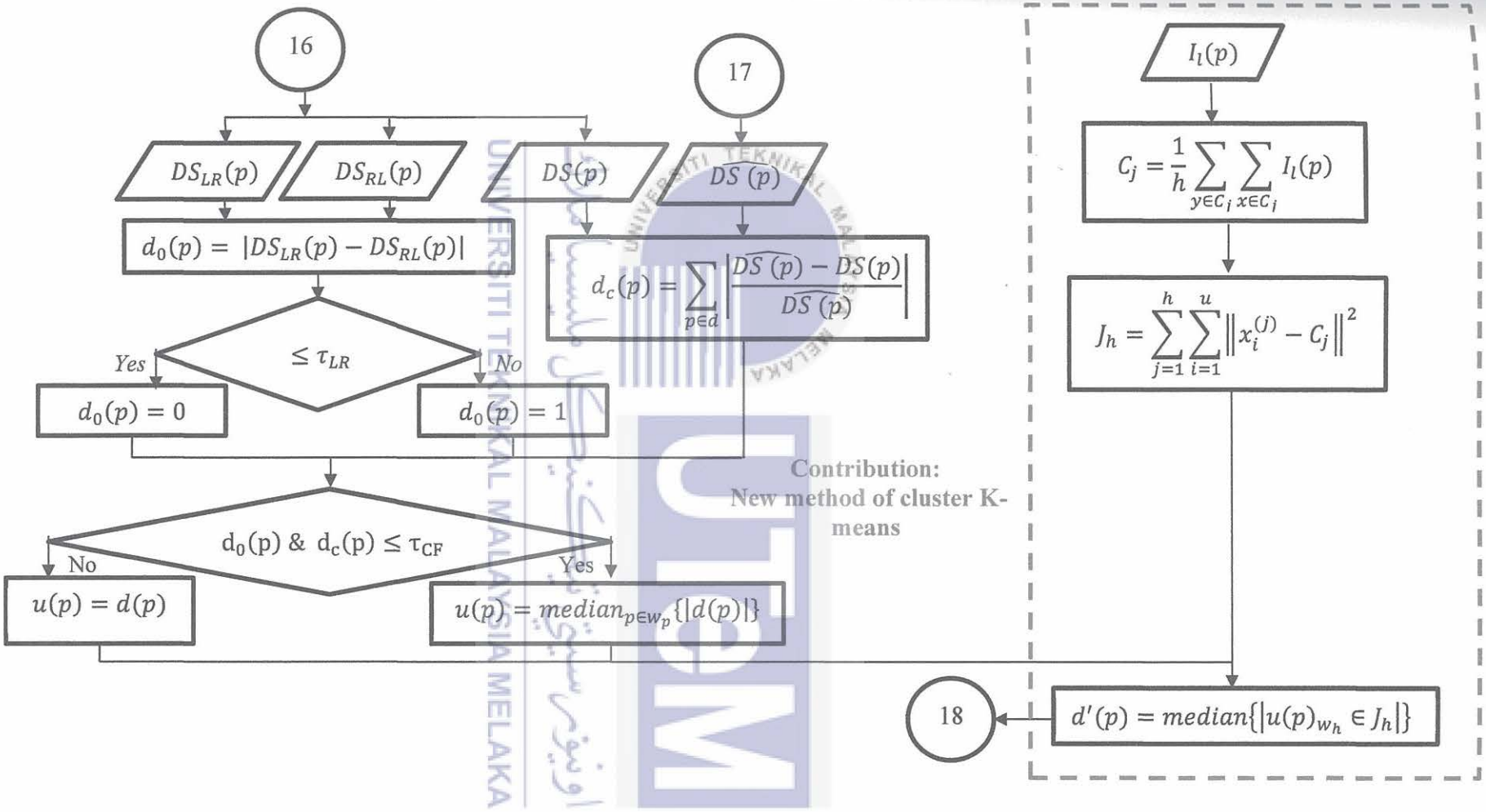


Figure 3.11: Disparity Refinement Flowchart – Left-Right Check, Disparity Confidence, Median Interpolation, and K-means

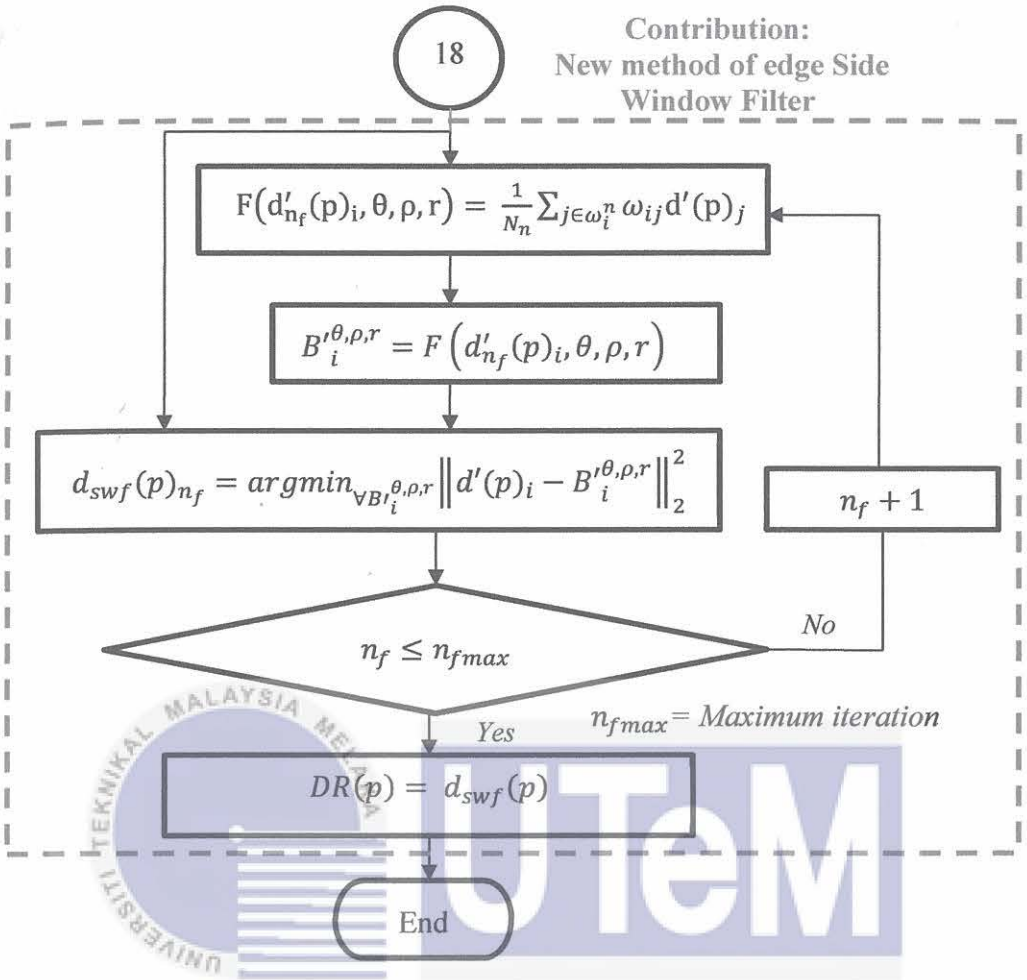


Figure 3.12: Disparity Refinement Flowchart – SWF

The final disparity value of SWF is determined by the output of the side window that has the minimum L2 distance to the input intensity as the final output given in equation (3.71) and (3.72).

$$d_{swf}(p)_{n_f} = argmin_{\forall B'^{\theta,\rho,r}_i} \left\| d'(p)_i - B'^{\theta,\rho,r}_i \right\|_2^2, \tag{3.71}$$

$$DR(p) = d_{swf}(p), \tag{3.72}$$

Figures 3.11 and 3.12 show the disparity refinement in stages that comprises of the left-right check consistency, disparity confidence computation and filling, k-means clustering and SWF.

3.7 3D Surface Reconstruction

In this research, the 3D surface reconstruction is performed as a use case for the proposed SMA. The reconstruction is based on the work developed by Fan et al. (2018) which was explained in Section 2.5. The 3D surface is reconstructed, whereas each 3-D point $p^W = [x^W, y^W, z^W]^T$ can be computed from its projections. The projections on the left image πl and the right image πr are $p_L = [u_l, v_l]^T$ and $p_r = [u_r, v_r]^T$ using the intrinsic and extrinsic parameters of the stereo system, where v_r is equivalent to v_l, and u_r is associated with u_l by d. The disparity is defined as $d = u_l - u_r$. The projection of a horizontal plane on the v-disparity map is a linear pattern as expressed in equation (3.73).

$$d = -\frac{T_c n_l}{\beta}(f\sin\theta - v_0\cos\theta) - v\frac{T_c n_l}{\beta}\cos\theta = \alpha_0 + \alpha_1 v, \tag{3.73}$$

where θ is the pitch angle between the stereo rig and the object. f is the focus length of the cameras, T_c is the baseline, and (u_0, v_0) is the principal point in pixels. When $\theta = \pi/2$, $d = -\frac{T_c n_l}{\beta}$ is a constant. Otherwise, d is proportional to v. Now, the PT can be straightforwardly realised using parameters $\alpha = [\alpha_0, \alpha_1]^T$. α can be estimated by solving a least squares problem with a set of reliable correspondences $Q_l = [p_{l1}, p_{l2}, \ldots, p_{lm}]^T$ and $Q_r = [p_{r1}, p_{r2}, \ldots, p_{rm}]^T$. The roll angle γ can be estimated by fitting a linear plane ($d(u,v) = \gamma_0 + \gamma_{1u} + \gamma_{2v}$) to a small patch from the near field in the disparity map and $\gamma = \arctan(-\gamma_1/\gamma_2)$. The pitch angle θ can be estimated by rearranging equation 3.73 as equation 3.74, where the parameters $[\alpha 0, \alpha 1]^T$ have been approximated. The yaw angle ψ is assumed to be 0.

$$\theta = \arctan\left(\frac{1}{f}\left(\frac{\alpha_0}{\alpha_1} + v_0\right)\right), \tag{3.74}$$

Each 3D point $[X_w, Y_w, Z_w]^T$ can be transformed into $[X'_W, Y'_w, Z'_w]^T$ using equation (3.75). The rotation matrix, $R = R_\psi R_\theta R_\gamma$ is a SO(3) matrix as expressed in equation (3.76), (3.77) and (3.78). The rotation with R greatly facilitates the detection of the texture and edges.

$$\begin{bmatrix} X'_w \\ Y'_w \\ Z'_w \end{bmatrix} = R_\psi R_\theta R_\gamma \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}, \tag{3.75}$$

where

$$R_\psi = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix}, \tag{3.76}$$

$$R_\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}, \tag{3.77}$$

$$R_\gamma = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.78}$$

Furthermore, the integration of the 3D surface reconstruction in this research provides the dual purposes of producing a projection of SMA accuracy and qualitative measurement for input stereo images without ground truth.

3.8 Measurement Set Up

The Middlebury Stereo and KITTI Stereo, two standard online benchmarking datasets are used in this research. This standard online benchmarking datasets have two images that are needed to establish the disparity maps: the left image and the right image. Furthermore, these online platforms provide data such as the ground truth for the disparity map and the efficacy of other researchers' methods. Therefore, this is ideal for conducting comparison for the stereo vision research investigation. Middlebury Stereo has 30 input images (15 training datasets and 15 testing datasets), which can simulate parameter settings, whereas KITTI Stereo has 400 input images (200 training datasets and 200 testing datasets), that are employed in autonomous vehicle navigation and simulate an actual road scenario.

A setup was developed to aid in the development and testing of this SMA using three different datasets: Middlebury, KITTI and real stereo image from the Universiti Teknikal

Malaysia Melaka Laboratory (UTeMLab-Stereo). The Middlebury dataset, which provided the indoor scenes for this research was developed by Scharstein et al. (2014) consisting of 15 training images and 15 test images. Each training and test datasets provided the stereo pair images and ground-truth disparities in each set with different resolutions, disparity level, height, width, image size, and camera calibration parameters. The parameters of an algorithm were established using the training images, which were submitted repeatedly onto the Middlebury online platform to achieve the quantitative results. Only the final evaluation was performed using the test images. The Middlebury dataset contains 24 mobile stereo datasets created by Guanghan Pan et al. (2021), which were used in this research to obtain the qualitative results. In order to determine the ground-truth disparities for this dataset, an Apple iPod touch 6G placed on a UR5 robot arm was used in this research based on the portion subset of the structured lighting pipeline developed by Scharstein et al. (2014).

All the input images in the Middlebury dataset had distinctive characteristics and resolutions. The characteristics and resolutions of the final results were influenced by the disparity maps that were produced. These characteristics depended on the input images, such as the Jadeplant, Motorcycle, Playroom, Piano, and Shelves, which were utilised for complex scene artifacts. Meanwhile, the Adirondack, Recycle, and Vintage images were used to distinguish foreground objects from background objects. The images of the Motorcycle, Pipes, and Teddy were deployed to challenge the regions of depth discontinuity. The MotorcycleE and PianoL were used to evaluate the performance in terms of variation in illumination and radiometric differences.

The outdoor stereo scenes were established by the KITTI dataset and were obtained from the actual vehicle navigation data. The KITTI Stereo dataset was developed by Geiger et al. (2020) under the collaboration with the Karlsruhe Institute of Technology (KIT) and the Toyota Technological Institute. The dataset consists of 200 training images and 200

testing images with dynamic scenes of vehicle. The training images were utilised to determine the parameters and error rate of the SMA, while the testing images were used for the final evaluation and were uploaded onto the KITTI Vision Benchmark online platform for quantitative and qualitative results. The usage of the KITTI dataset enabled the testing of the algorithm's adaptability with more complex real-world stereo images in which the stereo sceneries were subjected to unpredictable lighting conditions attributed to the presence of the sun's natural light, vehicles, and tree shading. Additionally, the test also include dataset containing vast regions with low textures, including the walls, roads, repetitive patterns, and sky.

Real stereo images based on UTeMLab-Stereo indoor images were applied to evaluate the efficiency performance of the proposed SMA. The stereo sensor was used to capture six distinct scenes, each with a different layout and stereo challenges which were Kotak, Kotak2, Cube, Cube5, Cube22 and Cube33. The Bumblebee BB2 stereo vision camera was used to acquire all the images that were being displayed in the UTeMLab-Stereo; these images were not modified in any manner and did not contain any type of image enhancement. The parameters for the UTeMLab-Stereo images are equipped with resolution of 640 x 840, disparity range of 65, baseline of 12 mm and focal length of 6 mm. The Kotak and Kotak2 exhibited images taken with a stereo camera from two different distances. The boxes were organised to reflect the texture in the images and these images also had contrasting illumination in regions created by various lighting ambient noises. In this research, the images that were captured in the UTeMLab-Stereo were only allowed to be used for the purpose of providing a qualitative evaluation of real stereo images.

The total number of images used in this work, including those from Middlebury, KITTI, and UTeMLab-Stereo, is 460, which is sufficient to formulate and evaluate the performance of SMA due to the fact that evaluating stereo matching algorithms depends on

various factors, including the complexity of the algorithms being evaluated, the diversity and quantity of the datasets, and the specific goals of the evaluation. These datasets provide a diverse set of stereo vision datasets that include variations in lighting conditions, scene complexity, object types, occlusions, textures, etc. The datasets also ensure that the evaluation results are statistically significant, which shows that an algorithm might perform well on a small subset of data but might struggle with more complex or diverse scenes. Then, a reliable ground truth (accurate depth maps) is also provided by the datasets, which is essential for evaluating stereo matching algorithms. A good stereo-matching algorithm should generalise well to unseen data. Therefore, it's important to evaluate the algorithm's performance on datasets that it was not specifically trained on.

The parameters for this work were selected and optimised to cover the breadth of variations in the algorithm stage by stage to acquire the best accuracy. However, in many instances, these parameters took on different values. The parameter is optimised at every stage, where each parameter is executed using an incremental test from the minimum increase to the maximum value until it produces the lowest overall disparity map accuracy (bad pixel percentage error for all error). All the constant parameters were originally planned to have a minimum value and were gradually increased until the average error output attained a minimum value. Consequently, this method contributed to the efficient and maximum usage of the selected parameters. The selection of the parameters began with matching cost computation and continued one at a time until the disparity refinement stage. All the experiments were executed using the MATLAB and C++ programming languages and the hardware of a personal computer with a Central Processing Unit (CPU) of an Intel Xeon E5-2650, 2.6 GHz, and 64 GB of RAM.

3.9 Summary

This section has outlined the theoretical foundations for the methods used to design a new local stereo matching algorithm for disparity maps, as shown in Figure 3.2 and 3.3. Basically, the new SMA comprises of four stages: Stage 1 is the matching cost computation to determine the raw cost volume from the multi-cost per-pixel adjustment and fusion based on pyramid combination. Stage 2 is the cost aggregation, which is the advanced approach to reduce noise and to preserve the edge using the hybrid random aggregation between iNLGF, SLIC, graph segmentation and eRWR. Stage 3 involves employing a WTA strategy to select the minimum disparity value for the disparity refinement and confidence. Finally, Stage 4 is the final refinement applied to the disparity map values based on left-right check consistency, disparity confidence computation, median interpolation for invalid pixel filling, K-means clustering, and SWF to handle the occlusion, the boundaries and the low texture regions.

CHAPTER 4

RESULT AND DISCUSSION

This chapter presents the results of the experiment carried out to evaluate the performance of the proposed SMA algorithm. Additionally, this chapter aims to discuss the analysis and the accuracy assessment of the new SMA algorithm. First, Section 4.1 describes the type of the evaluation performed on the algorithm, which can be categorised into two types: quantitative and qualitative. Furthermore, section 4.2 presents the thorough manipulation of the SMA's algorithm, parameters and variables of SMA to obtain the optimal value. This is followed by Section 4.3 that discusses the performance of the SMA at each stage of the taxonomy; the Middlebury dataset, the KITTI dataset, the use of real stereo images and the 3D reconstruction. In the last section, Section 4.4, the comparison analysis based on the constraints is established using the Middlebury dataset. Among the five main components being analysed in this scenario include low texture region, repetitive patterns, depth discontinuity, radiometric differences, and occlusion.

4.1 Evaluation: Quantitative and Qualitative

In this research, there were two types of evaluations performed: quantitative and qualitative evaluation. The Middlebury and KITTI dataset offered a thorough quantitative assessment of the SMA algorithm using an online benchmark platform, enabling researchers to apply the SMA to generate the comparison results in the rank tables. The quantitative evaluation of the accuracy performance for each image was based on two attributes of bad pixel percentage, which were *nonocc* (i.e., the percentages of bad pixels among all pixels in

non-occluded regions) and *all* (i.e., the percentages of bad pixels among all pixels in all areas).

The lower bad pixel percentage for the stereo matching algorithm showed better disparity map accuracy. The quantified matching was obtained by comparing the experimental results with the ground truth provided by Middlebury and KITTI Platform, thus enabling the algorithm's accuracy to be evaluated objectively. Additionally, the Middlebury and KITTI provided the qualitative evaluation that was demonstrated in the comparison of the various stereo correspondence constraints. However, the Middlebury mobile dataset and the UTeMLab-Stereo images were only able to be evaluated qualitatively since there were no ground true images available as a reference. The elaboration about qualitative results is explained in the thesis by results from parameter tests. The qualitative result is based on disparity map observation, which focuses on invalid pixels (dark blue or dark brown) produced in the disparity map that have the characteristics of stereo correspondence constraints. Actually, to determine the optimised parameters, they are based on the quantitative result with the lowest bad pixel error (all error). The qualitative result is used only as a reference for these parameters, as it also produces the relationship between the number of invalid pixels and the pixel error percentage.

4.2 Parameters Optimisation

The first step in gaining the optimal parameters for the equations in this research was to determine the variables and parameters that were used in the equations. Since these parameters reflected the final accuracy of the SMA, it was essential to identify the parameters that were able to produce the best outcomes. The parameters were selected to cover the breadth of variations in algorithm stage by stage to acquire the best accuracy. However, in many instances these parameters took different values. The typical parameters used in this

research were determined from the experiment of Middlebury training dataset based on an average error of *nonocc* and *all* errors. All the constant parameters were originally planned to a minimum value and were gradually increased until the average error output attained a minimum value. Consequently, this method contributed to the efficient maximise usage of the selected parameters. The selection of the parameters began with matching cost computation and continued one at a time until disparity refinement stage. The following is a detailed description of the parameter settings employed in this research.

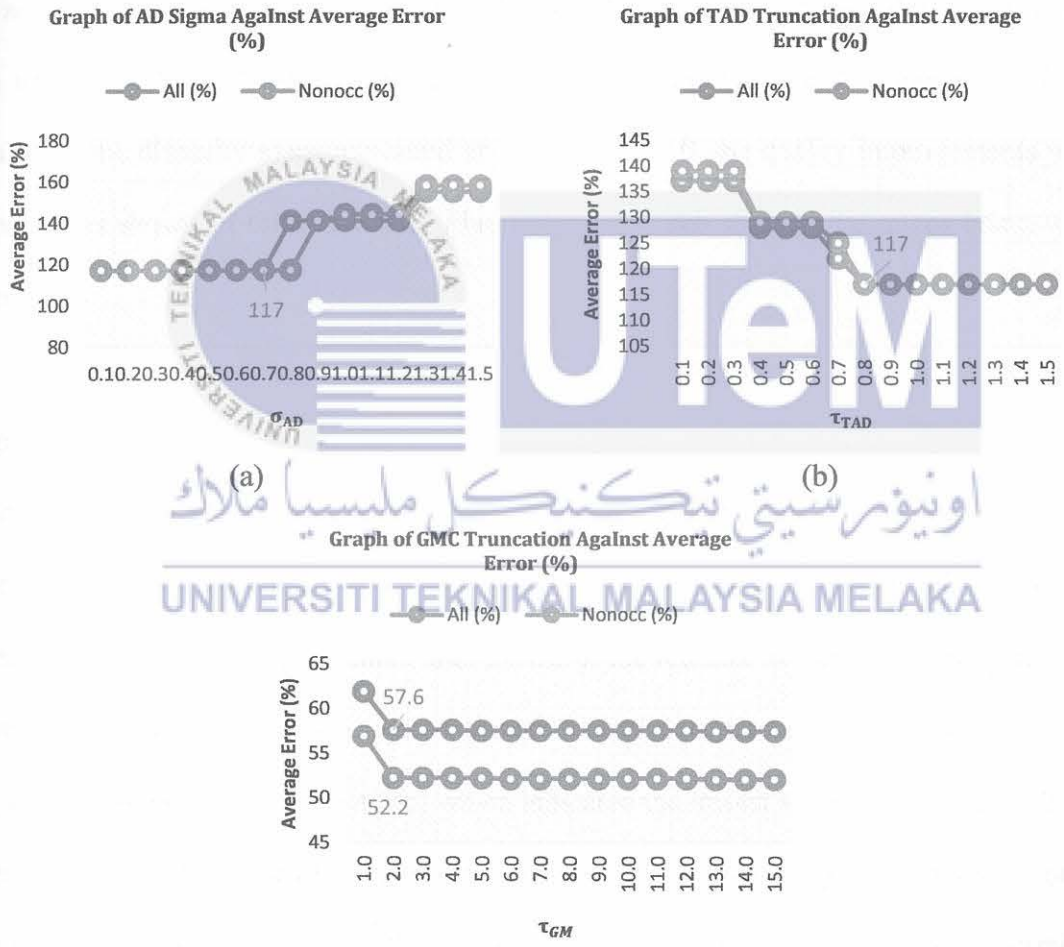
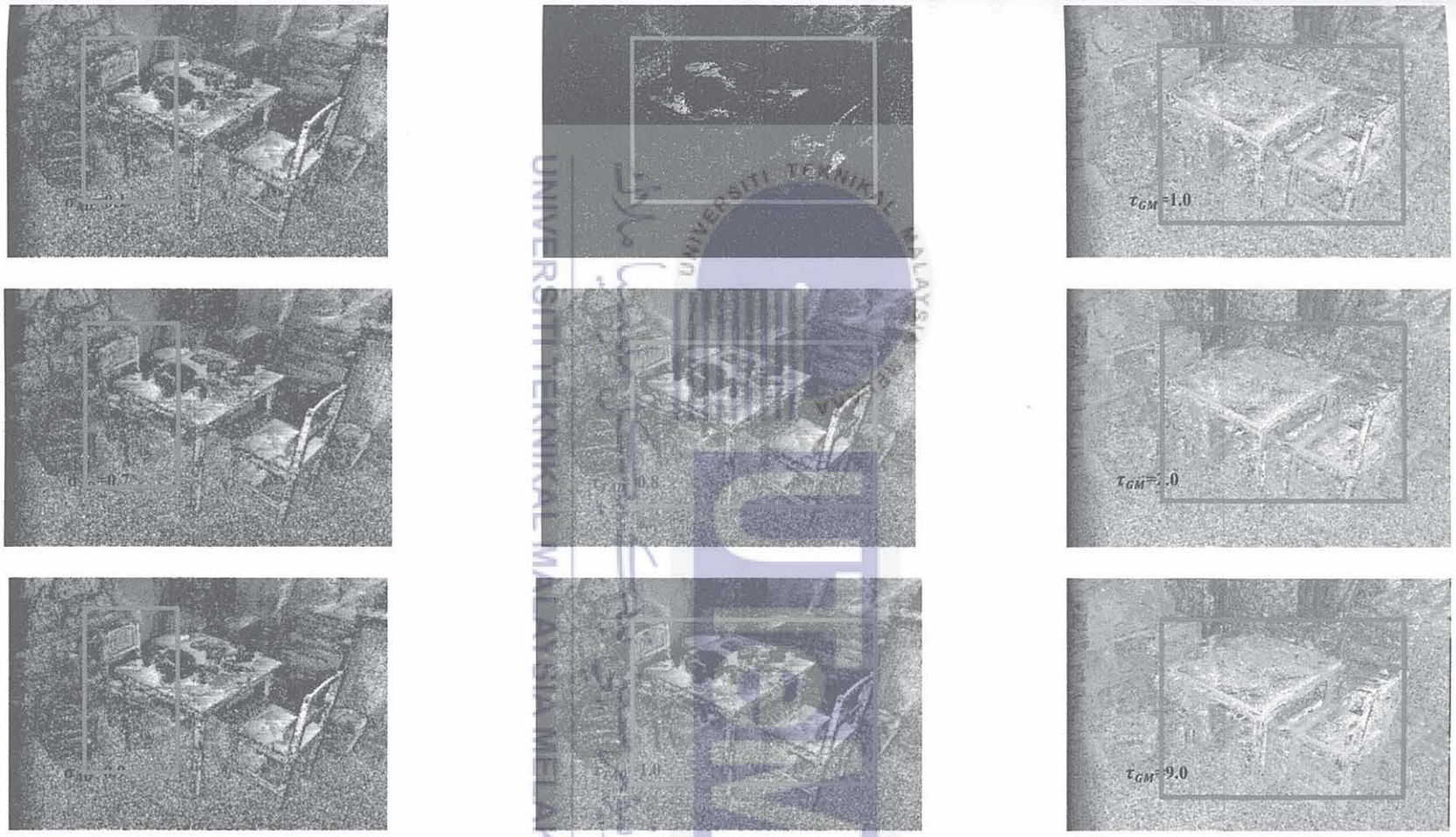


Figure 4.1: Line Graphs of Parameter Selection for TAD and GMC (a) σ_{AD} (b) τ_{TAD} (c) τ_{GM}

Stage 1 – Matching cost: 13 parameters were selected to determine the optimum cost volume as shown in Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, and Figure 4.6. At this stage, only matching cost computation stage and disparity selection stage were executed. The first parameter selected under the TAD cost computation which was the σ_{AD} as shown in Figure 4.1(a). There was a constant minimum value of 117% for *all* error and *nonocc* error from the value of 0.1 to 0.7. The parameter selected was σ_{AD}= 0.7. According to the graph of Figure 4.1(a), when σ_{AD} was increased uniformly in sequence, the average *all* and *nonocc* error remained constant and gradually increased in non-uniform sequence at the value of 0.8, while the differences between each interval steadily increased. When analysing the disparity maps measured at 0.1, 0.7 and 1.0, the quality improvements were apparent as shown in the red box in Figure 4.2 (a) which showed the edges being well-preserved.

The second parameter determined in TAD cost was the τ_{TAD} at 0.8 as shown in Figure 4.1(b) which shows a constant minimum value of 117% for *all* error and *nonocc* error from value 0.8 to 1.5. Based on the graph, the avg *all* and *nonocc* error reduced dramatically from 0.1 to 0.8 when τ_{TAD} increased in sequence and the minimum value remained constant after that. The quality improvement was shown in the red box in Figure 4.2 (b) where the texture and boundary were also well-preserved. The third parameter used was the τ_{GM} under the GMC cost shown in Figure 4.1(c) which indicated the lowest value of 52.0% for *all* and *nonocc*_errors at the value of 2.0 until 15.0. The results showed a significantly reduction in the accuracy value from 1.0 to 2.0 and remained constant after that. Next, the parameter selected was τ_{GM} = 2.0. Figure 4.2(c) shows the horizontal streaks in the red box were reduced after the truncation was set to 2.0 and above.



(a) (b) (c)

Figure 4.2: Qualitative Evaluation of Middlebury Playtable Parameter Selection (a) σ_{AD} (b) τ_{TAD} (c) τ_{GM}

There were 5 parameters determined under the MCE matching cost as shown in Figure 4.3. The w_{CE} was first selected and the window size was determined at 5 x 5 which contributed to the minimum value of an average *all* error of 45.7% and *nonocc* error of 38.9%. The line graph in Figure 4.3(a) depicts the variation in *all* error and *nonocc* error as the value of the window size was changed by the w_{CE}, indicating that the accuracy level was influenced by the window size dimension. The red box in Figure 4.4(a) shows the results indicating the significantly smoother texture and edges of the window size 5 x 5, which eliminated the salt and pepper noise compared to the window sizes of 1 x 5 and 3 x 3. The second parameter selected in the MCE was the τ_{CN} as shown in Figure 4.3(b). The constant minimum average error for *all* error and *nonocc* error were at 43.2% and 35.9%. Figure 4.3(b) shows the graph starting at τ_{CN} = 1.0 and increasing until 15.0. The results showed a rapid increase in accuracy until the peak value was reached and remained at τ_{CN} = 7.0 , thus, was selected as the parameter value. Figure 4.4(b) displays the disparity map for the Motorcycle with τ_{CN}=1.0, τ_{CN}=7.0 and τ_{CN}= 10.0. The results showed distorted edges and blurry edges shown clearly in the red box for τ_{CN}=7.0 and τ_{CN}= 10.0 compared with τ_{CN}=1.0, which texture was sharper and more apparent.

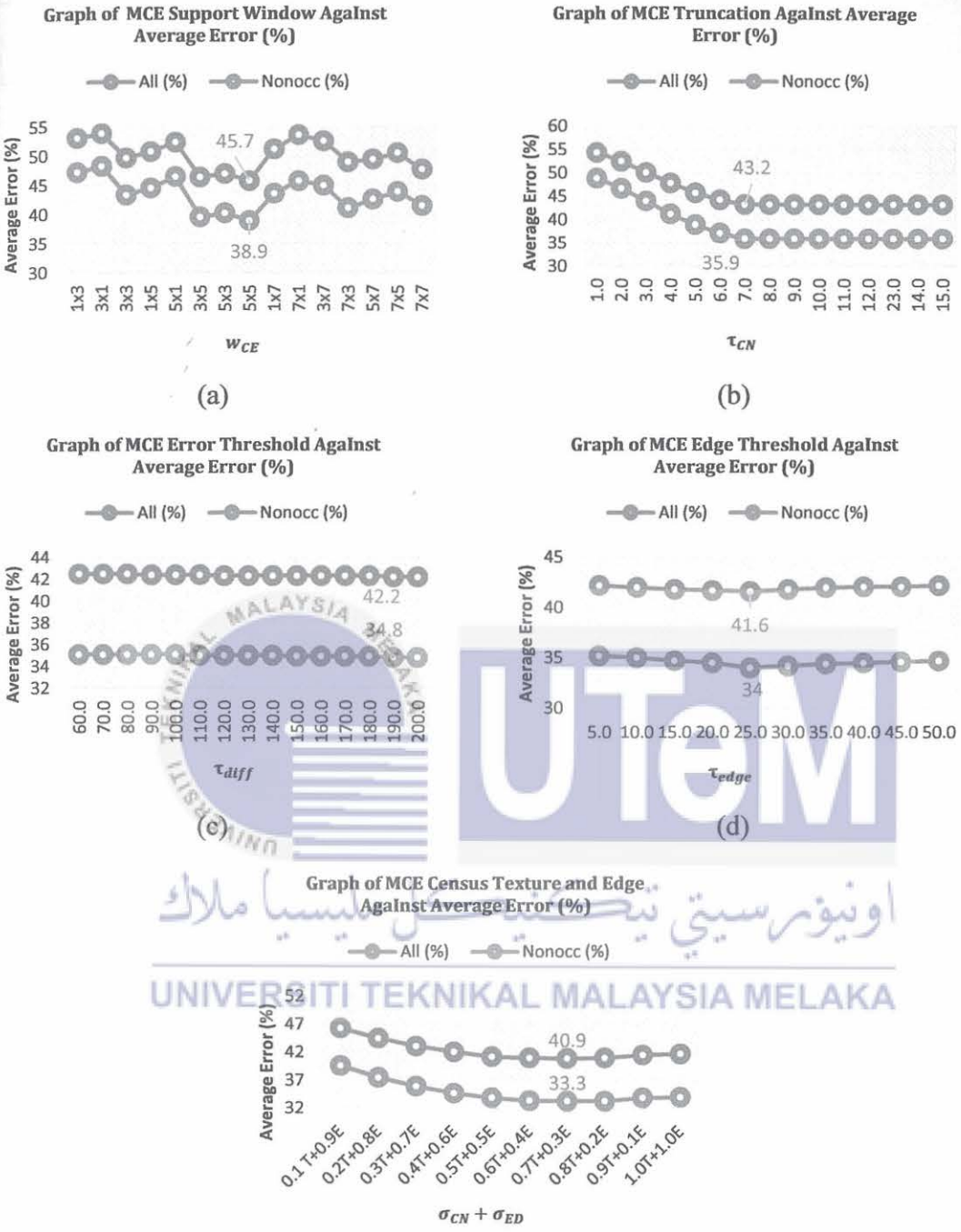


Figure 4.3: Line Graphs of Parameter Selection for MCE (a) w_{CE} (b) τ_{CN} (c) τ_{diff} (d) τ_{edge} (e) $\sigma_{CN} + \sigma_{ED}$

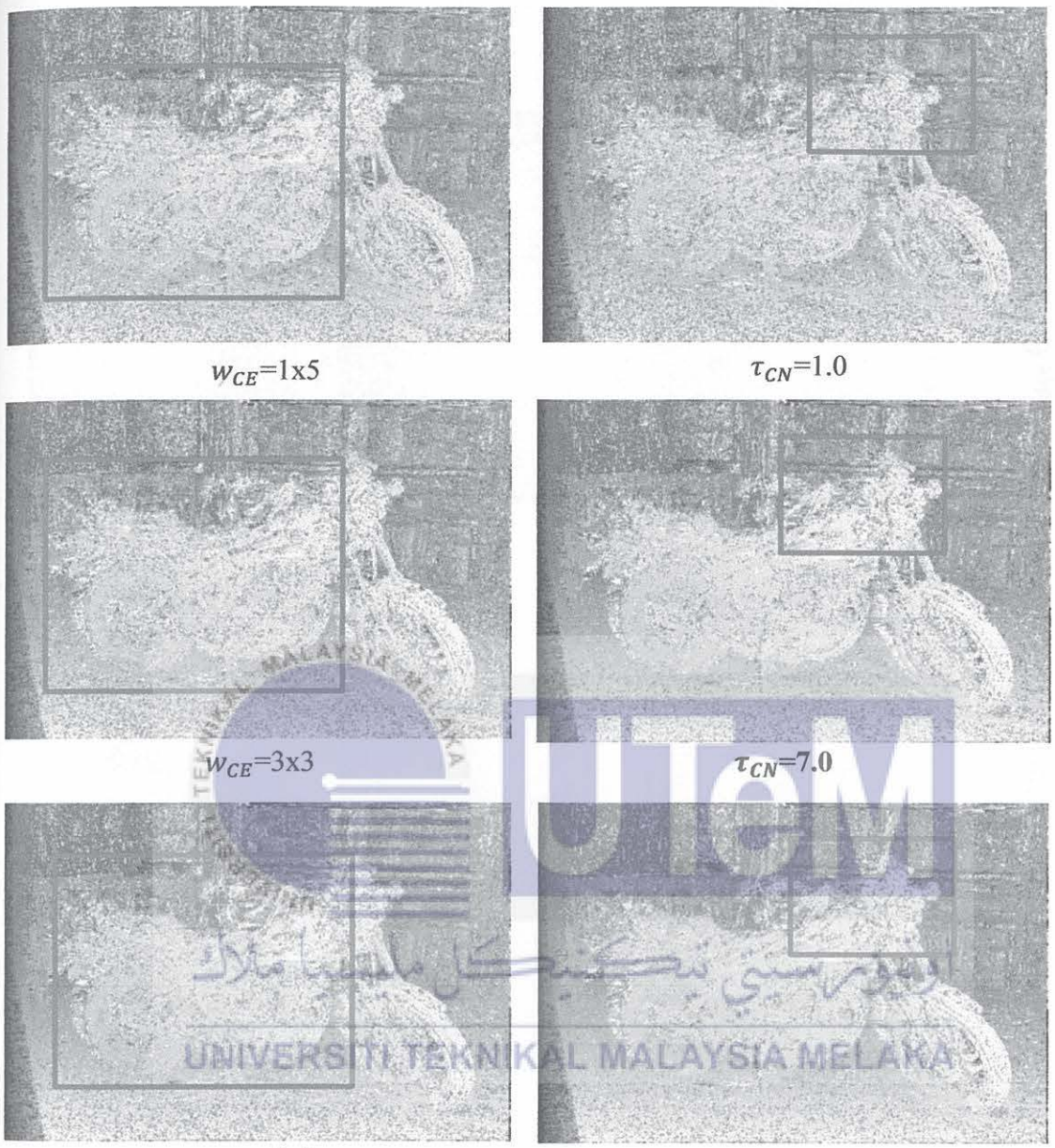


Figure 4.4: Qualitative Evaluation of Middlebury Motorcycle Parameter Selection (a) w_{CE} (b) τ_{CN}

Figure 4.5: Qualitative Evaluation of Middlebury Motorcycle Parameter Selection (a) τ_{diff} (b) τ_{edge} (c) $\sigma_{CN} + \sigma_{ED}$

Next, the parameter established in the MCE cost was the τ_{diff} given in Figure 4.3(c). A very little change was observed when τ_{diff} was increased from 60.0 to 200.0 which showed the accuracy peak at τ_{diff} = 190 for 42.2% *all* error and 34.8% *nonocc* error. As shown in Figure 4.5(a), there was a slightly significant quantitative improvement in *all* error accuracy around 1.0% from τ_{diff} = 60 until τ_{diff} = 190 and the disparity was not clearly visible. Then, Figure 4.3(d) displays the experimental results for parameter τ_{edge} which was the lowest of *all* error at 41.6% and *nonocc* error at 34.0% for τ_{edge}= 25.0. The line graph of τ_{edge} depicts the accuracy that showed a slight increase from τ_{edge}= 5.0 until τ_{edge}= 25.0 and a slight decrease until τ_{edge}= 50.0. Figure 4.5(b) illustrates the disparity maps of τ_{edge} for 5.0, 25.0, and 45.0 values showing the handle of the in the red box for τ_{edge}=25.0 having a smoother and sharper image than τ_{edge}=5.0 and τ_{edge}=45.0. The last parameter selected in the MCE was the combination census texture and edge, $\sigma_{CN} + \sigma_{ED}$. Figure 4.3(e) shows the downward tendency of accuracy for *all* and *nonocc* error from 46.3% to 40.9% and 39.6% to 33.3% at $\sigma_{CN} + \sigma_{ED}$= 0.7+0.3. Then, the line graph accuracy was stabilised to reach $\sigma_{CN} + \sigma_{ED}$= 1.0+1.0. The effect of $\sigma_{CN} + \sigma_{ED}$ was observed in the red box area of Figure 4.5(c), which demonstrated that the edge became smoother and sharper since the salt and pepper noise was reduced for $\sigma_{CN} + \sigma_{ED}$= 0.7+0.3 compared to other parameter value.

The execution of PPF in the matching cost required four parameters to be determined. The first parameter was the pyramid layers, k. Figure 4.6(a) graphically depicts this experimental result which provided the minimum constant value of 44.1% *all* error and 37.0% *nonocc* error at k=3. Between k=1 to k=3, the values were slightly decreased from 44.5% to 44.1% for *all* error and from 37.4% to 37.0% for *nonocc* error. However, from k=4 to k=10, the results became less accurate as the value dramatically increased reaching 57.5% and 50.0% for *all* and *nonocc* errors. Figure 4.7(a) presents a visual disparity map for k which displays

the Middlebury Recycle image. The red box indicates the surface of the front body dustbin was sharper and smoother for k = 3 compared with k = 1 and k = 10 since the edge distortion, salt and pepper noise were reduced. Next, 3 balancing parameters between cost volume differences in the pyramid were specified, $\sigma_{LT} = 0.5$, $\sigma_{LG} = 0.5$, and $\sigma_{LM} = 0.7$. Figure 4.6(b) shows the results of line graph of *all* error slightly dropped to the minimum from 40.8% to 40.6% for $\sigma_{LT} = 0.1$ and $\sigma_{LT} = 0.5$, while *nonocc* error slightly fell to the minimum from 33.2% to 33.1% for $\sigma_{LT} = 0.1$ and $\sigma_{LT} = 0.5$. Both line graphs were maintained after these values. Figure 4.6(c) shows the downward tendency to the minimum of σ_{LG} for *all* and *nonocc* error from 42.4% to 40.7% and 35.2% to 33.1% at $\sigma_{LG} = 0.5$ and slightly increased at $\sigma_{LG} = 0.6$ until $\sigma_{LG} = 1.0$.

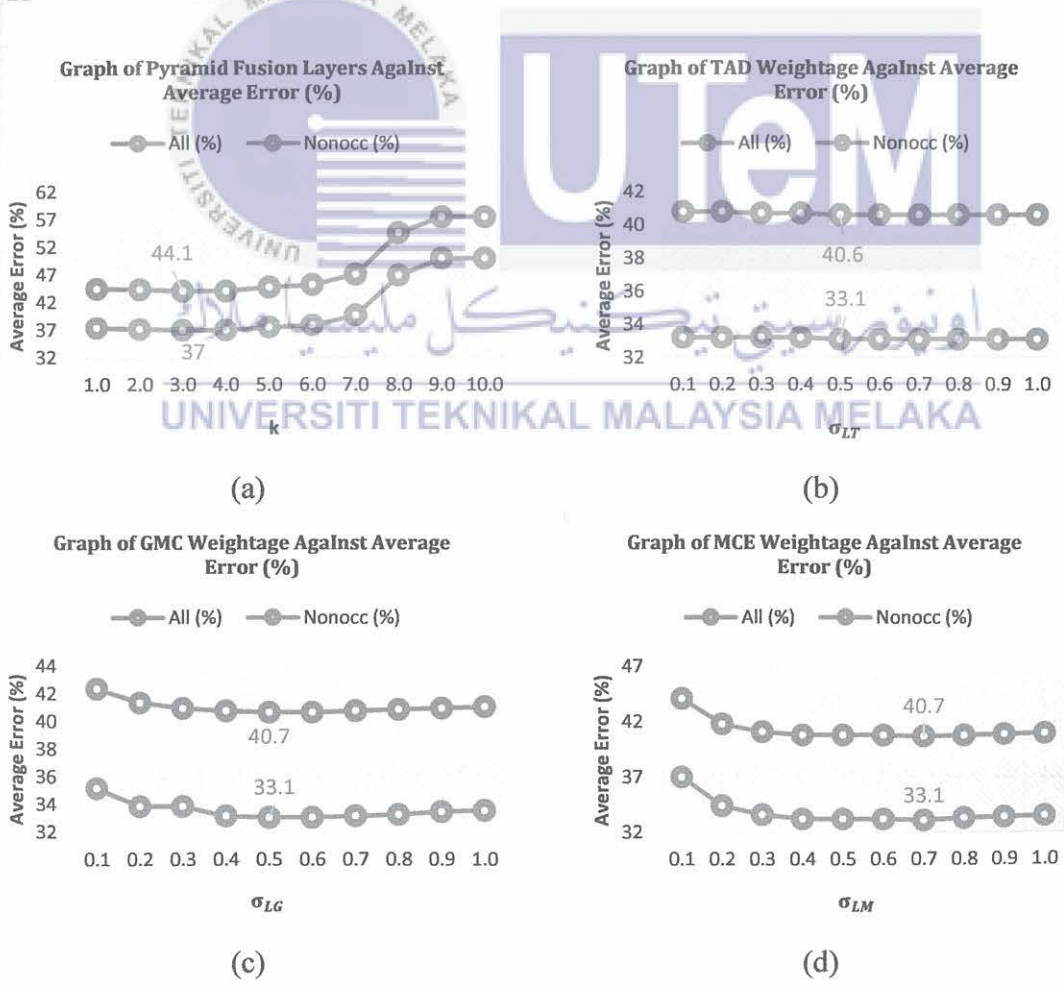


Figure 4.6: Line Graphs of Parameter Selection for PPF (a) k (b) σ_{LT} (c) σ_{LG} (d) σ_{LM}

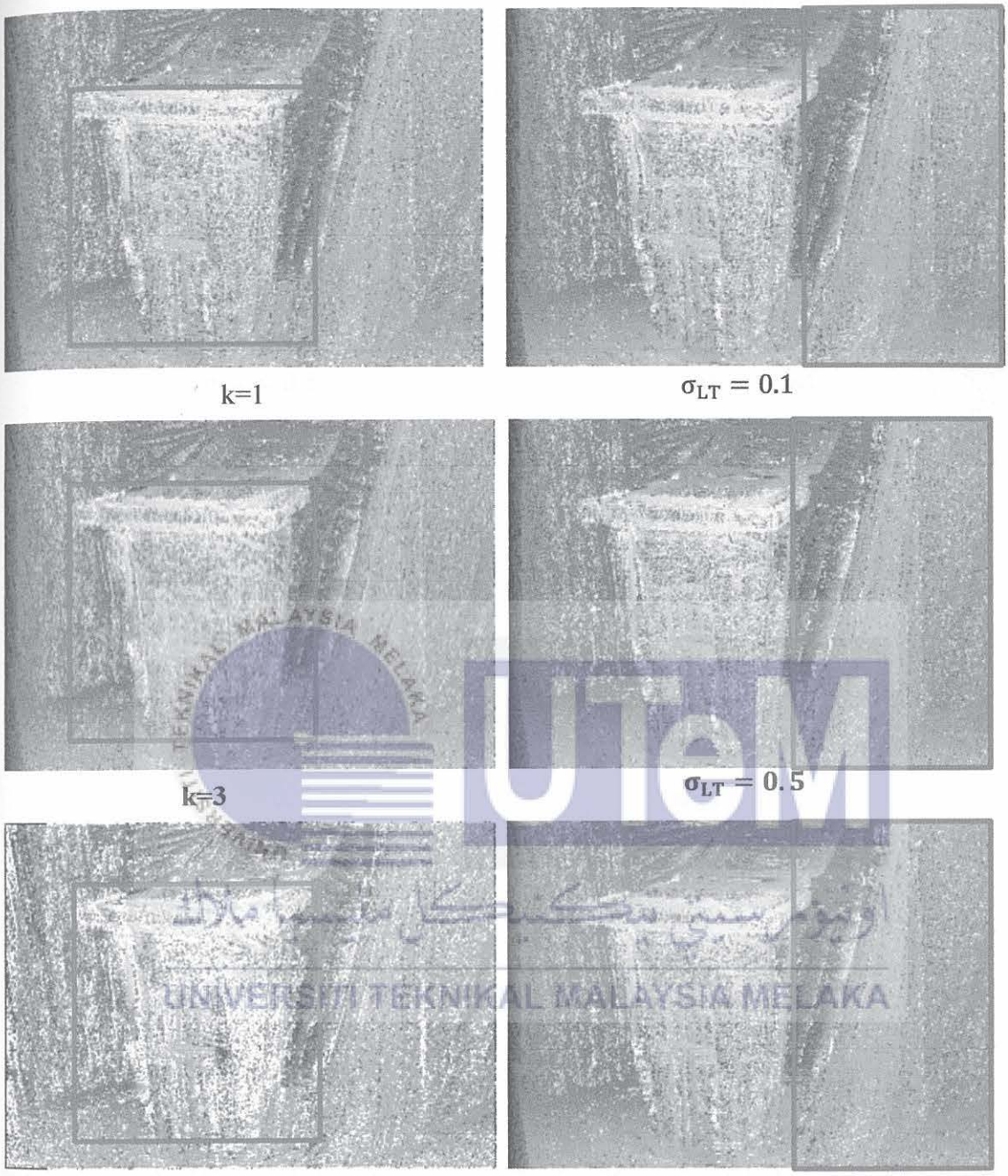


Figure 4.7: Qualitative Evaluation of Middlebury Recycle PPF Parameter (a) k (b) σ_{LT}

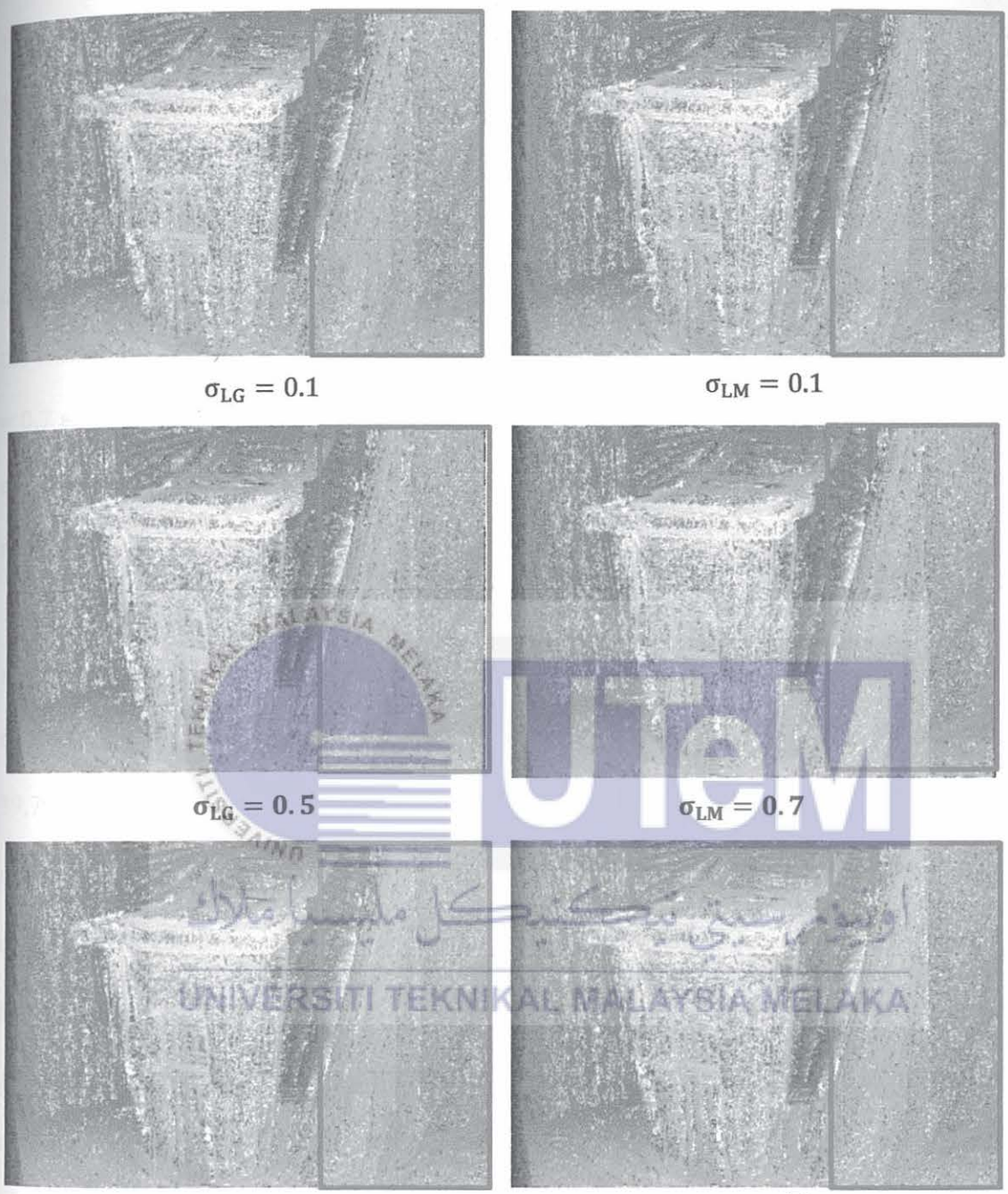


Figure 4.8: Qualitative Evaluation of Middlebury Recycle PPF Parameter (a) σ_{LG} (b) σ_{LM}

Figure 4.6(d) depicts a line graph for σ_{LM}, which shows the minimum constant value when $\sigma_{LM} = 0.7$, resulting in an *all* error of 40.7% and a *nonocc* error of 33.1%. The line graph tendency displayed a downward trend starting from $\sigma_{LM} = 0.1$ until $\sigma_{LM} = 0.7$ and the

trend continued to slightly increase after $\sigma_{LM} = 0.8$. The qualitative performance for the balancing parameters between cost volume differences in the pyramid can be observed in Figure 4.7(b), Figure 4.8(a), and Figure 4.8(b). There was no significant visible variation in the texture and edges for σ_{LT} since the accuracy value between them contributed only minor differences. As expected, there were significant visible variations for σ_{LG} and σ_{LM}, where the texture and edges are smoother for $\sigma_{LG} = 0.5$ and above, with the same condition for $\sigma_{LM} = 0.7$ and above, as presented in the red box. Although the salt and pepper noise were still apparent in the disparity map, the line balancing parameters successfully reduced the noise, respectively.

Stage 2 – Cost aggregation: The line graph iNLGF parameter selection at the cost aggregation stage is shown in Figure 4.9. The selection started with the determination of the radius of the search window, w_q. As shown in Figure 4.9(a), the lowest average error was 39.7% for *all* errors and 31.9% for *nonocc* errors at $w_q = 15.0$. The next parameter was the radius comparison window, w_p, which showed the lowest average error of 36.4% for *all* errors and 28.2% for *nonocc* errors at $w_p = 9.0$. Figure 4.9(b) shows a significant drop in accuracy from $w_p = 3.0$ to $w_p = 9.0$, followed by a slight increase until $w_p = 17.0$. Next, the selected parameter was the iteration for iNLGF, n. According to Figure 4.9(c), the lowest minimum value at n = 3 was 34.6% for *all* errors and 26.0% for *nonocc* errors. The accuracy value was uniformly decreased from n = 1 to n = 3 and showed a significant increase from n = 4 until n = 10.

Figure 4.9(d) depicts the smoothness term for iNLGF with a line graph showing a slight drop from $\varepsilon = 0.1$ to $\varepsilon = 0.3$ and then, remained stable. The selected parameter value for the smoothness term was at $\varepsilon = 0.3$. The last parameter determined for the iNLGF was the iNGF radius support window, w_g. The line graph started from $w_g = 3.0$, which indicating

the minimum value contributed optimum accuracy of 19.7% for *all* error and 11.9% for *nonocc* error. The trend of the line graph was uniformly increasing until w_g= 21.0. The quality evaluation when determining the iNLGF parameter is shown in Figure 4.10. Figure 4.10(a) compares the three disparity maps for the w_q = 7, w_q = 15, and w_q = 27. The disparity map in the red box shows the salt and pepper noise was reduced when w_q = 15, and w_q = 27, making the edges more apparent. The performance w_p can be observed in Figure 4.10(b) with three disparity maps for comparison, w_p = 3, w_p = 9, and w_p = 17. The disparity map in the red box at the background and cup for w_p = 9 shows less salt and pepper noise and was sharper compared with other disparity maps. The qualitative performance for parameter iNLGF iteration, n, is compared in Figure 4.11(a). With comparison to the other textures, the red box displays the texture of the chair in n = 3 has less edge distortion and noise. When n was increased to 10, the chair texture continued to worsen. Similar results were obtained by increasing the smoothness term parameter, ε from 0.3 to 1.0, as shown in Figure 4.11(b). The radius support window, w_g as depicted in Figure 4.11(c), was the final parameter qualitatively evaluated in the iNLGF. The complex textures and edges, particularly at the red cup in the red box, were destructed, contributing to the low accuracy, even though w_g = 11 and w_g= 21 delivered significantly smoother textures than w_g = 3.0.

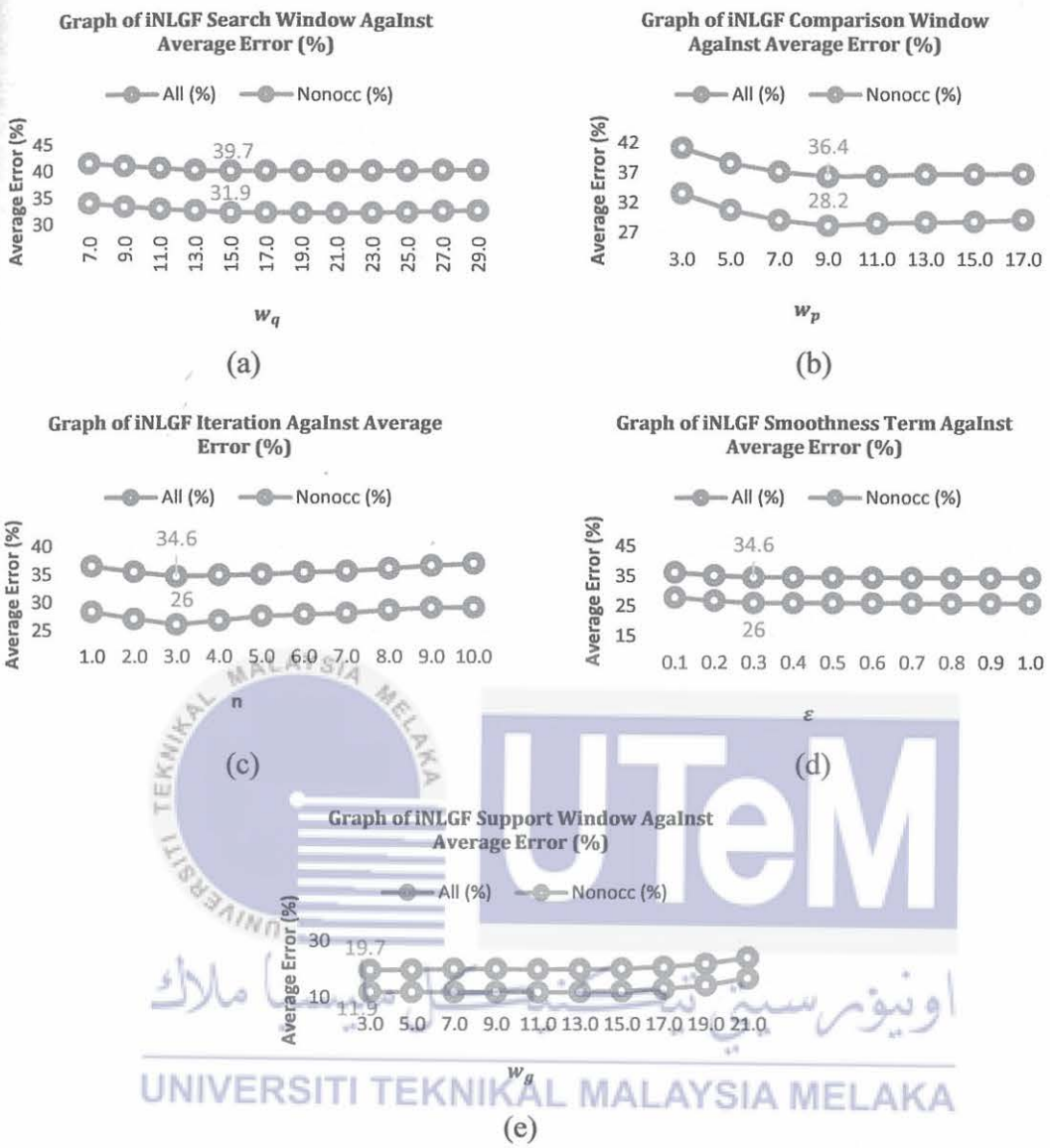
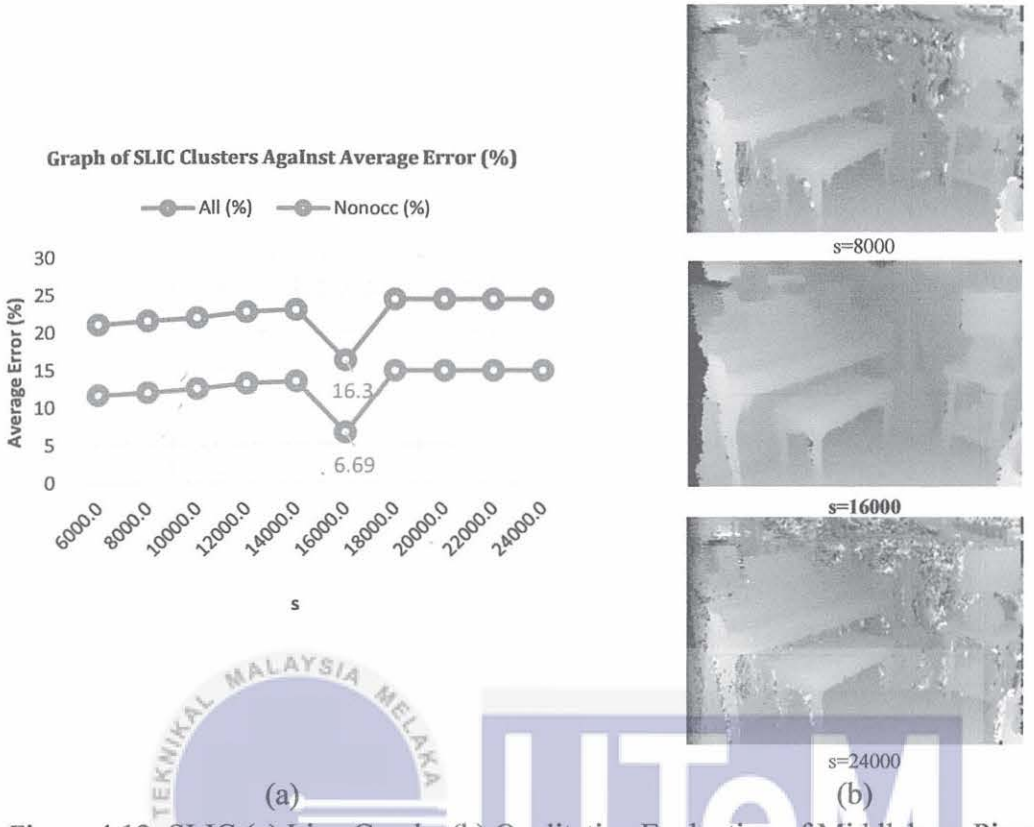


Figure 4.9: Line Graphs of Parameter Selection for iNLGF (a) w_q (b) w_p (c) n (d) ε (e) w_g

Figure 4.10: Qualitative Evaluation of Middlebury Adirondack iNLGF Parameter Selection

(a) w_q (b) w_p

Figure 4.11: Qualitative Evaluation of Middlebury Adirondack iNLGF Parameter Selection (a) n (b) ε (c) w_g



(a) (b)

Figure 4.12: SLIC (a) Line Graphs (b) Qualitative Evaluation of Middlebury Piano

Next, the parameter SLIC, s was determined in the cost aggregation. The SLIC parameter showed the lowest minimum constant accuracy at s = 16000, producing an average accuracy of 16.3% *all* error and 6.69% *nonocc* error. The line graph in Figure 4.12(a) shows the accuracy is increasing linearly from s = 6000, but the accuracy trend showed a massive drop at s = 16000, increased again after s = 18000, and stabilised at s = 22000. Figure 4.12(b) displays the comparison of disparity maps between s = 8000, s = 16000, and s = 24000 for Middlebury Piano dataset. As predicted, the lowest accuracy was obtained at s = 16000, which results in a texture that was sharp and a disparity map that preserved the boundaries. On the contrary, the s = 8000 and s = 24000, showed a lesser level of accuracy. Salt and pepper noise, horizontal streaks, and edge distortion were visible in both s = 8000 and s = 24000, particularly at the top right and bottom left, where there was an illumination variation constraint.

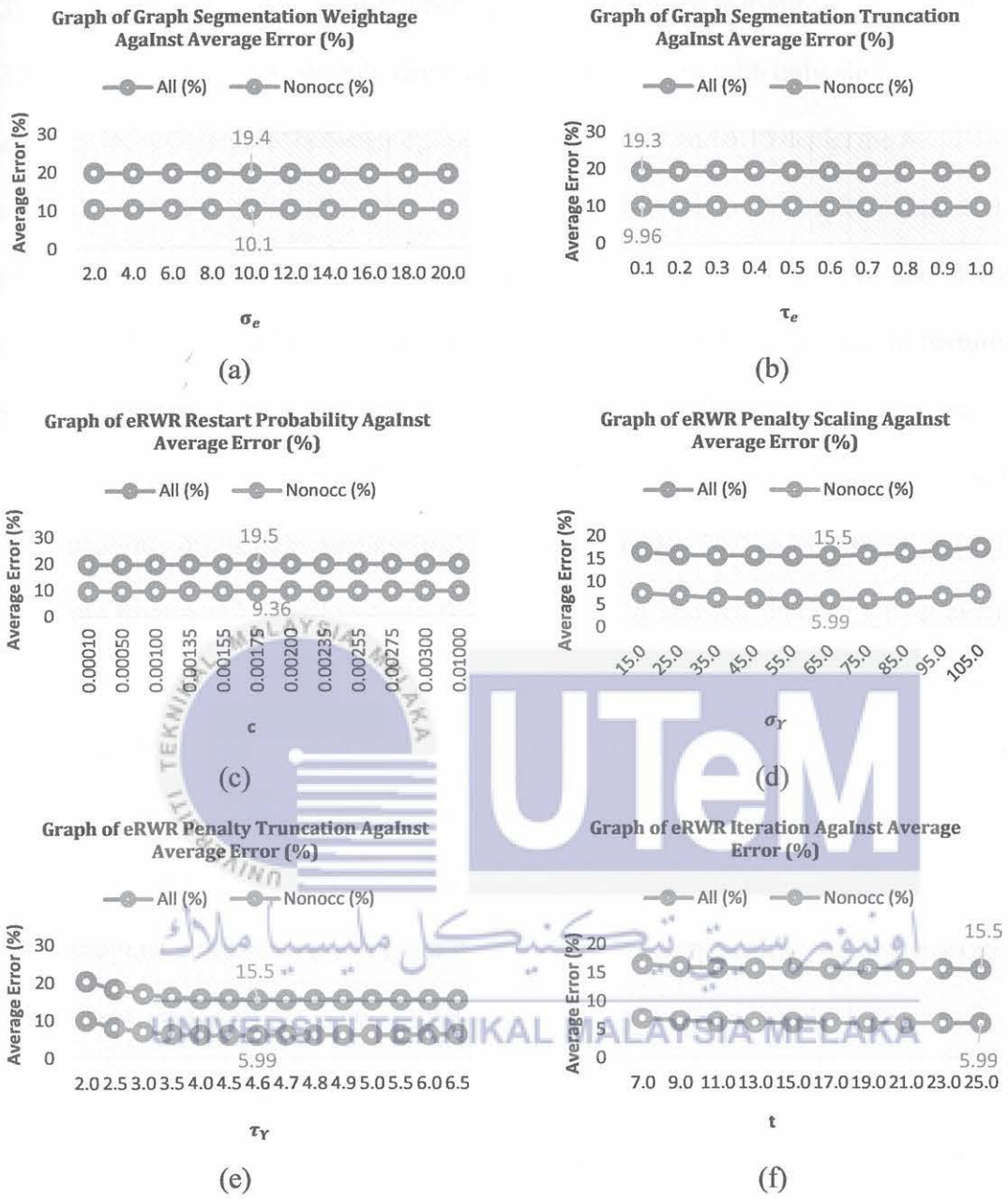


Figure 4.13: Line Graphs of Parameter Selection for Graph Segmentation and eRWR (a) σ_e (b) τ_e (c) c (d) σ_Y (e) τ_Y (f) t

Figure 4.13, Figure 4.14 and Figure 4.15 display the parameters established during the graph segmentation and eRWR. The graph segmentation consisted of two parameters: graph weightage, σ_e and graph truncation, τ_e. The lowest average accuracy for σ_e was 19.4% *all* error and 10.1% *nonocc* error at $\sigma_e = 10.0$ followed by an average accuracy for τ_e at

19.3% *all* error and 9.96 *nonocc* error when $\tau_e = 0.1$ as tabulated in Figure 4.13(a) and Figure 4.13(b). The line graph was slightly decreasing in both figures, with only slight differences in accuracy for each interval between σ_e and τ_e. Figures 4.14 and 4.15 depict the Middlebury Adirondack used to evaluate the disparity map's qualitative performance for parameters in graph segmentation and eRWR. The disparity map in Figures 4.14(a) and 4.14(b) demonstrates that, as predicted, there were no significant visible differences in texture or edges throughout the experimental observations due to the small accuracy differences.

Then, the parameters of eRWR were determined and established which include the restart probability (c), penalty scaling (σ_Y), penalty truncation (τ_Y) and the iteration (t). For experiments from c = 0.00010 to c = 0.01000, the results showed that the lowest average accuracy for restart probability was 19.5% for *all* error and 9.36% for *nonocc* at c = 0.00175 as shown in Figure 4.13(c). The *nonocc* error contributed 0.1% between the lowest and maximum accuracy, whereas *all* error, at 19.5%, was the value measured throughout the entire interval. Thus, there was no apparent change in the disparity map between c = 0.0001, 0.00175 and 0.01 as displayed in Figure 4.14(c). Next, penalty scaling, σ_Y and truncation, τ_Y were determined with the lowest accuracy of 15.5% for *all* error and 5.99% for *nonocc* error at $\sigma_Y = 65.0$ and $\tau_Y = 4.6$. The line graph for σ_Y showed an increase in a uniform manner while the line graph for τ_Y showed a decreasing trend, as presented in Figure 4.13(d) and Figure 4.13(e). The disparity map was more visible and sharper when $\sigma_Y = 65.0$, especially in the area of the cup, as shown in Figure 4.15(a), whereas when $\tau_Y = 4.6$, the area in the red box was smoother and sharper, with less edge distortion and noise as displayed in Figure 4.15(b).

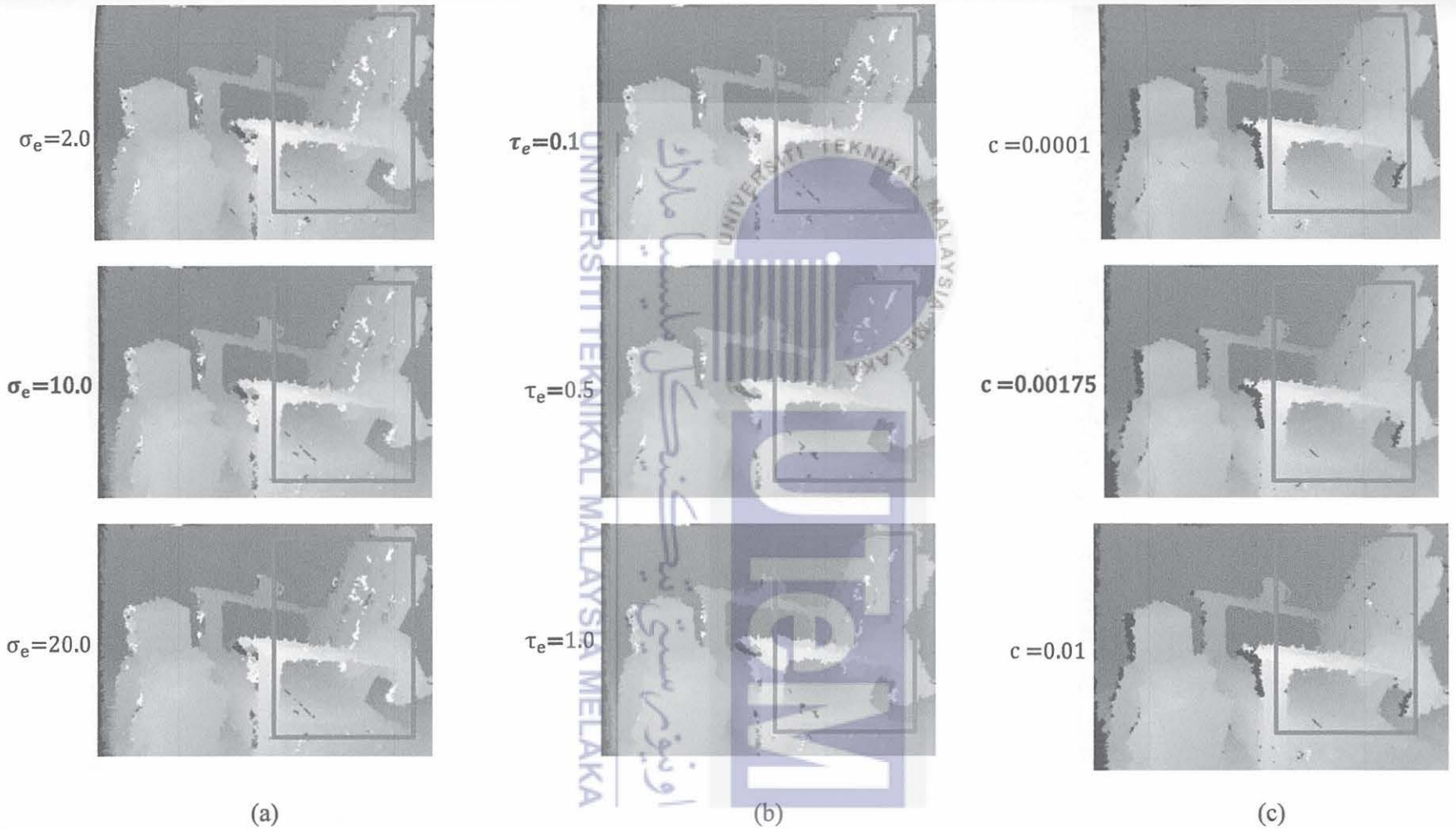


Figure 4.14: Qualitative Evaluation of Middlebury Adirondack Parameter Selection for Graph Segmentation and eRWR (a) σ_e (b) τ_e (c) c

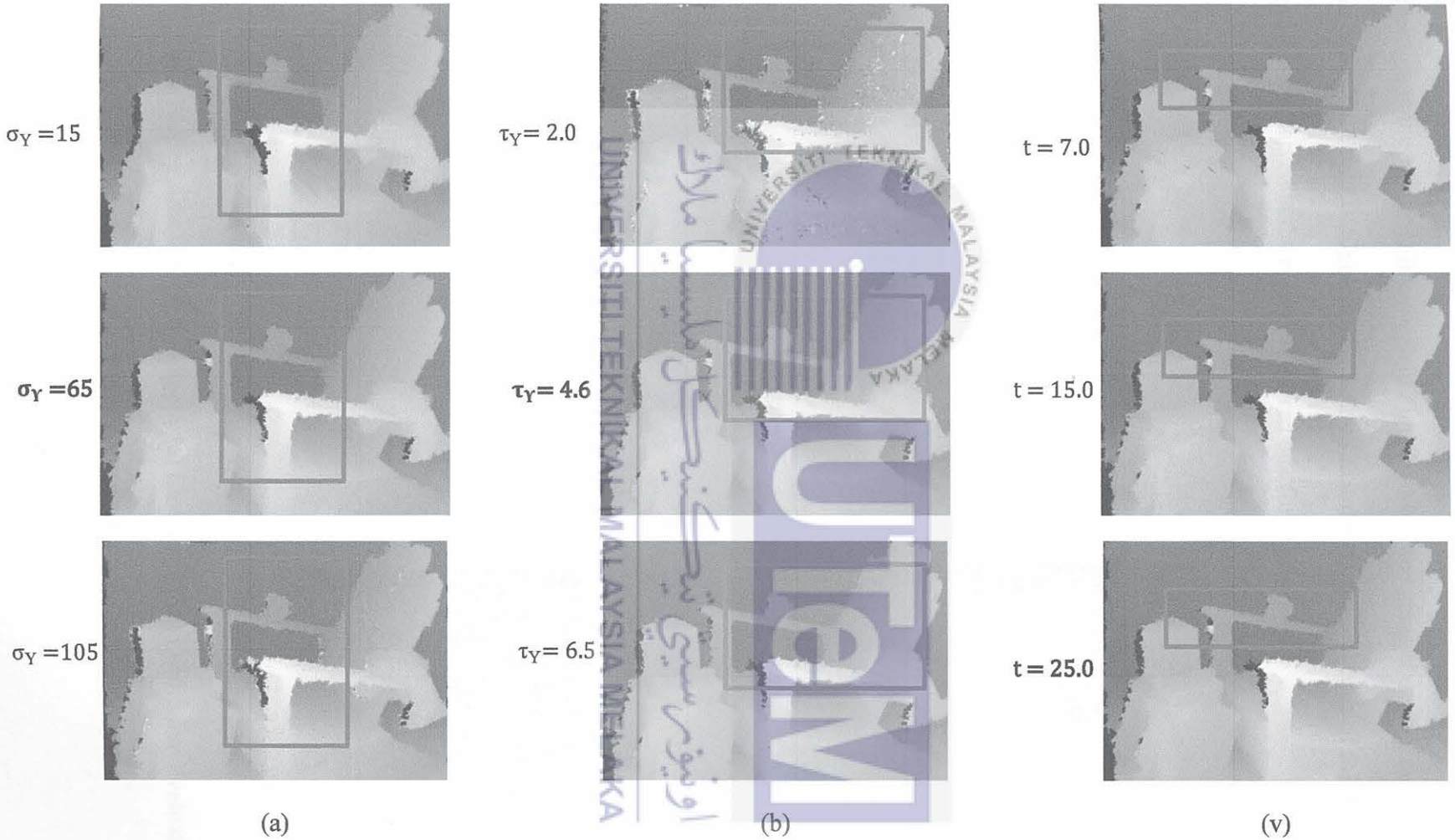


Figure 4.15: Qualitative Evaluation of Middlebury Adirondack Parameter Selection for eRWR (a) σ_Y (b) τ_Y (c) t

The final eRWR parameter, t, contributed to the lowest average accuracy of 15.5% for *all* error and 5.99% for *nonocc* error, with a minimum constant value at t = 25.0. According to the tabulated line graph in Figure 4.13(f), the graph showed a uniform increase from t = 7.0 to t = 25.0. Figure 4.15(c) shows that the texture and edge of the disparity map were well-preserved in the region of the red box for the cup.

(a) (b)

Figure 4.16: Segment Cost Weightage, γ (a) Line graphs (b) Qualitative Evaluation of Middlebury Teddy

The segment cost weightage, γ which was used to combine the segment cost and the pixel cost, was the last parameter in the cost aggregation stage. Line graph in Figure 4.16(a) indicated when the value γ was increased from 0.0001 to 0.0008, the accuracy line graph was stabilised. From 0.0010 to 0.03 there was a significant increase in the accuracy of the line graph. When the value, γ was equal to 0.008, the lowest average accuracy obtained was 15.5% for *all* error and 5.99% for *nonocc* error. The disparity map was compared to three measurements: $\gamma = 0.0001$, $\gamma = 0.0008$, and $\gamma = 0.03$ as seen in Figure 4.16(b). When compared to $\gamma = 0.03$ with $\gamma = 0.0008$, the implementation of this parameter greatly diminished the salt and pepper noise. Hence, the stage 1 final values were used as the parameters in the cost aggregation studies.

S*tage 3 – Disparity selection*: At this stage, there was no parameter employed and the disparity selection was based on the WTA method. The WTA technique was used to determine the amount of disparity value using the minimum value of the raw data from Stage 2.

Stage 4 - Disparity refinement: The results of the LR consistency checking method, τ_{LR} and invalid pixel fill-in, τ_{CF} are presented in Figure 4.17. This was composed of the black and white pixels which were corresponded as valid and invalid. The valid pixel values described in Section 3.6 were employed as substitutes to the white color pixels. The findings showed that the τ_{LR} produced the lowest average value at $\tau_{LR} = 1.0$ meanwhile $\tau_{CF} = 0.3$. The maximum and lowest accuracy for both line graphs showed only small differences at 0.1%, as shown in Figure 4.17(a) and Figure 4.17(b). In comparison to $\tau_{LR} = 0.3$ and $\tau_{LR} = 2.3$, as depicted in Figure 4.13(c), the algorithm was capable of identifying a greater number of invalid pixels at $\tau_{LR} = 1.0$. Figure 4.17(d) indicates that the disparity maps for τ_{CF} having an insignificantly qualitative difference when adjusted from 0.001 to 3.00, despite the fact that the quantitatively, the value did show some differences.

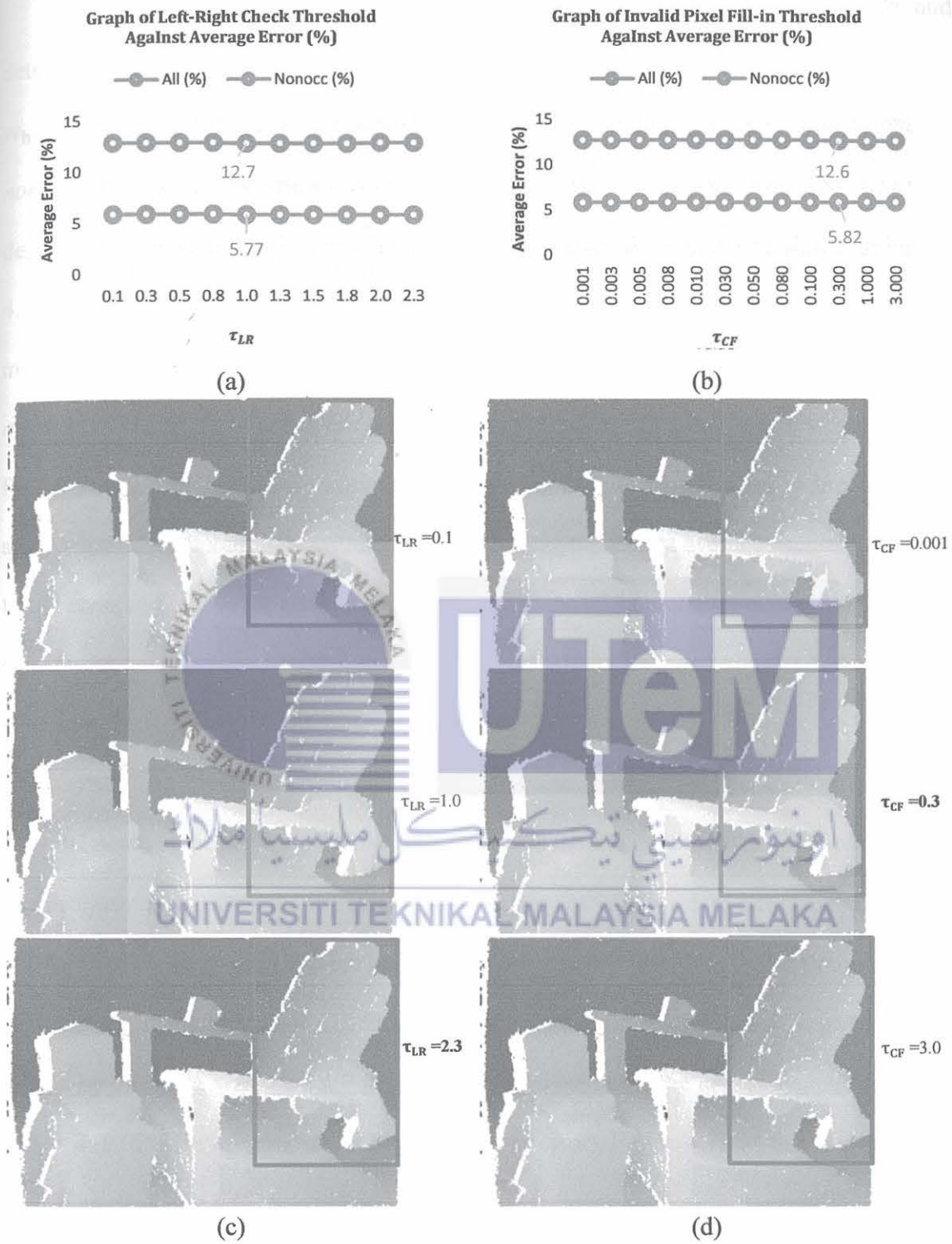


Figure 4.17: Parameter Selection for LR Consistency Checking and Invalid Pixel Fill-in (a) τ_{LR} Line Graph (b) Quality Evaluation of τ_{LR} (c) τ_{CF} Line Graph (d) Quality Evaluation of τ_{CF}

In this research, three vital K-means clustering parameters were selected and determined. The first K-means parameter was the radius of the support window denoted by w_h. When $w_h = 41.0$, the lowest average accuracy was 9.36% for *all* errors and 5.25% for *nonocc* errors. When the interval was increased from $w_h = 3.0$ to $w_h = 41.0$, the line graph depicted a downward trend, with a little increase in the trend thereafter as shown in Figure 4.18(a). Figure 4.18(b) shows that the line graph for the number of clusters, s, used in the K-means did not change significantly between the lowest and highest accuracy, with 0.14% for *all* error and 0.04% for *nonocc* error. At s = 32.0, the average accuracy was lowest at 9.22% for *all* errors and 5.21% for *nonocc* errors. Next, the third parameter of the K-means approach was the iteration, h, with h = 20.0 delivering the lowest average accuracy of 9.2% for all error and 5.80% for nonocc error. The line graph of the iteration between the highest and lowest accuracy was 0.04% for all error and 0.03% for nonocc error indicating no significant changes.

The final approach employed in this SMA was the SWF, which comprised of two essential parameters: the SWF window radius, r, and the SWF iteration, n_f. The parameter r was determined to 3 with the lowest average accuracy of 9.04% for all errors and 5.13% for *nonocc* error. When r was adjusted by an interval from r = 3 to r = 21, the accuracy value in the tabulated line graph shown in Figure 4.18(d) increased uniformly. The final parameter chosen by the SMA was the SWF iteration, which contributed the lowest average accuracy at $n_f = 4$ (9.02% for *all* error and 5.11% for *nonocc* error). Figure 4.18(e) shows a tabulated line graph depicting a slightly upwards trend in n_f accuracy from r = 4 to r = 40.

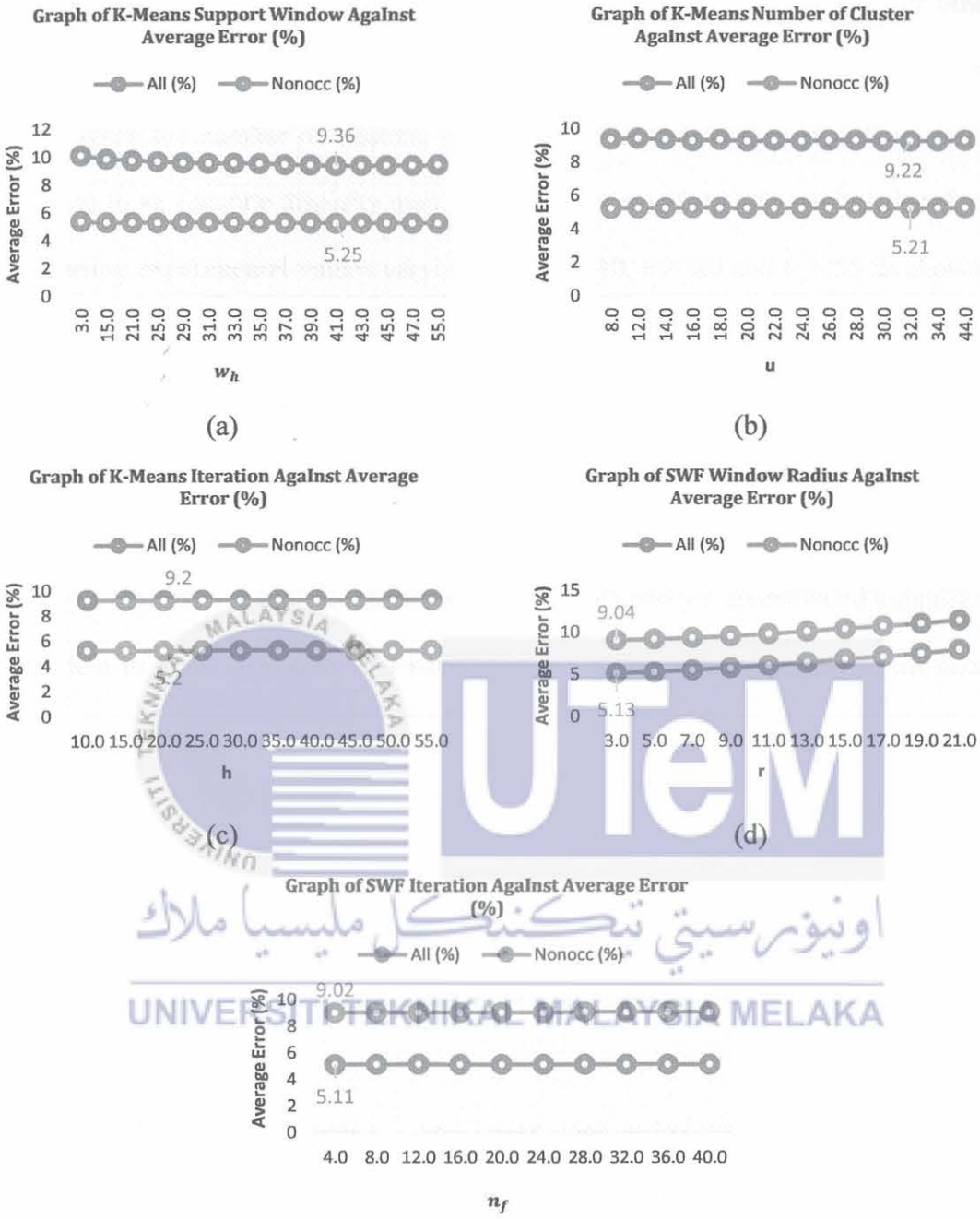


Figure 4.18: Line Graphs of Parameter Selection for K-means and SWF (a) w_h (b) s (c) h (d) r (e) n_f

The qualitative performance of the K-means and SWF parameters is shown in Figure 4.19 and Figure 4.20. When the disparity map for K-means, w_h was executed for $w_h = 3$, $w_h = 41$ and $w_h = 55$, as shown in Figure 4.19(a), the texture was smoother while the edge

was well-preserved for $w_h = 41$, especially at the chair in the red box compared with others. Figure 4.19(b) demonstrates the disparity map was smoother and the area of occlusion was improved when the number of clusters, u = 32, as opposed to u = 8 as well as u = 44, contributing to an accurate disparity map. The disparity map for K-means iteration, h was assessed using experimental values varying from h = 10, h = 20 and h = 55 as shown in Figure 4.19(c). There is a minor improvement where the disparity map in the red box for h = 20 was smoother and the texture was clearer with an improved occlusion region compared to others.

Figure 4.20(a) displays the disparity map execution for the radius window of the SWF parameter, r. Observation indicated that when r = 3, the disparity map exhibited a significant improvement in terms of texture and edge. The smoothing and sharpening of the chair's edges were visibly observed in the red box region compared to other regions where the chair was oversmoothed by the increase size of radius window. Figure 4.20(b) shows the SWF iteration disparity map, which concluded the parameter selection study. The occlusion was minimised when $n_f = 4$. This was visible within the red box region. While increasing n_f to $n_f = 24$ and $n_f = 40$, the texture and edges were oversmoothed, resulting in an inaccurate disparity map.

The final values from stages 1, 2, and 3 were employed as evaluations for the disparity refinement parameters. The disparity maps had a great number of invalid pixels at the beginning of the selection process, therefore producing lower quality disparity maps. However, as more optimal settings were continuously applied to the parameters of the proposed algorithm, the disparity maps became clearer and sharper, improving the depth maps' quality and accuracy. Table 4.1 provides an overview of the parameters used and respective values in this thesis.

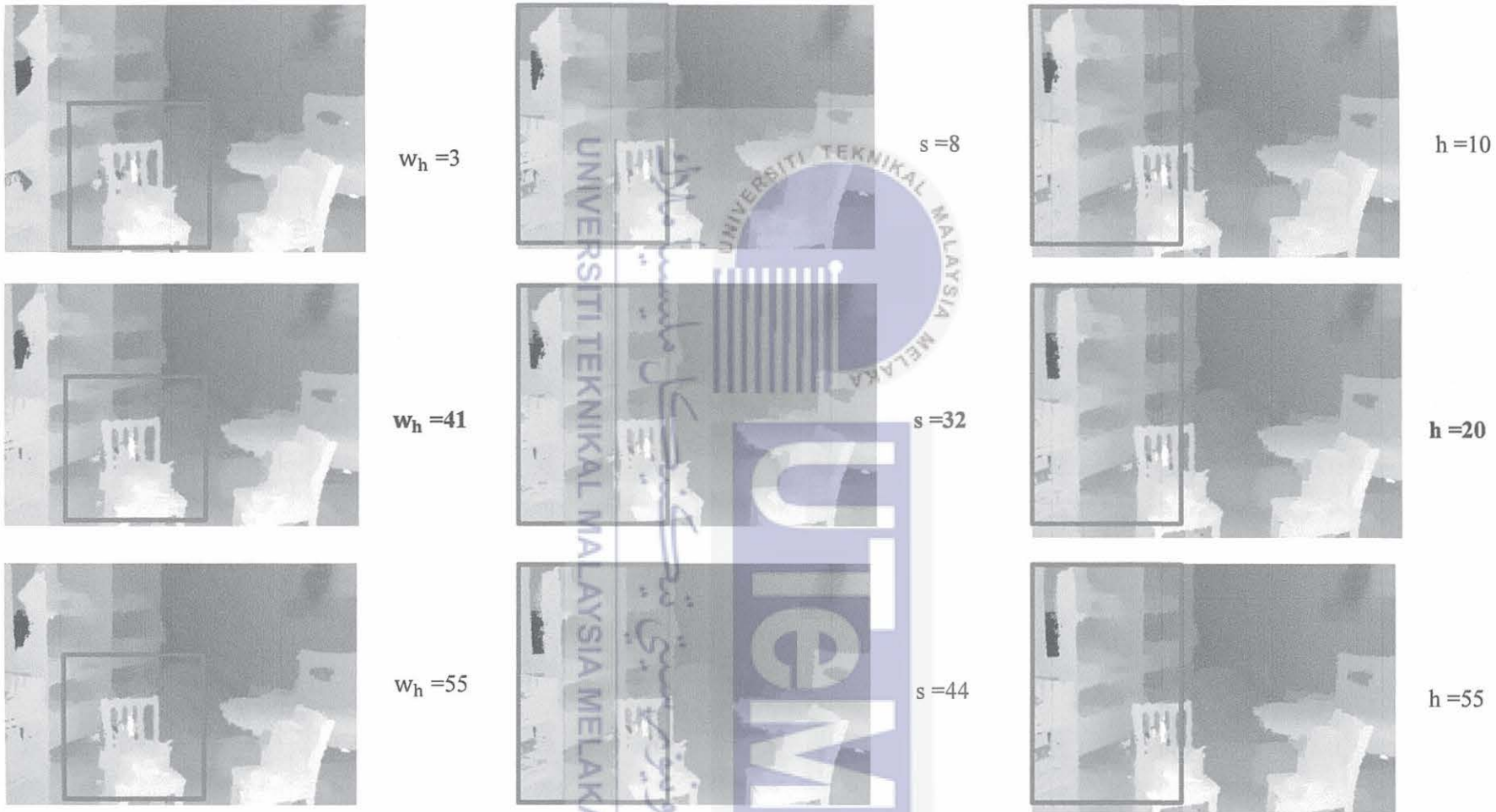


Figure 4.19: Qualitative Evaluation of Middlebury Playroom Parameter Selection for K-means (a) w_h (b) s (c) h

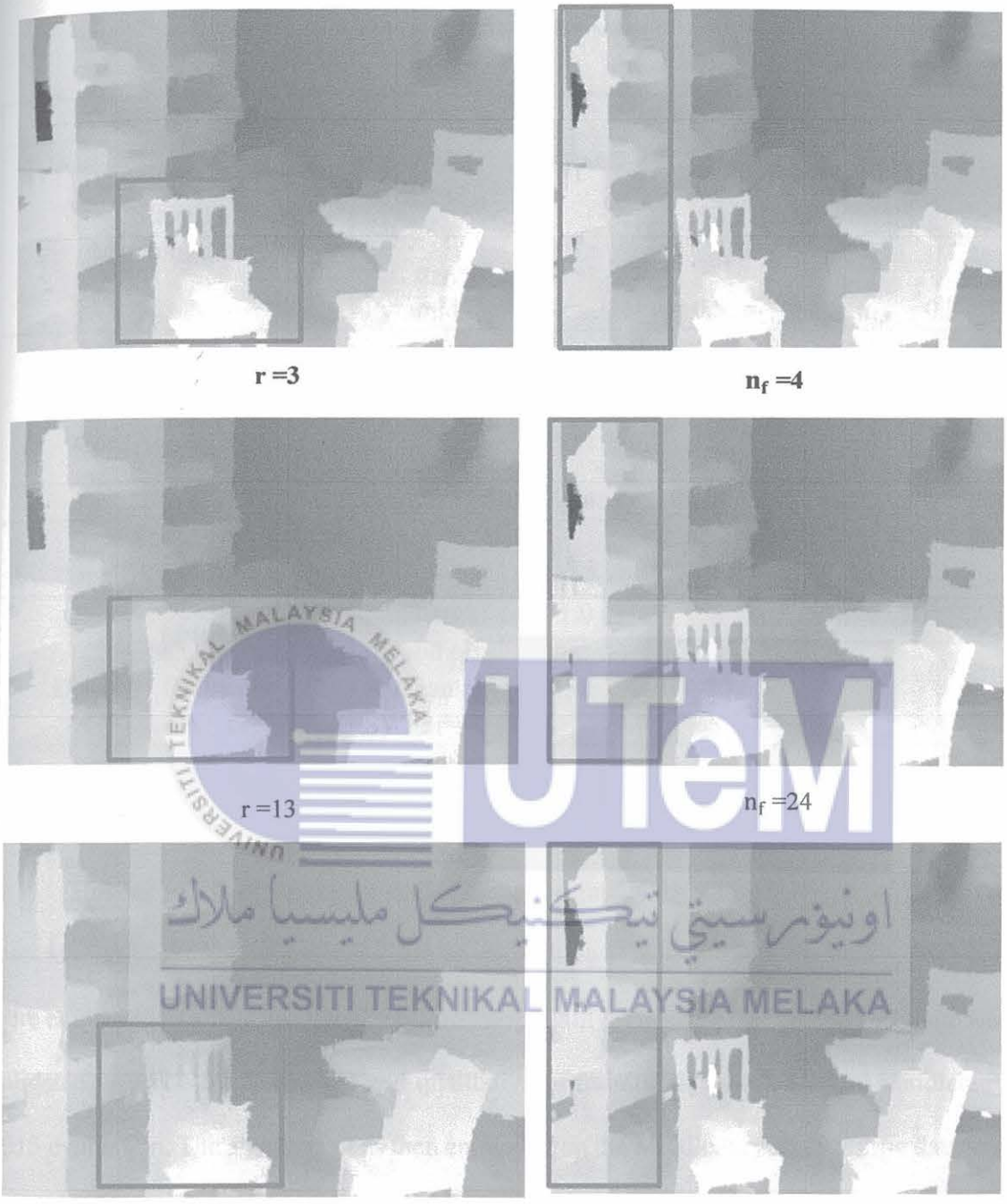


Figure 4.20: Qualitative Evaluation of Middlebury Playroom Parameter Selection for SWF

(a) r (b) n_f

Table 4.1: Summary of Parameters Used in This Work

Algorithm Stage	Parameters
Stage 1 Matching Cost	TAD: σ_{AD}= 0.7, τ_{TAD}= 0.8 GMC: τ_{GM}= 2.0 MCE: w_{CE} = 5, τ_{CN} = 7.0, τ_{diff} = 190, τ_{edge}= 25, σ_{CN}= 0.7, σ_{ED}= 0.3 Pyramid Fusion: k= 3, σ_{LT} = 0.5, σ_{LG}= 0.5, σ_{LM}= 0.7
Stage 2 Cost Aggregation	iNLGF: w_q = 15, w_p = 9, n = 3, ε= 0.3, w_g= 3 SLIC: s= 16000 eRWR: σ_e= 10.0, τ_e= 0.1, c= 0.00175, σ_γ= 65, τ_γ= 4.6, t = 25 Segment cost weight: γ= 0.0008
Stage 3 Disparity Selection	WTA
Stage 4 Disparity Refinement	L-R check: τ_{LR}= 1.0 Invalid pixel filling: τ_{CF}= 0.3 K-means: w_h= 41, u= 32, h = 20 SWF: r= 3, n_f= 4

4.3 Performance and Discussion

This section is essential for establishing and justifying the validity of the proposed research work. As stated earlier in this chapter, three prominent benchmarking databases were employed to determine the algorithm's adaptability. The performance of each stage was validated using both quantitative and qualitative of training datasets based on Middlebury 2015 evaluation. The algorithm was then employed on the Middlebury 2015 testing datasets, the KITTI 2015 training and testing datasets, the UTeMLab-Stereo datasets and 3D reconstruction.

4.3.1 Every Stage Performances

In this research, the performance for every stage was validated and analysed using the final parameter setting published in Table 4.1. The evaluation was carried out by

observing the average error accuracy for *all* error and *nonocc* error, as well as determining the quality of the disparity map view.

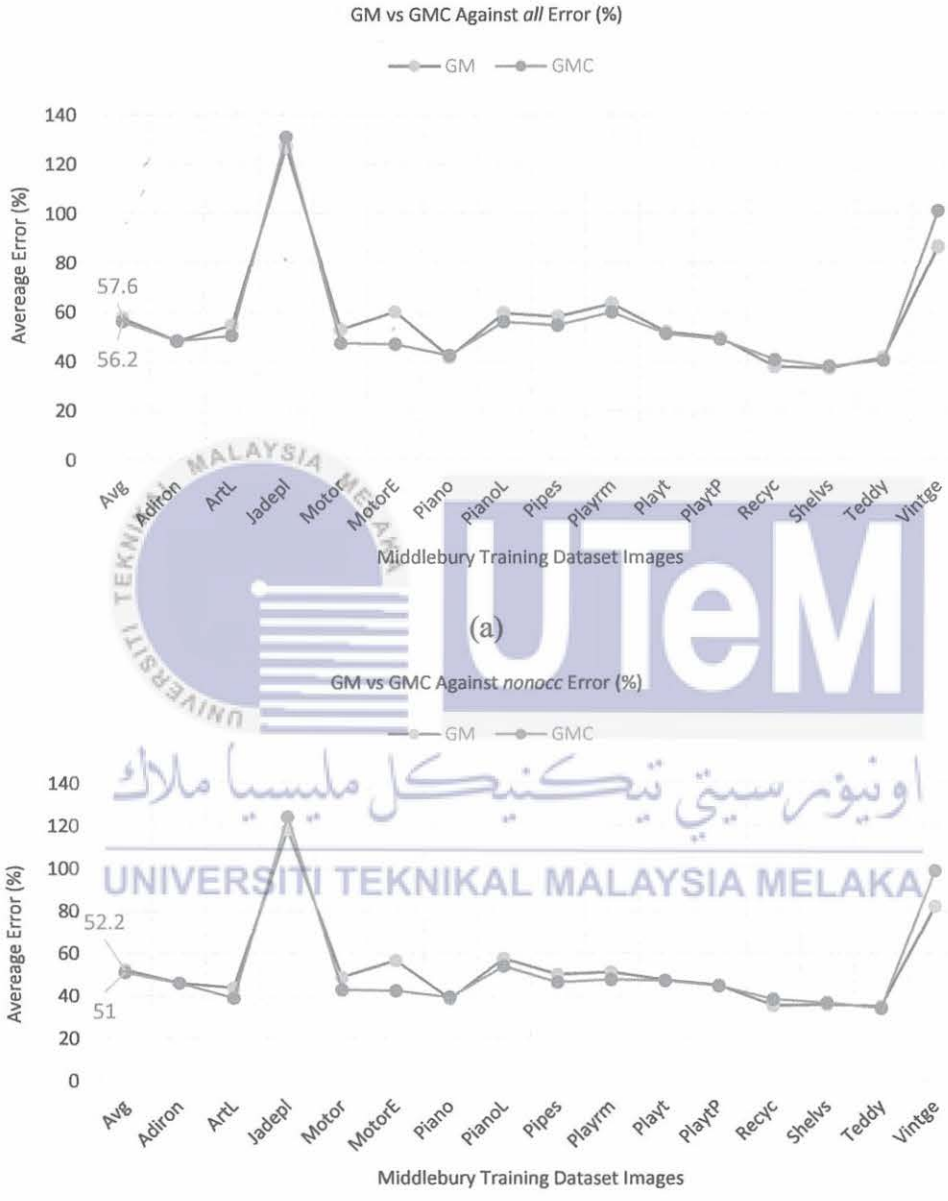


Figure 4.21: Quantitative Performance of GM with GMC (a) *all* errors (b) *nonocc* errors

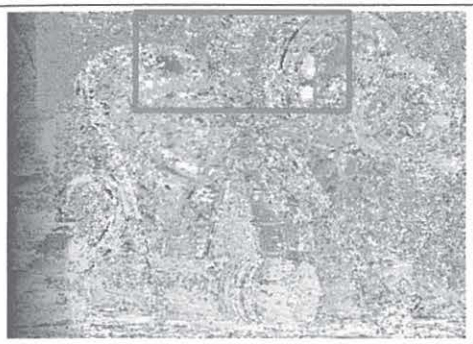
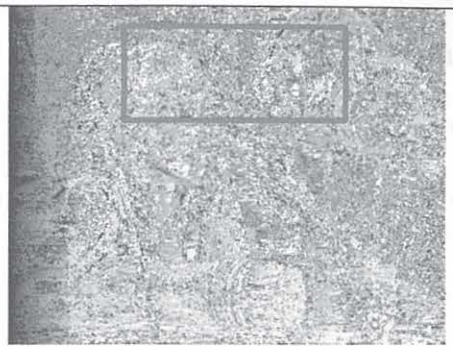
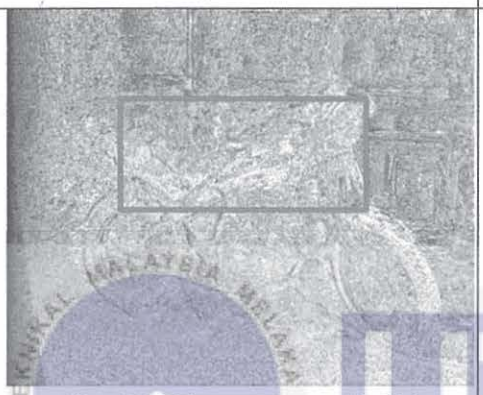
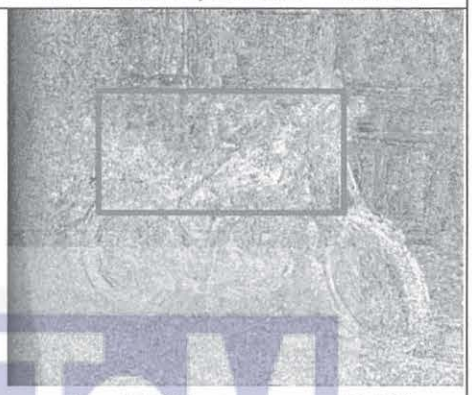
Image	GM	GMC
ArtL	 *all* = 54.4%, *nonocc* = 43.6%	 *all* = 50.2 %, *nonocc* = 38.8%
Motorcycle E	 *all* = 52.8%, *nonocc* = 48.6%	 *all* = 47.2%, *nonocc* = 42.7%
Teddy	 *all* = 41.5%, *nonocc* = 35.0%	 *all* = 40.3%, *nonocc* = 34.0%

Figure 4.22: Qualitative Performance of GM with GMC for Middlebury Images ArtL, MotorcycleE and Teddy

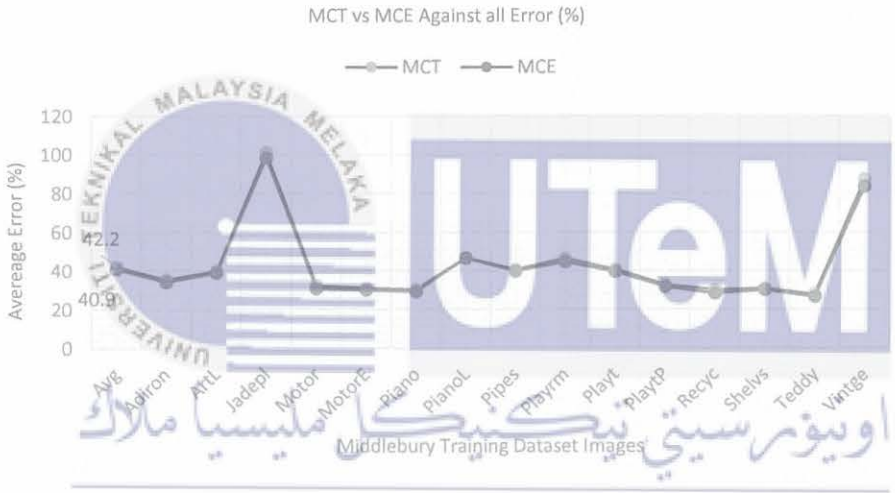
Stage 1: The evaluation was performed at this stage to determine the efficiency of the parameters in cost matching, the accuracy between GM and GMC, the accuracy between MCT and MCE, the performance of single cost, the achievement of multiple costs, which included TAD+GMC+MCE, and the performance of PPF. Figure 4.21 shows the

performance average error of *all* error and *nonocc* error for GM and GMC based on Middlebury training dataset. It was proven that the average accuracy of *all* error was improved by 1.4% from 57.6 to 56.2 and *nonocc* error by 1.2% from 52.2% to 51.0% when applying the GMC instead of GM. The accuracy had increased significantly, particularly for image with radiometric differences constraint such MotorcycleE, which reduced *all* error from 59.9% to 46.8% (3.1% change) and dropped *nonocc* error from 56.4% to 42.3% (4.1% change). The qualitative performance of GM and GMC is presented in Figure 4.22. In particular, the salt and pepper noise were distributed more uniformly at the edges and the texture for GMC was smoother. This can be seen clearly in the red box region for the disparity map of ArtL, MotorcycleE and Teddy. When GMC was introduced, the area with the black background for the GM was enhanced and replaced with a smoother background.

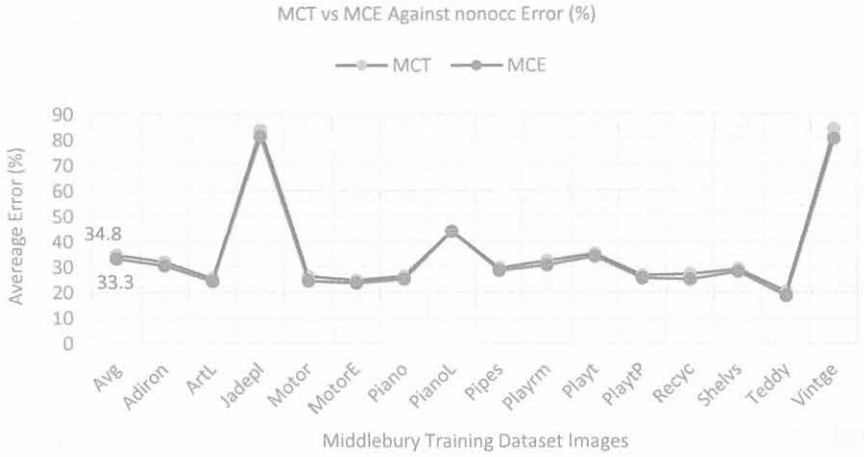
Figure 4.23 subsequently presents the MCT and MCE experimental results. When the MCE was being used instead of the MCT, the average accuracy increased from 42.2% to 40.9% (1.3% change) for *all* error and from 34.8% to 33.5% (1.3% change) for *nonocc* error. Thus, the MCE demonstrated the ability to improve the accuracy and reduce errors. The disparity map between the MCT and MCE for three images: Jadeplant, MotorcycleE, and Vintage, is displayed in Figure 4.24. There were numerous low texture regions in the images of Jadeplant and Vintage, which implied that the texture in the red box region of the MCE disparity map was sharper, and the edge was well preserved compared to the MCT. Radiometric differences between the left and right images were visible in the MotorcycleE image. In general, the texture was smoother in the MCE disparity map than in the MCT disparity map, as shown by the red box region.

Then, the performance of each single matching cost and the multiple matching cost (TAD+GMC+MCE) was analysed as shown in Figure 4.25. The line graph showed the multiple matching cost was producing the highest average accuracy for both *all* error and

nonocc error compared with the single matching cost. The single matching cost accuracy for *all* error was TAD (117%), GMC (64.2%), MCE (51.4%), and for *nonocc* error was TAD (117%), GMC (59.5%), and MCE (45.3%). Meanwhile, the multiple matching cost accuracy was 44.5% for *all* error and 37.4% for *nonocc* error. Hence, this showed a significant improvement in the average accuracy from 6.9% to 72.5% for *all* error and from 7.9% to 79.6% for *nonocc* error. This also clearly indicated the efficiency of combined matching costs, which were capable of improving the accuracy of the SMA especially in the matching cost stage.



(a)



(b)

Figure 4.23: Performance of MCT with MCE (a) *all* errors (b) *nonocc* errors

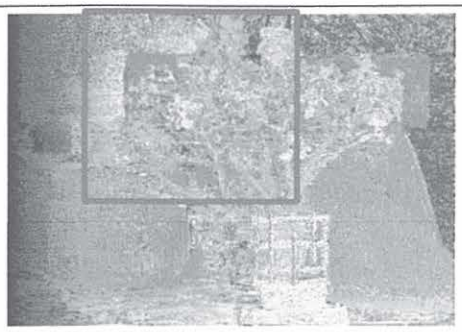
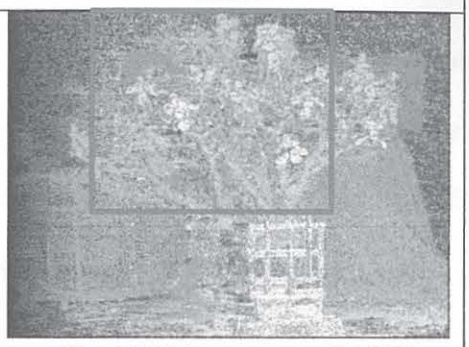
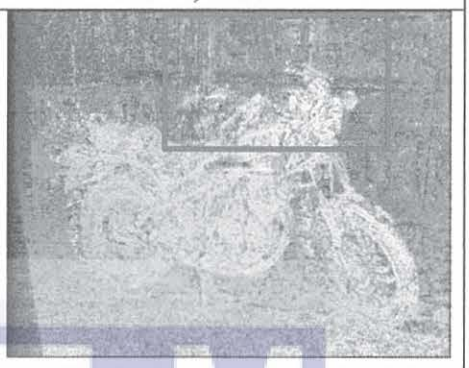
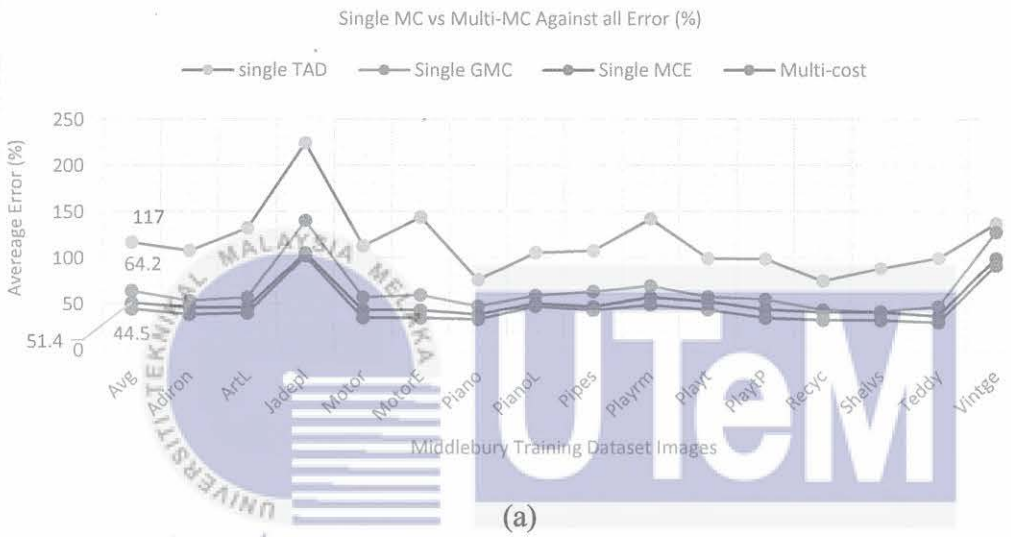
Image	MCT	MCE
Jadeplant	 *all* = 101%, *nonocc* = 83.7%	 *all* = 98.3%, *nonocc* = 81.2%
Motorcycle E	 *all* = 32.5%, *nonocc* = 26.2%	 *all* = 30.9%, *nonocc* = 24.5%
Vintage	 *all* = 87.3%, *nonocc* = 84.2%	 *all* = 83.7%, *nonocc* = 80.5%

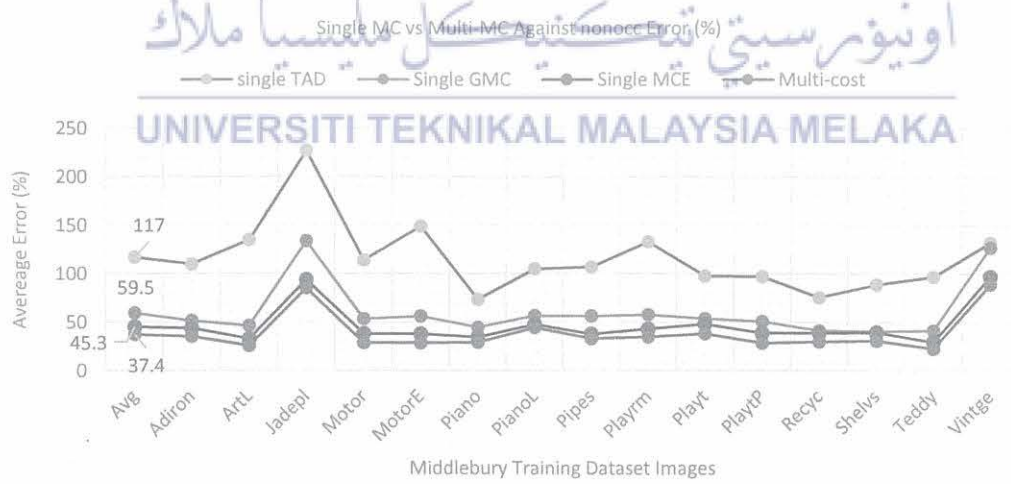
Figure 4.24: Qualitative Performance of MCT with MCE for Middlebury Images Jadeplant, MotorcycleE and Vintage

Figure 4.26 depicts the three disparity maps of Middlebury Jadeplant, MotorcycleE, and PianoL for qualitative performance when comparing single matching cost against multiple matching cost. The TAD single matching cost had the most severe disparity map, followed by GMC and MCE. This was expected due to the fact that the TAD used only per-pixel differences, meanwhile the GMC used directional pixel gradient differences, and the

MCE employed the brightness variation between neighbouring pixels. Despite the fact that the GMC disparity was smoother than TAD, the texture was blurry and there were numerous horizontal streaks in comparison to the MCE, which showed finer texture and boundaries. Therefore, multiple matching costs comprised of the optimum disparity map, which generated a sharper texture by reducing horizontal streaks and preserving edges, as displayed by the region within the red box.



(a)



(b)

Figure 4.25: Performance of Single Matching Cost with Multiple Matching Cost (a) *all* errors (b) *nonocc* errors

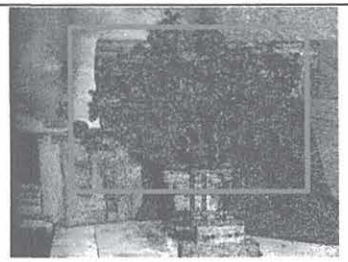
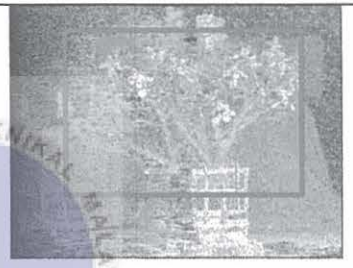
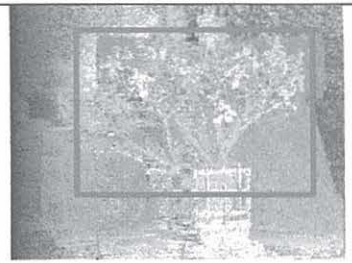
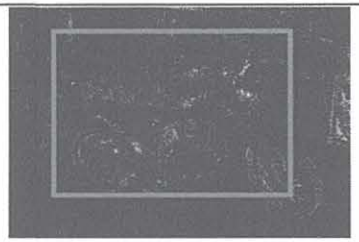
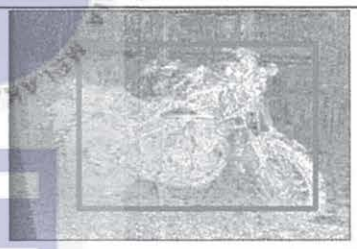
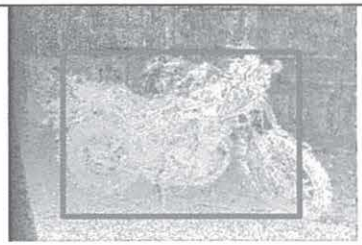
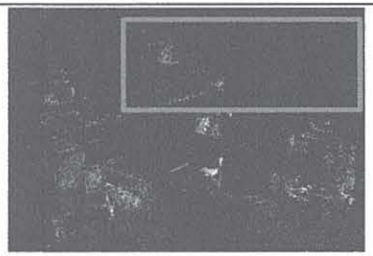
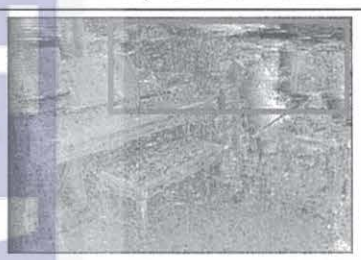
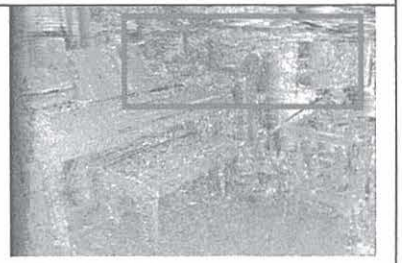
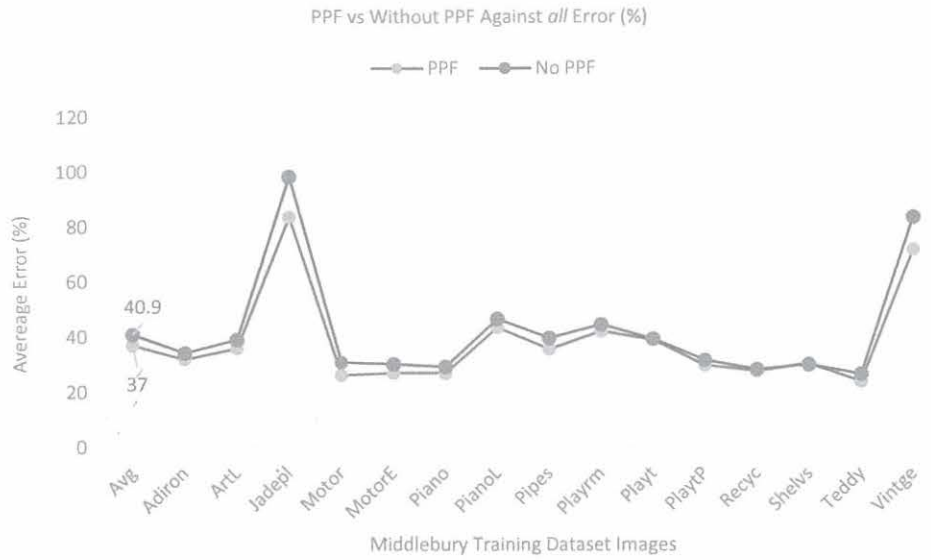
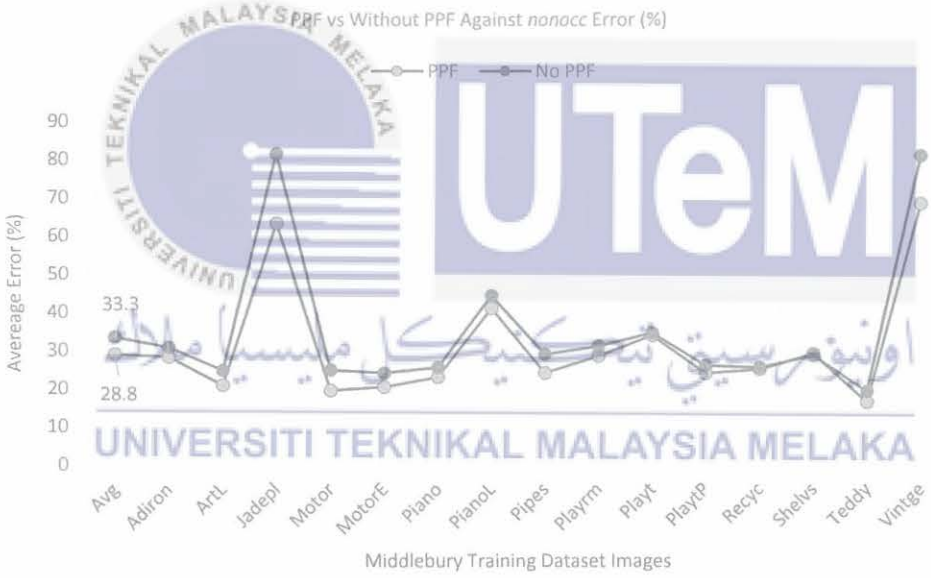
Image	Single MC - TAD	Single MC - GMC	Single MC - MCE	Multi-cost (TAD+GMC+MCE)
Jadeplant	 *all* = 224%, *nonocc* = 227%	 *all* = 140%, *nonocc* = 134%	 *all* = 105%, *nonocc* = 93.9%	 *all* = 38.5%, *nonocc* = 35.3%
MotorcycleE	 *all* = 113%, *nonocc* = 114%	 *all* = 59.6%, *nonocc* = 56.2%	 *all* = 43.4%, *nonocc* = 38.3%	 *all* = 34.8%, *nonocc* = 28.8%
PianoL	 *all* = 106%, *nonocc* = 105%	 *all* = 59.3%, *nonocc* = 57.1%	 *all* = 50.8%, *nonocc* = 48%	 *all* = 47.5%, *nonocc* = 44.8%

Figure 4.26: Qualitative Performance of MCT with MCE for Middlebury Images Jadeplant, MotorcycleE and PianoL



(a)



(b)

Figure 4.27: Performance of PPF and without PPF (a) *all* errors (b) *nonocc* errors

The final performance analysis in the matching cost was the implementation of PPF to combine the multiple costs in the pyramid approach. This was also the implementation of balancing parameter σ_{LT}, σ_{LG}, and σ_{LM} in the PPF. Figure 4.27, which shows the line graph of average accuracy for each image in the Middlebury training dataset. The results showed that there was an improvement in *all* error accuracy from 40.9% to 37.0% (3.9% change)

and *nonocc* error from 33.3% to 28.8% (4.5% change), especially for Jadeplant and Vintage image, which showed a significant improvement. Figure 4.28 is a presentation of the qualitative performance of the PPF consisting of the disparity maps for the Middlebury Jadeplant, Playroom, and Vintage environments.

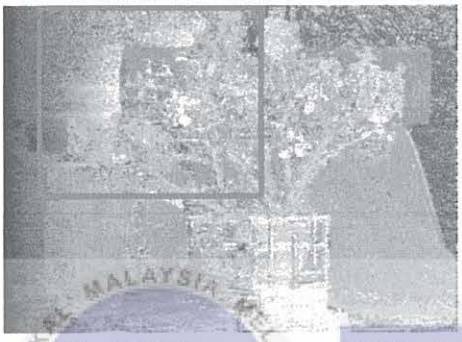
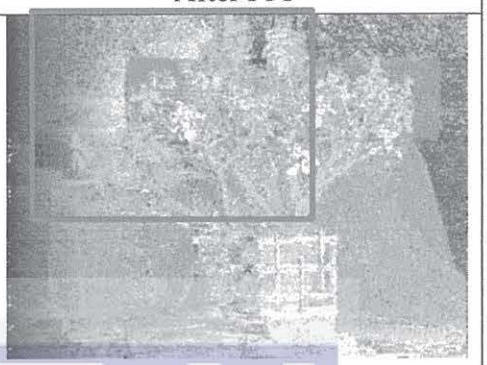
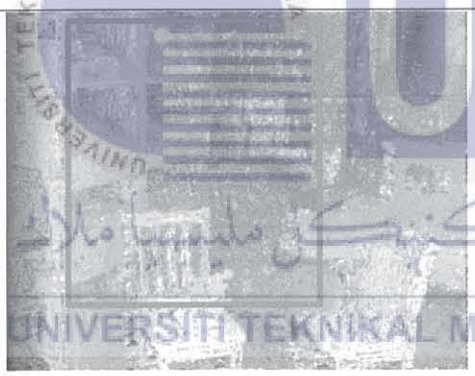
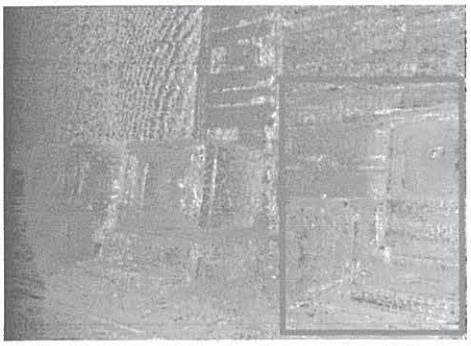
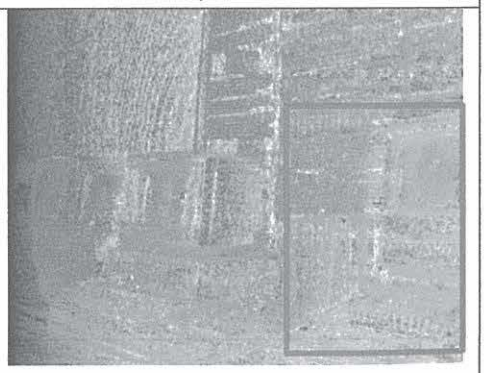
Image	No PPF	After PPF
Jadeplant	 *all* = 98.3%, *nonocc* = 81.2%	 *all* = 83.7%, *nonocc* = 62.9%
Playroom	 *all* = 44.8%, *nonocc* = 30.7%	 *all* = 42.4%, *nonocc* = 28.1%
Vintage	 *all* = 83.7%, *nonocc* = 80.5%	 *all* = 71.9%, *nonocc* = 68.0%

Figure 4.28: Qualitative Performance of PPF and without PPF for Middlebury Images Jadeplant, Playroom and Vintage

Table 4.2: Quantitative Performance *all* errors of GM, GMC, MCT and MCE Matching Cost Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
GM	57.6	48.2	54.4	127	52.8	59.9	41.7	59.4	58	63.2	51.8	49.4	37.8	37.1	41.5	86.4
GMC	56.2	48.2	50.2	131	47.2	46.8	42.3	55.9	54.5	59.8	51	48.7	40.6	37.9	40.3	101
MCT	42.2	35.5	39.9	101	32.5	31.1	30.4	46.9	40.8	46.5	40.8	32.8	30.2	31	28	87.3
MCE	40.9	34.2	39	98.3	30.9	30.2	29.3	46.8	39.9	44.8	39.6	31.8	28.4	30.1	26.7	83.7

Table 4.3: Quantitative Performance *nonocc* errors of GM, GMC, MCT and MCE Matching Cost Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
GM	52.2	45.9	43.6	118	48.6	56.4	38.3	57.3	49.9	51.1	47.3	44.7	35.2	35.7	35	82
GMC	51	45.9	38.8	124	42.7	42.3	39.2	53.8	46.2	47.5	47	44.4	38.2	36.6	34	98.8
MCT	34.8	32	25.5	83.7	26.2	24.7	26.3	44.1	29.6	32.4	35.3	26.7	27.1	29.2	20.3	84.2
MCE	33.3	30.5	24.4	81.2	24.5	23.7	25.2	44	28.6	30.7	34.1	25.6	25.1	28.2	18.8	80.5

Table 4.4: Quantitative Performance *all* errors of Single Cost and Multi-cost Matching Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
Single - TAD	117	108	132	224	113	144	76.3	106	108	142	99.5	98.8	74.8	88.6	99.8	137
Single - GMC	64.2	53.4	56.7	140	57	59.6	47.8	59.3	63.5	69.4	57.6	55	43.7	41.8	47.3	128
Single - MCE	51.4	46.2	46.1	105	43.4	43.2	38.8	50.8	47.1	57.3	52.7	44	41.1	41.1	36.9	99.5
Multi-cost	40.9	34.2	39	98.3	30.9	30.2	29.3	46.8	39.9	44.8	39.6	31.8	28.4	30.1	26.7	83.7

Table 4.5: Quantitative Performance *nonocc* errors of Single Cost and Multi-cost Matching Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
Single - TAD	117	110	135	227	114	149	73.6	105	107	133	97.7	97.1	75.5	88.8	96.8	132
Single - GMC	59.5	51.4	46.3	134	53.3	56.2	44.8	57.1	56.6	58	53.8	51	41.4	40.6	41.4	127
Single - MCE	45.3	43.9	33	93.9	38.3	38.1	34.8	48	37.9	43.1	48.4	39.1	39.1	39.7	29.7	97
Multi-cost	33.3	30.5	24.4	81.2	24.5	23.7	25.2	44	28.6	30.7	34.1	25.6	25.1	28.2	18.8	80.5

Table 4.6: Quantitative Performance *all* errors of PPF and without PPF Based on Middlebury Training Dataset

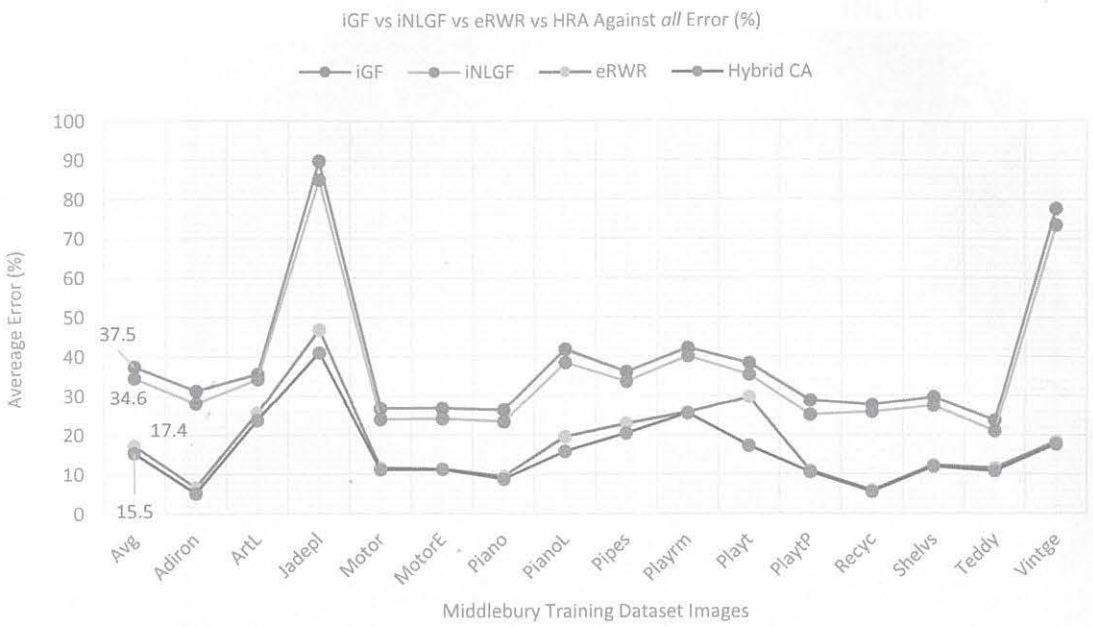
Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
No PPF	40.9	34.2	39	98.3	30.9	30.2	29.3	46.8	39.9	44.8	39.6	31.8	28.4	30.1	26.7	83.7
PPF	37	31.9	35.9	83.7	26.2	27	26.9	43.7	35.7	42.4	39.2	29.9	27.9	30.6	24	71.9

Table 4.7: Quantitative Performance *nonocc* errors of PPF and without PPF Based on Middlebury Training Dataset

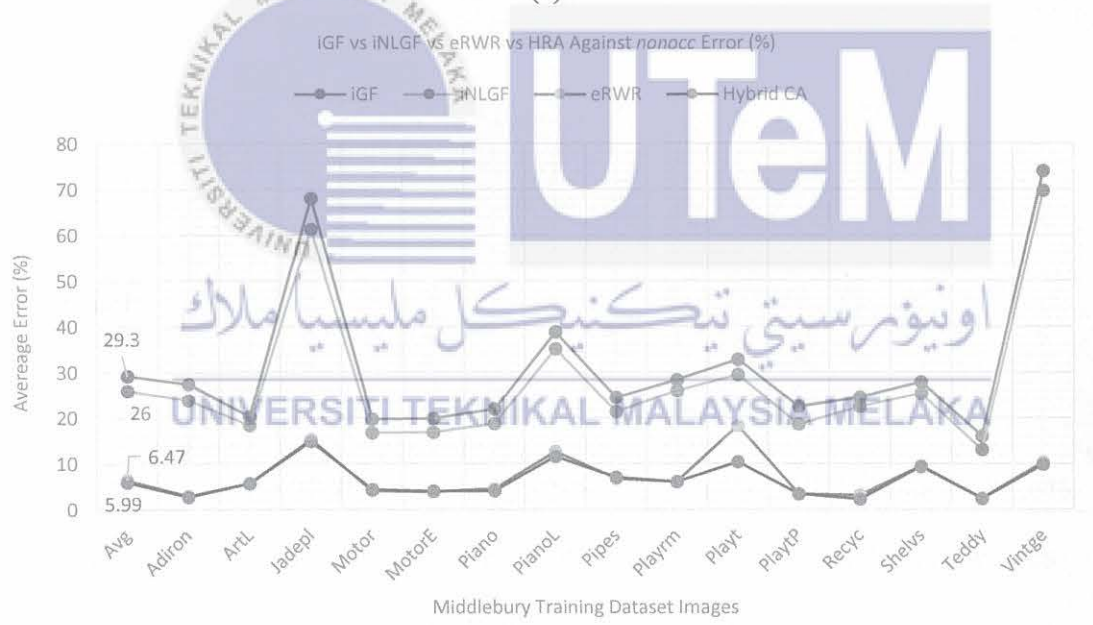
Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
No PPF	33.3	30.5	24.4	81.2	24.5	23.7	25.2	44	28.6	30.7	34.1	25.6	25.1	28.2	18.8	80.5
PPF	28.8	28	20.5	62.9	19.1	20	22.6	40.6	23.7	28.1	33.5	23.5	24.6	28.7	16	68

When compared with the disparity that existed without PPF, the low texture of the leaf of the Jadeplant appeared to be smoother and more visible in the area highlighted by the red box. In the meantime, the PPF disparity map for the Playroom displayed the shelves along the background, and the texture was more apparent, and reduced down the horizontal streaks in the red box region of the Vintage disparity map. The average value accuracy of *all* errors and *nonocc* error is displayed in Tables 4.2 and 4.3 for GM, GMC, MCT, and MCE respectively. Then, Table 4.4 and Table 4.5 display the average value accuracy of all error and nonocc error for both single cost and multiple cost settings, respectively. Following this, Table 4.6 and Table 4.7 summarises the average value accuracy of all error and nonocc error for the PPF implementation in the matching cost stage, respectively.

Stage 2: There were four performance evaluations executed in the cost aggregation to determine the optimal average accuracy for iGF, iNLGF, eRWR, and hybrid random aggregation (HRA; iNLGF + eRWR). Figure 4.29(a) shows the average accuracy result for *all* error between iGF, iNLGF, eRWR, and HRA. The *all* error was at its lowest when the SMA employed only the iGF, which contributed to 37.5% accuracy. However, when the SMA used the new enhanced iNLGF, *all* error was increased to 34.6% with a 2.6% improvement. The algorithm produced *all* error at 17.4% when the algorithm used only eRWR, but when the HRA was employed with a combination of iNLGF and eRWR, *all* error was improved by 1.9%, from 17.4% to 15.5%. There was a significant increase in *all* error from iGF to HRA at 22.0%, down from 37.5% to 15.5%. Meanwhile, Figure 4.29(b) presents the accuracy of *nonocc* error between iGF, iNLGF, eRWR, and HRA. The line graph for *nonocc* error trend was similar to that of *all* error; the lowest *nonocc* error was the iGF at 29.3%, followed by the iNGLF at 26.0%, the eRWR at 6.47%, and the HRA at 5.99%. Furthermore, there was a massive improvement in *nonocc* error from iGF to HRA, resulting in an accuracy difference of 23.31% for the disparity map between iGF and iNLGF.



(a)



(b)

Figure 4.29: Performance of Cost Aggregation for iGF, iNLGF, eRWR and Hybrid Cost Aggregation (iNLGF + eRWR) Based on Middlebury Dataset (a) *all* errors (b) *nonocc* errors

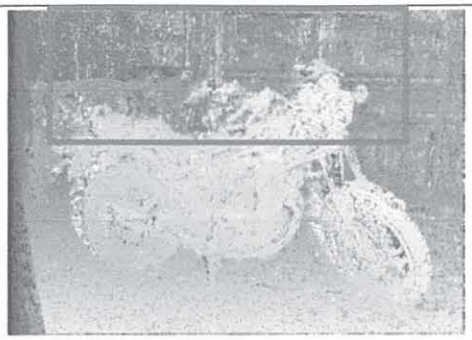
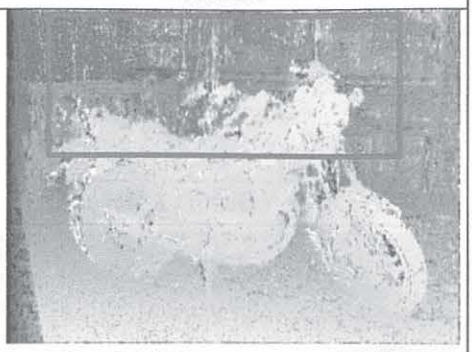
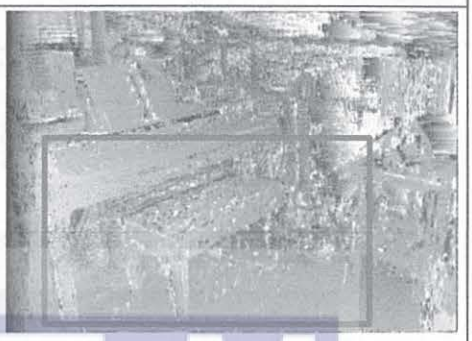
Image	iGF	iNLGF
Motorcycle E	 *all* = 26.8%, *nonocc* = 19.8%	 *all* = 24.2%, *nonocc* = 16.9%
PianoL	 *all* = 41.9%, *nonocc* = 38.9%	 *all* = 38.5%, *nonocc* = 35.2%
PlaytableP	 *all* = 28.8%, *nonocc* = 22.5%	 *all* = 25.2%, *nonocc* = 18.6%

Figure 4.30: Qualitative Performance of iGF and iNLGF for Middlebury Images MotorcycleE, PianoL and PlaytableP

Figure 4.30 displays the Middlebury MotorcycleE, PianoL, and PlaytableP disparity map between iGF and iNLGF. All of these images were influenced by illumination variations and radiometric differences. Based on the result obtained, the disparity map for iNLGF was of better quality (i.e., reduced blur), with the edges and shapes of the MotorcycleE, PianoL, and Playtable were preserved compared with the iGF. The horizontal streaks on the disparity

map; PianoL, in the area of the lamp, were significantly reduced and visibly apparent for qualitative evaluations.

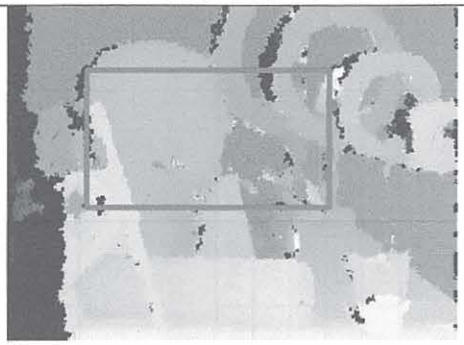
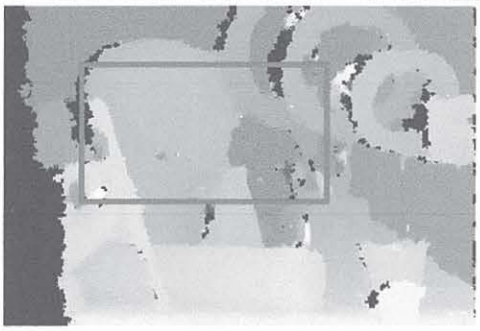
Image	Only eRWR	After HRA (iNLGF + eRWR)
ArtL	 *all* = 25.7%, *nonocc* = 5.73%	 *all* = 23.8%, *nonocc* = 5.72%
PianoL	 *all* = 19.6%, *nonocc* = 12.7%	 *all* = 15.9%, *nonocc* = 11.6%
Playtable	 *all* = 29.7%, *nonocc* = 18.3%	 *all* = 17.3%, *nonocc* = 10.4%

Figure 4.31: Qualitative Performance of eRWR and HRA for Middlebury Images ArtL, PianoL and Playtable

Table 4.8: Quantitative Performance *all* errors of Cost Aggregation Analysis Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
iGF	37.5	31.4	35.6	89.8	26.8	26.8	26.4	41.9	36.2	42.3	38.4	28.8	27.7	29.7	23.7	77.8
iNLGF	34.6	28.1	34.3	85	24	24.2	23.4	38.5	33.7	40.2	35.5	25.2	25.9	27.5	21	73.6
eRWR	17.4	6.7	25.7	46.9	11.7	11.4	9.47	19.6	22.9	25.8	29.7	11	5.99	12.3	11.6	18.3
HRA: iNGF + eRWR	15.5	5.2	23.8	41.1	11.2	11.3	8.87	15.9	20.5	25.6	17.3	10.6	5.61	12	10.8	17.7

Table 4.9: Quantitative Performance *nonocc* errors of Cost Aggregation Analysis Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
iGF	29.3	27.5	20.3	68	19.8	19.9	22	38.9	24.5	28.4	32.8	22.5	24.5	27.9	16.1	74.1
iNLGF	26	23.9	18.4	61.3	16.8	16.9	18.8	35.2	21.4	26	29.5	18.6	22.5	25.4	13.1	69.7
eRWR	6.47	2.96	5.73	15.3	4.11	3.92	4.58	12.7	6.81	6.1	18.3	3.36	3.12	9.52	2.51	10.3
HRA: iNGF + eRWR	5.99	2.8	5.72	14.9	4.42	4.09	4.1	11.6	7.05	6.01	10.4	3.36	2.26	9.39	2.39	9.85

Figure 4.31 examines the qualitative disparity map findings when using only eRWR and HRA. The results were comprised of Middlebury ArtL, PianoL, and Playtable, which exhibited an improvement in areas with low texture, occlusion, and discontinuity. The HRA approach detected the head and circle on the Artl disparity map with much less occlusion than the eRWR approach independently. Compared to the eRWR, the region with illumination variations around the lighting for PianoL had also been corrected and improved. The low texture region in the red box of the PlaytableP disparity had been efficiently recovered, the occlusion had been improved, and the edges had been preserved. Both the eRWR and HRA disparity maps were able to recover and improve the discontinuity regions just as well because both used eRWR. Tables 4.8 and 4.9, respectively, present the average value accuracy of *all* error and *nonocc* error for iGF, iNLGF, eRWR, and HRA.

Stage 3: The minimum cost volume value obtained in the cost aggregation stage provided a solid basis for the selection and optimization of raw data for disparity value at this stage. The average accuracy for *all* errors was 12.7%, and 5.83% for *nonocc* error, as determined by the quantitative evaluation. Similar to Emlek et al. (2018), Q. Li et al. (2018), and Zhou et al. (2019), this technique used the WTA method, which was the best strategy for maximising the disparity selection for the local method. The WTA approach was presented in ***Appendix A***, which used the raw data collected at each location during the cost aggregation stage to determine the minimum number of raw data required to determine the disparity value for stage 4.

Stage 4: Figure 4.32 and Figure 4.33 display the performance evaluation for the disparity refinement. Based on the line graph in Figure 4.32(a), the initial evaluation was the left and right check consistency to detect the invalid pixels in the disparity map. The average accuracy was 5.83% for *nonocc* error and 12.7% for *all* error. Afterwards, the invalid fill-in pixel was applied, replacing any detected invalid pixels from the left and right check

processes. This was accomplished using the median interpolation technique. The average accuracy at this point for *all* error was 12.6%, and for *nonocc* error was 5.82%, shown in Figure 4.32(b). The disparity map's left and right checks for consistency only detected a minor invalid pixel, as both errors only contributed a 0.1% accuracy improvement.

Figure 4.32(c) exhibits the quantitative findings for *all* and *nonocc* errors for Stage 4 using the K-means clustering. The settings and parameters used in the rest of the stages were those described in Section 4.3. There was a significant improvement in average accuracy when the SMA used the K-means clustering to refine the disparity map after invalid pixel fill-in. This can be compared with the results from Figure 4.32(b), which indicates the result of the disparity accuracy without the clustering techniques. The average accuracy was improved by 3.4% fr *all* error from 12.6% to 9.2% and by 0.62% for *nonocc* error from 5.82% to 5.20%. According to the result, the K-means clustering was proven as excellent to decrease the error rate and provide a quality disparity map.

Finally, the SWF was the last strategy that was utilised in the disparity refinement, and contributing significantly to the results obtained which were displayed in Figure 4.32(d). The SWF was tested in order to observe the disparity map's various impacts while preserving the edges and also smoothing them out. When the SWF was implemented, the average accuracy value showed an increase of 0.18%, falling significantly from 9.20% to 9.02% for *all* error, and 0.09%, falling from 5.20% to 5.11% for *nonocc* error. However, the disparity refinement (left and right checks, median interpolation, K-means clustering, and SWF) showed a significant improvement in accuracy by 3.68% for *all* error and 0.72% for *nonocc* error. In summary, the proposed K-means and SWF (Hierarchical Cluster-Edge) in this research was successful to increase the accuracy and reduce almost all errors in all of the Middlebury training images compared to those without clustering and filtering.

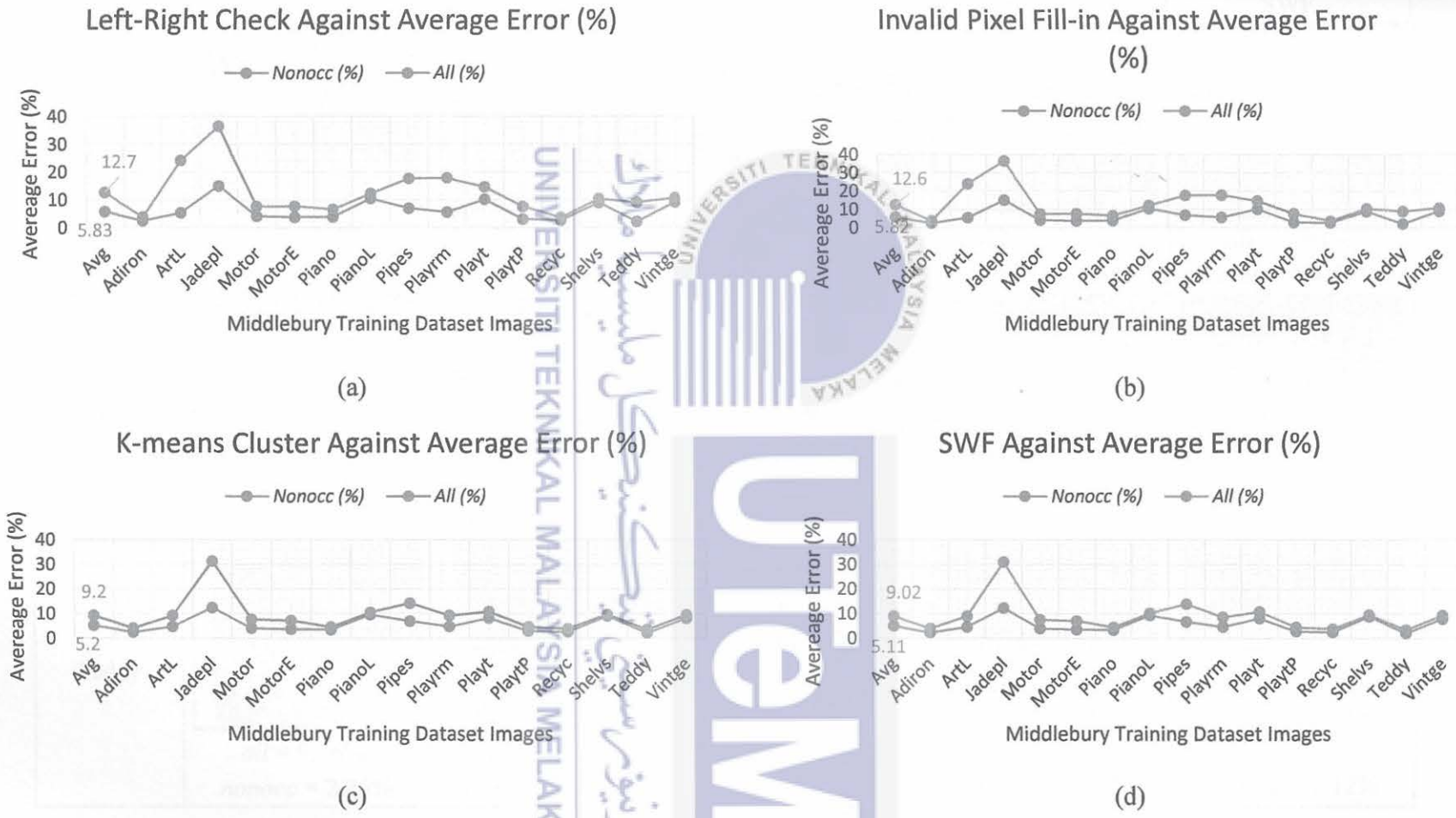


Figure 4.32: Performance Result of Disparity Refinement Step by Step Based on Middlebury Dataset (a) Left-right Check (b) Median Interpolation Invalid Pixel Fill-in (c) K-means Clustering (d) SWF

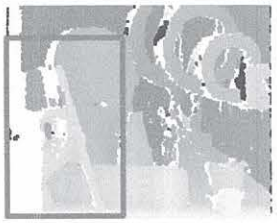
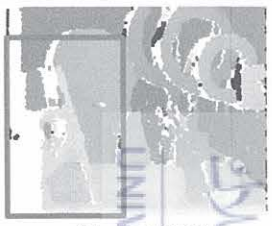
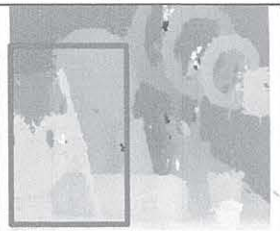
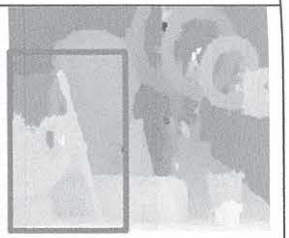
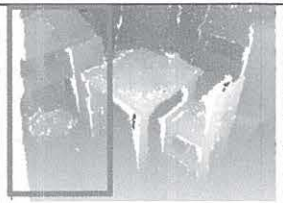
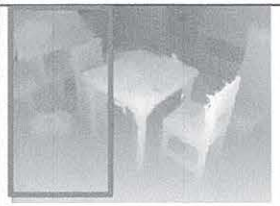
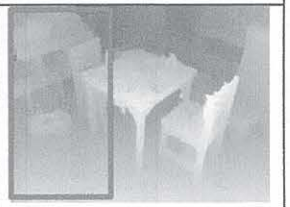
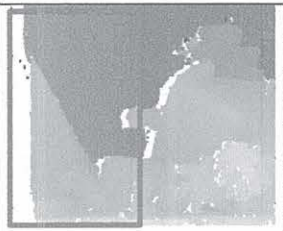
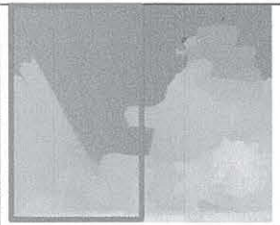
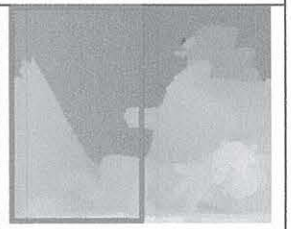
Image	Left-right check	Invalid pixel fill-in	K-means Clustering		SWF
ArtL	 *all* = 24.2%, *nonocc* = 5.53%	 *all* = 24.2%, *nonocc* = 5.51%		 *all* = 9.25%, *nonocc* = 4.78%	 *all* = 8.94%, *nonocc* = 4.69%
PlaytableP	 *all* = 7.71%, *nonocc* = 3.12%	 *all* = 7.70%, *nonocc* = 3.11%		 *all* = 4.63%, *nonocc* = 2.90%	 *all* = 4.51%, *nonocc* = 2.83%
Teddy	 *all* = 9.16%, *nonocc* = 2.26%	 *all* = 9.15%, *nonocc* = 2.24%		 *all* = 3.49%, *nonocc* = 2.15%	 *all* = 3.46%, *nonocc* = 2.12%

Figure 4.33: Qualitative Performance of Disparity Refinement Stage for Middlebury Images ArtL, PlaytableP and Teddy

Figure 4.33 depicts the qualitative performance outcome of three disparity maps during the disparity refinement stage for Middlebury ArtL, PlaytableP, and Teddy. The aims of disparity refinement were to enhance low texture regions, occlusions, horizontal artefacts, texture smoothing, and the preservation of edges. Based on the qualitative result, there were visible occlusions and horizontal artefacts in the disparity maps after the left-right checks and median interpolation invalid pixel fill-in. When the K-means clustering was performed based on the input image, the area of occlusion and horizontal artefacts was greatly improved, as shown in the red box regions. Additionally, the texture and edge of the brush in ArtL disparity and the table in PlaytableP disparity were well preserved, which contributed to a large accuracy improvement of all error by 14.96% for ArtL and 4.21% for PlaytableP. The final implementation of SWF further reduced the horizontal artefacts, smooths the low texture, and improves the sharpness of the edges, which can be seen in the three disparity maps. Tables 4.10 and Table 4.11 present the average accuracy values of *all* error and *nonocc* error for the disparity refinement beginning with the left and right consistency checks, followed by median interpolation invalid pixel fill-in, K-means clustering, and finally the SWF.

In this section of the thesis, error improvements on every stage of algorithm development are explained. Table 4.12 presents the findings of *all* error and *nonocc* error based on the sequential algorithm transformation at every stage. Fundamentally, Stage 1 possessed high error rates when pixel matching was used. Once Stage 2 (HRA) was applied, the number of errors was drastically reduced. In Stage 3, the cost volume was optimised and the disparity value was able to be selected. The remaining noise was then refined and eliminated in Step 4, delivering the final disparity map and average accuracy. Figure 4.34 presents images of the results from every stage.

Table 4.10: Quantitative Performance *all* errors of Disparity Refinement Analysis Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
L-R check	12.7	3.93	24.2	36.5	7.66	7.76	6.6	12.3	17.8	18	14.8	7.71	3.73	10.5	9.16	10.8
L-R check + Invalid pixel fill-in	12.6	3.92	24.2	36.5	7.64	7.74	6.58	12.3	17.8	18	14.8	7.7	3.71	10.5	9.15	10.8
L-R check + Invalid pixel fill-in + K-means clustering	9.2	4.15	9.25	31.3	7.67	7.11	4.58	10.5	14.2	9.16	10.8	4.63	3.87	9.61	3.49	9.53
L-R check + Invalid pixel fill-in + K-means clustering + SWF	9.02	4.01	8.94	30.9	7.59	7.07	4.4	10.2	13.9	8.56	10.8	4.51	3.79	9.54	3.46	9.33

Table 4.11: Quantitative Performance *nonocc* errors of Disparity Refinement Analysis Based on Middlebury Training Dataset

Method	Avg	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge
L-R check	5.83	2.58	5.53	15.1	4.21	3.87	3.95	10.6	7.06	5.68	10.3	3.12	2.78	9.09	2.26	9.23
L-R check + Invalid pixel fill-in	5.82	2.56	5.51	15.1	4.19	3.85	3.92	10.6	7.06	5.65	10.3	3.11	2.75	9.08	2.24	9.21
L-R check + Invalid pixel fill-in + K-means clustering	5.2	2.42	4.78	12.4	4.05	3.62	3.44	9.51	6.93	4.76	8.23	2.9	2.56	8.96	2.15	8.09
L-R check + Invalid pixel fill-in + K-means clustering + SWF	5.11	2.32	4.69	12.3	4	3.6	3.33	9.28	6.76	4.66	8.14	2.83	2.45	8.9	2.12	7.91

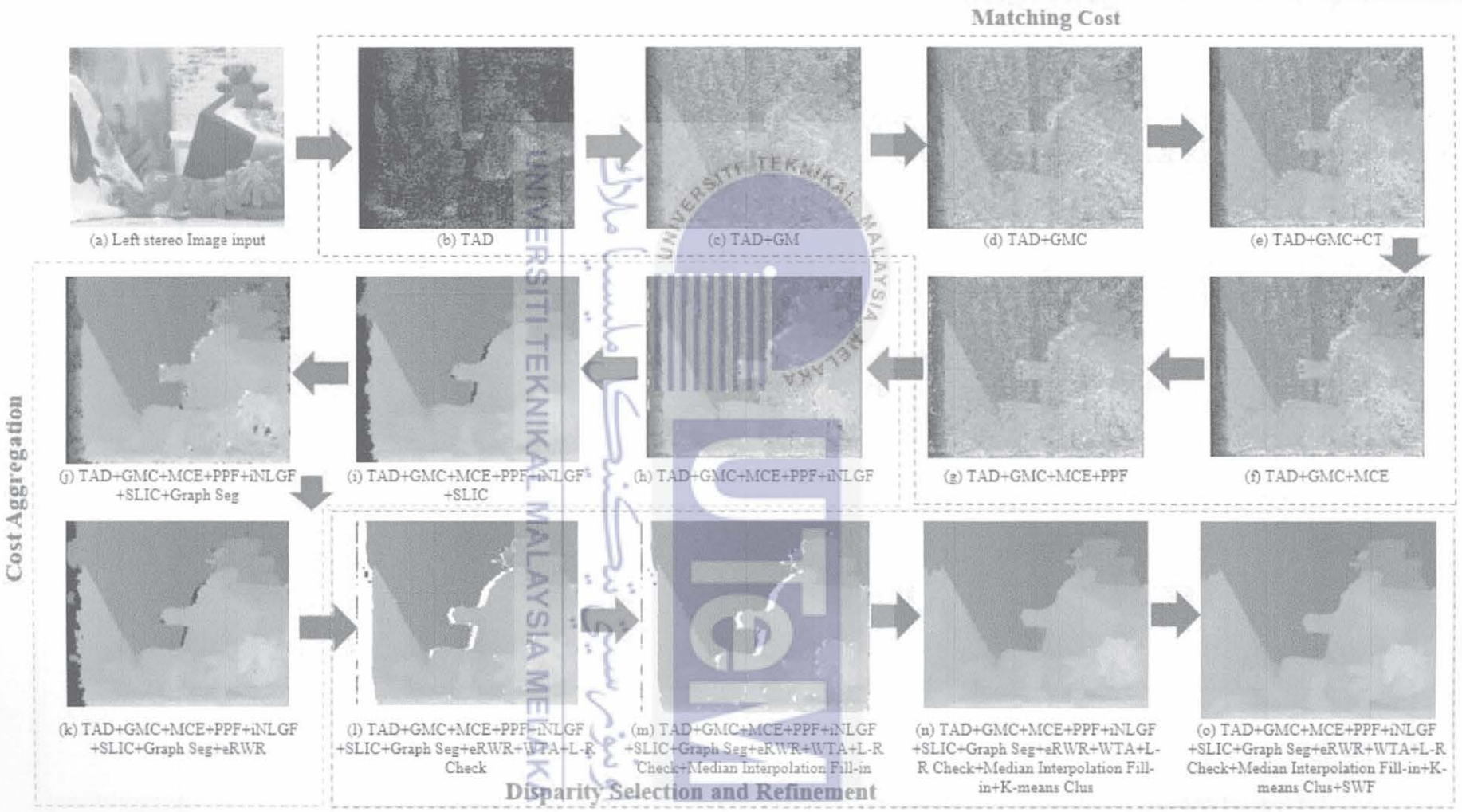


Figure 4.34: Stage-by-Stage SMA Transformation from Matching Cost until Disparity Refinement

Table 4.12: Summary of Stage-by-Stage SMA Improvement

No	Matching Cost	Cost Aggregation	Disparity Selection	Disparity Refinement	*all* errors (%)	*nonocc* errors (%)
a	Left Stereo Image					
b	TAD	None	WTA	None	117.00	117.00
c	TAD + GM	None	WTA	None	57.60	52.20
d	TAD + GMC	None	WTA	None	56.20	51.00
e	TAD + GMC + MCT	None	WTA	None	42.20	34.80
f	TAD + GMC + MCE	None	WTA	None	40.90	33.30
g	TAD + GMC+ MCE + PPF	None	WTA	None	37.00	28.80
h	TAD + GMC+ MCE + PPF	iNLGF	WTA	None	34.60	26.00
i	TAD + GMC+ MCE + PPF	iNLGF + SLIC	WTA	None	16.30	16.90
j	TAD + GMC+ MCE + PPF	iNLGF + SLIC + Graph Seg	WTA	None	19.20	9.95
k	TAD + GMC+ MCE + PPF	iNLGF + SLIC + Graph Seg + eRWR	WTA	None	15.50	5.99
l	TAD + GMC+ MCE + PPF	iNLGF + SLIC + Graph Seg + eRWR	WTA	L-R Check	12.70	5.83
m	TAD + GMC+ MCE + PPF	iNLGF + SLIC + Graph Seg + eRWR	WTA	L-R Check + Median Interpolation Fill-in	12.60	5.82
n	TAD + GMC+ MCE + PPF	iNLGF + SLIC + Graph Seg + eRWR	WTA	L-R Check + Median Interpolation Fill-in + K-means Clus	9.20	5.20
o	TAD + GMC+ MCE + PPF	iNLGF + SLIC + Graph Seg + eRWR	WTA	L-R Check + Median Interpolation Fill-in + K-means Clus + SWF	9.02	5.11

Table 4.12 shows the performance of each stage in the proposed algorithm based on the quantitative measurement of the Middlebury Stereo Benchmarking. When generating the initial disparity maps, the single TAD algorithm generated a lot of noise, especially the salt and pepper noises, and horizontal artifacts, which contributed to the high value of *all* and *nonocc* errors. However, the combination of TAD and GM was able to reduce the noises by 59.4%, and they were further reduced by 1.4% when the GM was replaced with GMC. The texture was improved but still blurry, and there were numerous horizontal streaks. The matching cost was then combined with the MCE and the PPF, which contributed significantly to noise reduction by 21.6%. The implementation of MCE contributed to finer texture and boundaries. Consequently, the optimum disparity map was composed of several matching costs which delivered a sharper texture by minimising the horizontal streaks and preserving the edges.

When stage 2 was introduced, the noises were significantly reduced even further. This demonstrated the significance of stage 2 to the proposed algorithm. The implementation of iNLGF contributed to a reduction of noise of 2.4% for *all* error, while still preserving the edges and shapes. The iNLGF was also resistant to radiometric variations and horizontal streaks. The combination of iNLGF and eRWR under HRA further reduced the noise in the disparity map by 19.1% for *all* error and 20.01% for *nonocc* error. The findings clearly showed that the combination of the aggregation methods of iNLGF and eRWR performed better compared with the application using only filter-based aggregation. Concisely, the HRA was successful to correct the illumination variations, recovered the low texture regions and improved the occlusion. In addition, the HRA was also able to recover and improve the discontinuity regions.

The noise was further reduced at stage 4, when the left-right consistency check and median interpolation for pixel fill were employed. There was an improvement in accuracy

of 2.9% for *all* error and 0.17% for *nonocc* error. A hierarchical cluster-edge method based on K-means clustering and SWF was added, resulting in a noise reduction of 3.8% for *all* error and 0.7% for *nonocc* error. This resulted in producing a hierarchical cluster-edge method which improved the area of occlusion and horizontal artefacts, preserved the texture and edge, reduced the horizontal artefacts, and smoothed the low texture. As a result, the cooperation of all stages showed 5.11% for *nonocc* error and *9.02%* for all error.

4.3.2 Standard Benchmarking Dataset and Real Stereo Images Performances

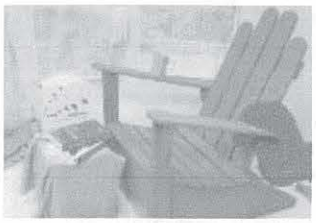
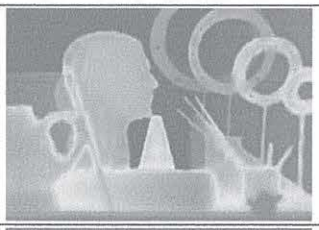
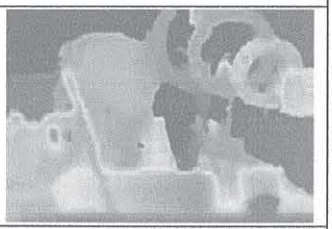
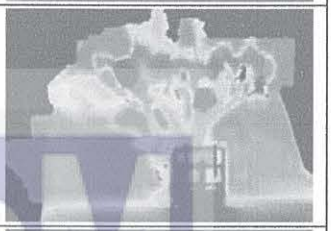
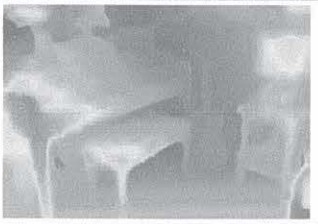
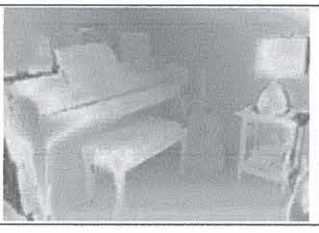
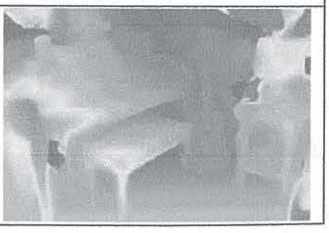
The capability of the proposed work in this research is described and validated in this section. Three databases were employed, as stated earlier, to demonstrate the adaptability of the proposed algorithm. The performance of the proposed algorithm was evaluated using the Middlebury datasets (i.e., training and testing images), the KITTI datasets (i.e., training and testing images), and the UTeMLab-Stereo images. The results from these databases were compared and discussed with several established methods accordingly.

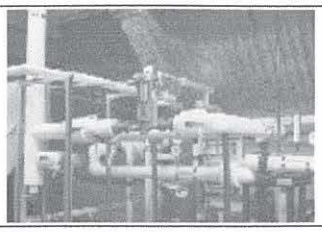
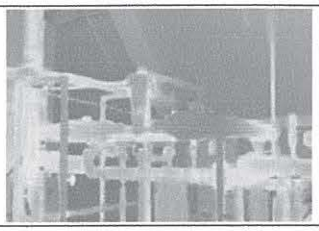
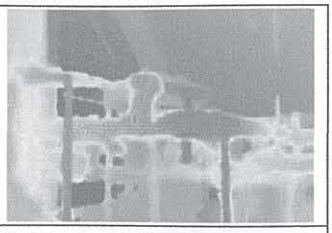
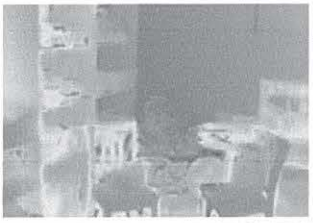
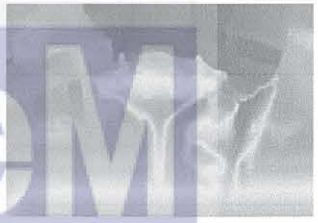
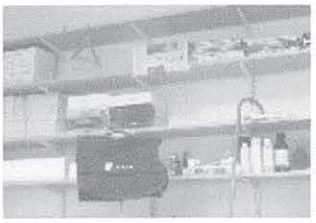
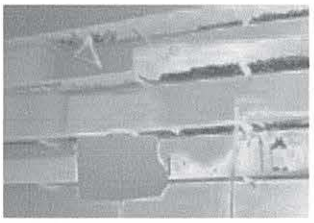
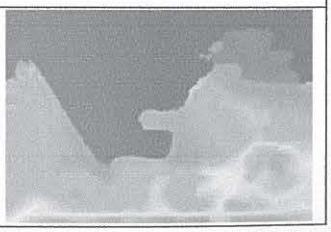
4.3.2.1 Middlebury Dataset

In this thesis, the performance of the SMA was evaluated qualitatively based on the Middlebury training and test datasets, as shown in Figures 4.35 and 4.36. As shown in Tables 4.13 and 4.14, the SMA was also compared with other established SMA quantitative performances in descending order. The other SMA approach was comprised of many categories, including local, global, semiglobal, and machine learning, as described in Chapter 2. The input images, resolution, disparity range, ground truth, and disparity maps, which comprise 15 datasets in the training and test of the Middlebury Stereo dataset, were presented in Figures 4.35 and 4.36. All the input images in the dataset had distinctive characteristics and resolutions. The characteristics and resolutions of the final results were

influenced by the disparity maps that were produced. These characteristics depended on the input images, such as the Jadeplant, Motorcycle, Playroom, Piano, and Shelves, which were utilised for complex scene artifacts. Meanwhile, the Adirondack, Recycle, and Vintage images were used to distinguish foreground objects from background objects. The images of the Motorcycle, Pipes, and Teddy were deployed to challenge the regions of depth discontinuity. The MotorcycleE and PianoL were used to evaluate the performance in terms of variation in illumination and radiometric differences.

Figure 4.35 displays the strong disparity maps of the SMA, which exhibit the contours and disparity regions smoothly and closely with the ground-truth disparity map. Large low texture regions were observed in The Piano, Playtable, and Recycle as well. On the contrary, the proposed SMA showed outstanding disparity maps result. The foreground objects were clearly distinguished from the background with exact contours and precise disparity values in accordance with respective to the depth order, based on the outcomes of the proposed SMA for the Adirondack, Jadeplant, Motorcycle, Playtable, and Recycle images. Further, the disparity map for images containing complex scene features, such as Jadeplant, Playroom, Pipes, and Shelves were reconstructed based on their respective depths and provides acceptable disparity in depth discontinuity regions. The images containing illumination variations and radiometric differences, for example, the MotorcycleE, Piano, and PianoL also produced as smooth disparity maps with detailed and clear contours. The Middlebury test dataset for disparity map outcomes is displayed in Figure 4.36. For the images of Australia, AustraliaP, Bicycle2, Classroom2, Classroom2E, Computer, Djembe, DjembeL, Hoops, Livingroom, Newkuba, Plants, and Stairs, smooth disparity maps were produced and reconstructed based on their respective depths. The performance of the proposed SMA in the two Middlebury datasets indicated the achievement of the effort to develop a precise local SMA for disparity mapping.

Image Res (D_{max})	Left image	Ground truth	Result
Adirondack 718 x 49 (73)			
ArtL 347 x 277 (64)			
Jadeplant 659 x 497 (160)			
Motorcycle 741 x 497 (70)			
MotorcycleE 741 x 497 (70)			
Piano 707 x 481 (65)			
PianoL 707 x 481 (65)			

Pipes 735 x 485 (75)			
Playroom 699 x 476 (83)			
Playtable 699 x 476 (83)			
PlaytableP 699 x 476 (83)			
Recyle 720 x 486 (65)			
Shelves 738 x 497 (60)			
Teddy 450 x 375 (64)			

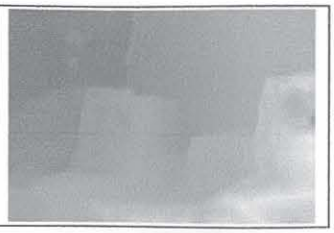
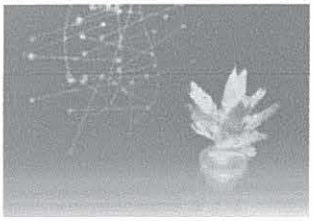
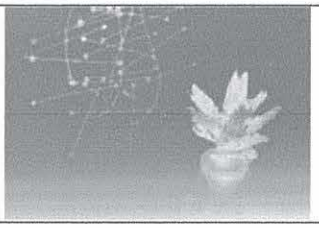
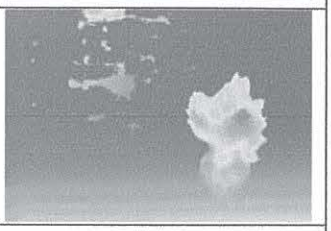
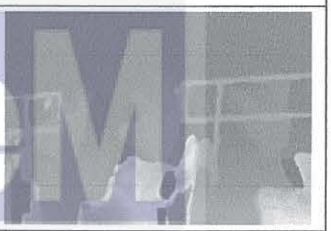
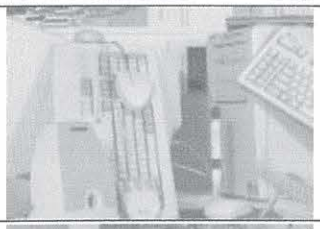
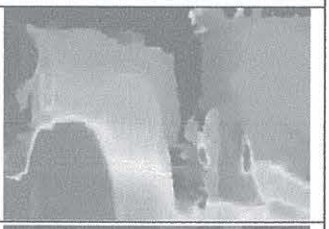
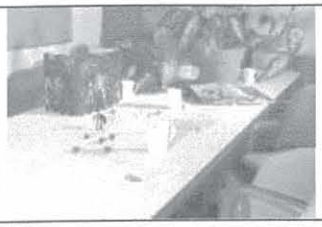
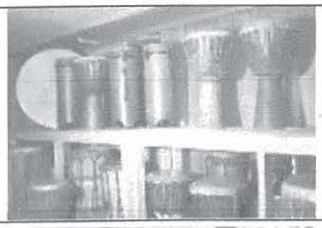
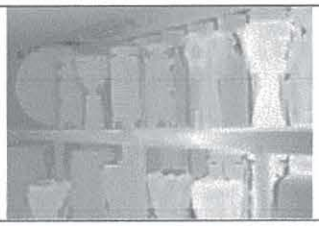
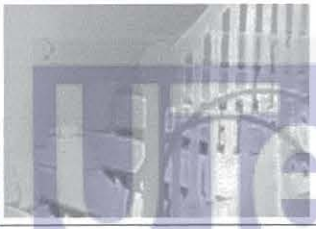
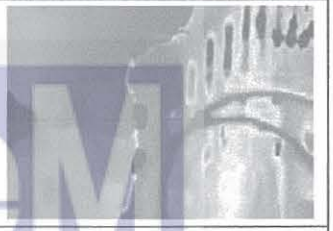
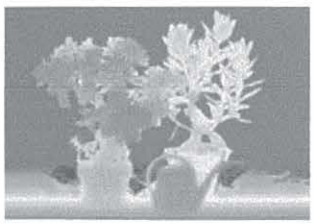
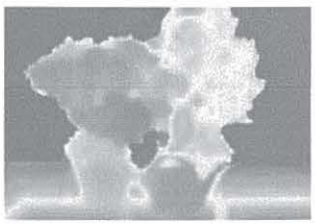
Vintage 722 x 480 (190)			

Figure 4.35: The Result of the Middlebury Training Dataset

Image Res (D_{max})	Left image	Ground truth	Result
Australia 715 x 492 (73)			
AustraliaP 715 x 492 (73)			
Bicycle2 713 x 488 (63)			
Classroom2 750 x 474 (153)			
Classroom2E 750 x 474 (153)			
Computer 322 x 277 (64)			
Crusade 720 x 474 (200)			

CrusadeP 720 x 474 (200)			
Djembe 719 x 494 (80)			
DjembeL 719 x 494 (80)			
Hoops 721 x 498 (103)			
Livingroom 742 x 496 (80)			
Newkuba 701 x 487 (143)			
Plants 710 x 496 (80)			

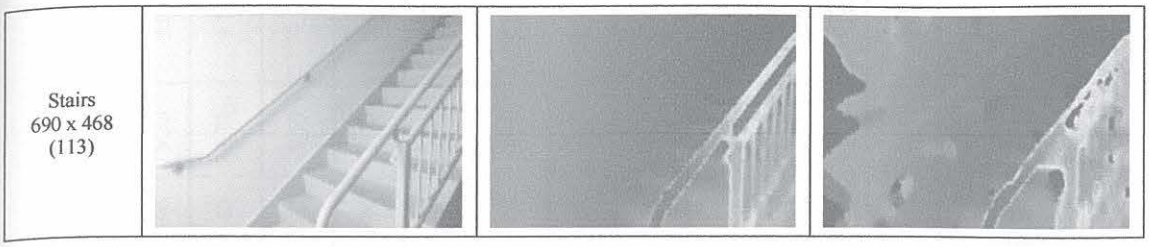


Figure 4.36: The Result of the Middlebury Test Dataset

Furthermore, the performance of the proposed SMA was quantitatively compared to those of other well-established methods, such as the local, global, semiglobal, and machine learning methods. Table 4.13 presents the Middlebury training dataset's qualitative performance based on *all* error, whereas Table 4.14 provides the *nonocc* error. Tables 4.15 and 4.16 exhibit the qualitative performance of the Middlebury test dataset based on *all* error and *nonocc* error. The local method comparison consisted of the Adaptive Cross-region based Guided Image Filtering (ACR-GIF-OW), Two Phase Optimisation AD-Census and Gradient Fusion (ADSG), Multi-Block Matching (MBM), Efficient Large-scale Stereo Matching (ELAS_RVC), Recursive Edge-Aware Filters Aggregation (REAF) and Shift Adapted Weighted Aggregation and variational completion (DAWA-F).

The global methods comparison include Two-step Energy Minimization MRF (TSGO), Top-down Cues Stereo Reconstruction (HLSC cor), DP and MRF Multiple Disparity Proposal (MDP). Additionally, SGM Precomputed Surface Orientation Priors (SGMEPi) and State-of-the-Art SGM (SGM RVC) were the sources of the semiglobal method. As for the machine learning algorithm, the methods chosen include the Domain Transform Solver (DTS), Cascaded Multi-scale and Multi-dimension CNN (MSMD ROB), Dense-CNN, Coalesced Bidirectional Matching Volume (CBMV), Cascaded Regression and Adaptive Refinement (CRAR), and End-to-end Hybrid CNN-CRF (JMR).

Table 4.13: Quantitative Performance of Middlebury Training Dataset Based on *all* error %

Algorithm	Method	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge	Weight ave
Proposed SMA	Local	4.01	8.94	30.9	7.59	7.07	4.4	10.2	13.9	8.56	10.8	4.51	3.79	9.54	3.46	9.33	9.02
DTS (Bapat and Frahm, 2019)	ML	2.67	18.2	20.3	6.1	6.14	5.24	7.66	11.7	16.6	8.12	7.69	3.52	4.53	9.62	8.71	9.12
MSMD_ROB (Lu et al., 2018a)	ML	2.85	8.58	45.1	5.12	4.99	3.75	7.18	11	6.86	9.74	9.32	2.74	3.56	3.02	9.59	9.2
ACR-GIF-OW (Kong et al., 2021)	Local	4.53	8.41	22.1	7.93	7.88	6.36	27.7	11	8.51	16.1	6.6	4.26	13.1	2.86	7.77	9.48
JMR (Knöbelreiter et al., 2017)	ML	2.17	18	24.7	5.98	6.9	6.14	7.27	11	17.5	8.18	7.44	2.96	7.81	8.98	10.3	9.57
HLSC_cor (Hadfield et al., 2017)	Global	3.35	9.7	35	6.85	6.87	3.92	7.3	13.8	10.1	16.6	3.9	3.55	11.7	2.99	14.6	9.61
ADSG (Liu et al., 2021)	Local	4.91	9.46	27.9	6.25	6.3	6.59	17.4	12.8	12.4	20.9	6.37	4.55	11.1	4.01	9.4	9.98
MBM (Chang and Maruyama, 2018)	Local	4.39	8.8	37.6	5.76	5.56	6.67	12.4	11.8	12.9	12	6.37	3.67	11.8	3.74	14.1	10.1
TSGO (Mozerov and Van De Weijer, 2015)	Global	2.41	4.71	55.6	5.02	4.77	3.5	16.6	8.88	5.69	20.7	2.95	2.66	8.86	2.88	13.5	10.1
DAWA-F (Navarro and Buades, 2019)	Local	4.37	13	44.4	7.29	7.04	3.27	21.7	15.9	8.86	6.39	3.34	2.89	11.1	3.93	6.48	10.6
MDP (Li et al., 2016)	Global	1.56	7.37	53.8	5.89	6.18	4.04	8.81	14.2	11.0	15.8	4.19	4.00	9.24	3.95	15.2	10.8
SGMEPi (Scharstein et al., 2018)	SGM	5.65	18.2	30.8	9.18	9.02	8.49	14.7	15.8	21.0	10.7	9.76	5.80	11.0	10.7	31.9	13.4

Table 4.14: Quantitative Performance of Middlebury Training Dataset Based on *nonocc* error %

Algorithm	Method	Adiron	ArtL	Jadepl	Motor	MotorE	Piano	PianoL	Pipes	Playrm	Playt	PlaytP	Recyc	Shelvs	Teddy	Vintge	Weight ave
JMR (Knöbelreiter et al., 2017)	ML	0.92	2.18	6.01	1.26	1.27	2.21	4.03	2.12	1.94	2.2	1.65	1.3	5.51	1.15	3.73	2.30
DTS (Bapat and Frahm, 2019)	ML	1.1	3.25	10.6	1.5	1.5	2.07	4.55	3.05	3.5	2.51	2.33	2.08	3.32	1.08	4.43	3.02
SGMEPi (Scharstein et al., 2018)	SGM	1.72	3.36	9.72	1.79	1.87	3.20	9.97	3.66	3.48	2.70	2.64	2.09	8.04	1.75	26.4	4.57
MDP (Li et al., 2016)	Global	1.21	3.24	14.9	2.33	2.36	2.53	7.6	5.04	3.46	11.2	2.17	2.2	8.2	1.97	12.3	4.75
Proposed SMA	Local	2.32	4.69	12.3	4	3.6	3.33	9.28	6.76	4.66	8.14	2.83	2.45	8.9	2.12	7.91	5.11
ADSG (Liu et al., 2021)	Local	2.35	4.97	11.4	2.69	2.49	5.02	16.2	5.08	5.07	16.9	3.13	2.47	10.1	2.11	7.17	5.55
ACR-GIF-OW (Kong et al., 2021)	Local	3.01	3.91	11.2	2.81	2.91	4.95	27.1	4.59	5.49	12.3	2.58	2.5	12.6	1.86	6.58	5.78
MBM (Chang and Maruyama, 2018)	Local	3.11	5.05	15.7	3.07	2.93	5.56	11.5	5.85	6.93	10.3	4.81	3.02	11.6	2.52	13.3	6.28
HLSC_cor (Hadfield et al., 2017)	Global	2.46	5.41	18.1	4.22	4.23	3.43	6.89	7.48	6.49	14.6	2.77	3.16	11.3	1.99	13.8	6.38
MSMD_ROB (Lu et al., 2018a)	ML	2.5	5.5	30	3.42	3.35	3.2	6.86	6.4	4.11	7.09	6.74	2.4	3.26	2.3	8.62	6.46
DAWA-F (Navarro and Buades, 2019)	Local	2.56	5.27	27.3	3.41	3.29	2.45	21.3	7.73	7.08	2.88	2.03	1.85	10.8	1.78	4.68	6.48
TSGO (Mozerov and Van De Weijer, 2015)	Global	2.02	3.07	32.5	3.12	2.94	2.96	16.1	4.90	4.02	18.7	2.20	2.33	8.34	2.46	12.6	7.07

Table 4.15: Quantitative Performance of Middlebury Test Dataset Based on *all* error %

Algorithm	Method	Austr	AustrP	Bicyc2	Class	ClassE	Compu	Crusa	CrusaP	Djemb	DjembL	Hoops	Livgrm	Nkuba	Plants	Stairs	Weight ave
Proposed SMA	Local	11.2	7.43	8.8	18	22.6	12.2	15.7	16.4	2.91	11	18.6	6.74	12.4	20.7	20.7	13.1
ELAS_RVC (Geiger, et al., 2011)	Local	11.7	7.81	6.2	17.3	51.9	10.7	9.27	9.39	2.45	15.1	19.9	9.33	13.1	26.9	12.5	13.4
MDP (Li et al., 2016)	Global	8.08	5.63	6.62	13.2	21.3	11.5	19.8	23.4	2.53	13.4	24.7	9.68	12	19.1	26.3	13.6
SGM_RVC (Hirschmüller, 2008)	SGM	11.1	6.09	5.97	17.4	22.1	17.2	25	21.8	2.31	11.4	13.3	15.8	12.2	18.6	13.4	14.2
CBMV (Batsos et al., 2018)	ML	7.64	7.28	6.12	18.6	32.7	12	20.8	20.8	2.62	21.2	18.9	8.83	16.2	18.6	16.4	14.4
DTS (Bapat and Frahm, 2019)	ML	7.69	7.4	5.51	17.5	27.9	14.5	33.3	34.8	2.66	5.81	14.2	11.9	10.4	10.3	13.9	14.6
CRAR (Zeng and Tian, 2022)	ML	6.54	5.91	4.13	9.99	10.7	7.79	50.5	56.8	1.84	3.98	10.8	7.32	8.17	9.16	12.9	14.7
REAF (Çiıla, 2015)	Local	18.6	11.4	7.1	35.6	54.2	10.6	12.1	10.3	2.11	18.8	16.3	7.51	13.2	16.2	15.1	15.0
ACR-GIF-OW (Kong et al., 2021)	Local	11.8	7.95	8.59	11	45.1	13.1	27	26.8	2.24	17.3	23.1	9.18	10.7	16.5	19.2	15.3
ADSG (Liu et al., 2021)	Local	12.6	8.84	9.51	18.3	28.8	13.3	23.2	24.7	2.63	18.7	21.5	9.29	14.3	18.4	23.1	15.6
JMR (Knöbelreiter et al., 2017)	ML	6.72	6.87	6.58	13.2	13.1	13.8	40.3	41.4	2.88	5.81	18	9.67	15.3	14.5	19.5	15.7
DAWA-F (Navarro and Buades, 2019)	Local	9.58	6.51	5.29	13.9	31.4	19.4	30.5	39.1	1.90	13.3	19.7	15.6	10.7	21.8	22.0	17.0
Dense-CNN (Zhang et al., 2021)	ML	8.20	7.92	6.09	22.7	33.2	14.2	35.8	36.3	3.05	10.3	21.3	12.9	16.7	13.1	17.6	17.1

Table 4.16: Quantitative Performance of Middlebury Test Dataset Based on *nonocc* error %

Algorithm	Method	Austr	AustrP	Bicyc2	Class	ClassE	Compu	Crusa	CrusaP	Djemb	DjembL	Hoops	Livgrm	Nkuba	Plants	Stairs	Weight ave
JMR (Knöbelreiter et al., 2017)	ML	2.66	2.67	2.42	1.63	1.93	3.42	2.7	2.59	1.02	3.98	3.86	3.99	4.44	3.91	5.5	3.02
Dense-CNN (Zhang et al., 2021)	ML	4.03	3.63	1.7	2.92	13.9	1.64	2.44	1.93	0.96	7.96	3.98	3.57	5.56	3.13	5.92	3.63
DTS (Bapat and Frahm, 2019)	ML	3.82	3.49	1.71	3.06	13.7	2.45	5.34	5.54	1.05	4.07	3.31	3.15	3.28	3.33	4.93	3.78
CBMV (Batsos et al., 2018)	ML	4.27	4.04	2.87	2.3	17	2.88	2.6	2.33	0.91	20.1	4.21	3.57	5.01	6.96	5.27	4.71
MDP (Li et al., 2016)	Global	5.08	3.32	4.07	6.76	12.9	2.17	4.43	3.39	1.2	12.7	8	3.26	6.41	6.54	10.1	5.28
SGM_RVC (Hirschmüller, 2008)	SGM	8.75	3.88	3.03	5.95	10.4	4.69	5.45	4.51	1.11	9.94	4.91	6.7	5.95	8.64	7.57	5.66
Proposed SMA	Local	8.91	5.58	6.15	4.93	9.21	5.5	3.82	3.52	1.97	10.2	7.06	4.46	7.21	10.8	11.9	6.21
ADSG (Liu et al., 2021)	Local	9.2	5.71	5.93	6.21	17.6	3.85	4.09	3.43	1.63	17.8	10	6.35	6.98	8.84	11.8	6.90
DAWA-F (Navarro and Buades, 2019)	Local	7.61	4.49	3.36	3.57	21.6	4.39	4.66	13	1.1	12	10.9	6.53	6.97	12.3	9.81	7.30
ACR-GIF-OW (Kong et al., 2021)	Local	7.62	4.99	6.58	4.29	35.5	3.57	6.19	5.49	1.44	17.1	13.7	8.01	6.59	7.86	13.5	7.90
CRAR (Zeng and Tian, 2022)	ML	5.19	4.55	3.13	6.11	6.40	3.33	28.9	32.7	1.15	3.42	6.11	3.87	6.23	3.91	6.78	8.63
ELAS_RVC (Geiger et al., 2011)	Local	9.87	6.00	4.18	12.1	47.8	4.31	4.91	4.51	1.59	14.9	14.0	6.49	9.41	18.3	8.13	9.52
REAF (Çiıla, 2015)	Local	17.0	9.81	4.32	29.6	46.1	4.62	6.23	4.35	1.39	18.6	8.79	6.51	8.40	7.59	10.4	10.7

The proposed SMA was ranked highly when compared to the other methods used according to the evaluation tables of all error for both training and test datasets. The proposed SMA produced an *all* error rate of 13.10% for the test and 9.02% for the training dataset. The closest accuracy for local was the ACR-GIF-OW at 9.48% for training and the ELAS_RVC at 13.4% for test. The highest accuracy for the global method was the HLSC_cor at 9.61% for training and for the MDP at 13.6% for test, while the semiglobal had the highest accuracy at 13.4% for training and at 14.2% for the test, contributed by SGMEPi and SGM_RVC. The ML with the best performance was the DTS with a training accuracy of 9.12% and CBMV with a test accuracy of 14.4%, which was still less accurate than the proposed SMA. Contrary, the *nonocc* error evaluation tables for both training and test datasets indicated the JMR, DTS, SGMEPi, MDP, Dense-CNN, CBMV, and SGM RVC were the only methods more accurate than the proposed SMA. However, the proposed new SMA remained among the most accurate. Figure 4.37 compares the performance of disparity maps with several methods, including ACR-GIF-OW, JMR, HLSC corr, and SGMEPi. The results displayed a smooth disparity map with clear and detailed contours for the proposed SMA in comparison to other method. Hence, the Middlebury datasets indicated the proposed SMA outperformed numerous established standard methods on both qualitative and quantitative accuracy.

Appendix B presents the detail of Middlebury dataset results including the input images, disparity maps, ground truths, error maps for training, test, and mobile datasets.

4.3.2.2 KITTI Dataset

The outdoor KITTI dataset was employed in the experimentation with the proposed SMA as well. The purpose of this experiment was to further evaluate the accuracy of the proposed SMA.

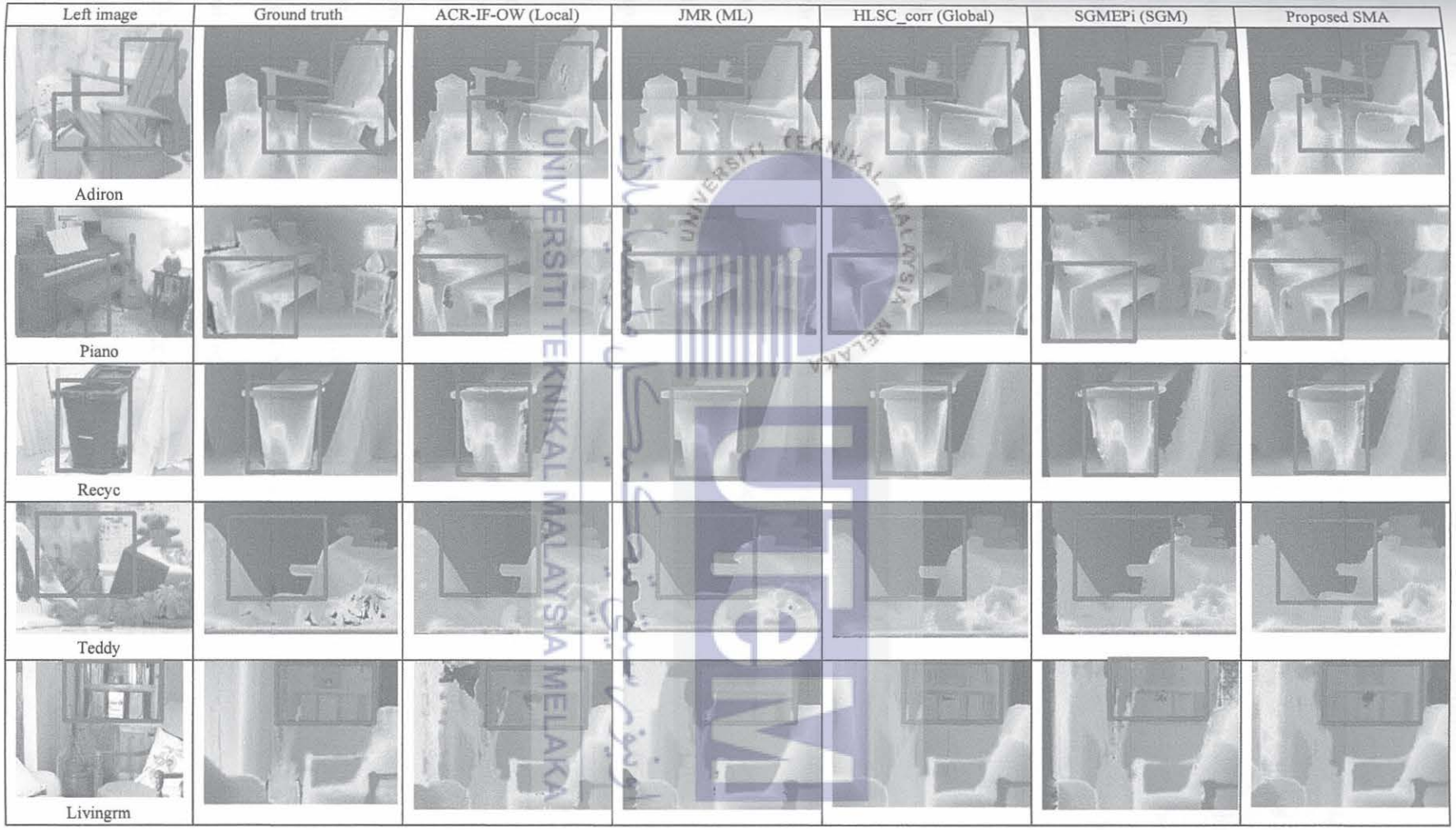


Figure 4.37: Middlebury Qualitative Performance Comparison with Other Methods

The usage of the KITTI dataset enabled the testing of the algorithm's adaptability with more complex real-world stereo images in which the stereo sceneries were subjected to unpredictable lighting conditions attributed to the presence of the sun's natural light, vehicles, and tree shading. Additionally, the test also include dataset containing vast regions with low textures, including the walls, roads, repetitive patterns, and sky. Figure 4.38 exhibits the left reference images, the ground truths, and the generated disparity map results from the training dataset, whereas Figure 4.39 shows the findings from the KITTI testing dataset with almost the same attributes as Figure 4.38, additional error maps, *all* error percentage, and *nonocc* error percentage. The image sequences were 1242 x 375 in size, with a maximum disparity range of 256 levels.

The proposed SMA was able to produce a smooth disparity map and reduced both errors of *all* and *nonocc* based on the quantitative and qualitative measurement. Smooth disparity maps were generated using the dataset's sequential training images, and the ground truth's colour texture and contour levels were accurately reflected. The disparity and error maps produced from the testing images demonstrated that the proposed SMA significantly improved accuracy in regions exhibiting edge discontinuities and low texture. Table 4.17 compares the quantitative measurement results, which represents *all* and *nonocc* errors, to highlight the accomplishment of the proposed SMA through the testing images and other relevant methods. The comparison consisted of the Multi-Block Matching (MBMGPU), Efficient Large-scale Stereo Matching (ELAS_RVC) and Recursive Edge-Aware Filters Aggregation (REAF), which were local method categories. The global methods involved Global 3D Triangular Mesh (MeshStereo), while the semiglobal method included Four Path SGM (CSCT+SGM+MF). ***Appendix C*** presents the detail of KITTI results for training and testing datasets.

Image number	Left image	Ground truth	Result
000010_10			
000011_10			
000012_10			
000013_10			
000014_10			
000015_10			
000016_10			
000017_10			
000018_10			
000019_10			

Figure 4.38: The Result of KITTI Training Dataset

Image number	Left image	Result	Error map	*all* error (%)	*nonocc* error (%)
000010_10				5.36	4.81
000011_10				3.76	3.21
000012_10				2.38	1.70
000013_10				2.56	2.03
000014_10				3.05	2.44
000015_10				5.88	4.87
000016_10				6.97	6.31
000017_10				3.23	2.32
000018_10				11.06	10.58
000019_10				3.76	2.90

Figure 4.39: The Result of KITTI Testing Dataset

Table 4.17: Performance Comparison Based on *all* and *nonocc* errors from the KITTI

Algorithm	Method	Average error	
		all (%)	*nonocc* (%)
Proposed SMA	Local	7.90	7.07
Self-Superflow (Bendig et al., 2022)	ML	8.06	6.93
CSCT + SGM + MF (Hernandez-Juarez et al., 2016)	SGM	8.11	6.56
MBMGPU (Chang and Maruyama, 2018)	Local	8.24	7.47
MeshStereo (C. Zhang et al., 2015)	Global	8.29	7.89
PCOF + ACTF (Derome et al., 2016)	ML	8.38	8.03
PCOF + LDOF (Derome et al., 2016)	ML	8.46	8.03
OASM-Net (Li and Yuan, 2019)	ML	8.65	7.39
ELAS_RVC (Geiger et al., 2011)	Local	9.67	8.80
REAF (Çiŧla, 2015)	Local	10.11	9.29

The Self-supervised Scene Flow (Self-Superflow), Optical Flow Prediction-Correction (PCOF+ACTF) and (PCOF+LDOF), Cooperative Unsupervised Learning (OASM-Net) were among the machine learning methods used. These outcomes established the proposed SMA in the top three of the most accurate methods, with average *nonocc* and *all* errors at 7.07% and 7.90%, respectively, the lowest accuracy values. Recent published methods such OASM-Net (Hernandez-Juarez et al., 2016) and Self-Superflow (Bendig et al., 2022) were outperformed by the proposed SMA. The proposed SMA delivered smooth disparity maps and edge contour details for low-textured objects when compared with other methods, as exhibited in the red box sections in Figure 4.40. The proposed SMA also was able to recover repetitive patterns on the road surfaces.

Figure 4.40: KITTI Qualitative Performance Comparison with Other Methods

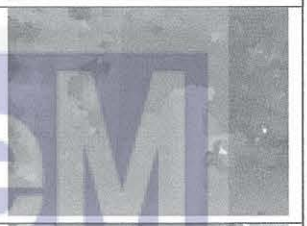
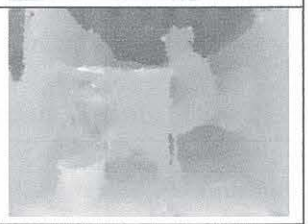
In summary, the KITTI experiment successfully demonstrated that the proposed SMA outperformed other well-established, robust methods in both qualitative and quantitative comparisons.

4.3.2.3 Real Stereo Images Performance

Real stereo images based on UTeMLab-Stereo indoor images were applied to evaluate the efficiency performance of the proposed SMA. The stereo sensor was used to capture six distinct scenes, each with a different layout and stereo challenges. Figure 4.41 shows the left and right image for the UTeMLab-Stereo real stereo images which were Kotak, Kotak2, Cube, Cube5, Cube22 and Cube33 along with the disparity map result. All of the images that were presently shown in the UTeMLab-Stereo were captured using the Bumblebee BB2 stereo vision camera; these images were not modified in any way and did not comprise any kind of image enhancement. The Kotak and Kotak2 exhibited images taken with a stereo camera from two different distances. The boxes were organised to reflect the texture in the images and these images also had contrasting illumination in regions created by various lighting ambient noises. When the stereo camera was nearer to the target, a smooth disparity map was obtained and images comprising of illumination regions were improved. Moreover, the SMA was capable of determining the various dimensions of the targets such as the boxes based on the input images. The disparity map also provided information regarding the edge contours of the target area, such as dividing walls and boxes. However, the illumination and texture regions were compromised when the stereo camera was farther away from the target.

The images Cube, Cube5, Cube22, and Cube33 featured human representatives, various arrangements of objects, and objects of different size, distance, and shape within a cube-shaped region. The Cube and Cube5 scenes showed a large brown box in the cubical

space with different textures of a shirt, waste bin, cubical partition, and chair. On the surface floor, these images also indicated significant textureless regions. Furthermore, it was difficult to match the plain regions caused by the wall divider's colour. Meanwhile, the Cube22 and Cube33 had different places for the trash can, an added object calendar, and a different stereo camera location that was either closer or further away.

Image name	Left image	Right image	Disparity map
Kotak			
Kotak2			
Cube			
Cube5			
Cube22			

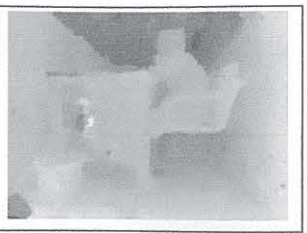
Cube33			

Figure 4.41: The Disparity Map Results of the UTeMLab-Stereo Images

The results, which are displayed in Figure 4.41, demonstrated the capability of producing an accurate disparity map in the region comprising of textureless and plainly coloured surfaces. A clear and detailed edge contour of the object, such as the human posture, chair, waste bin, box, and table in the cubical space, may be visible on the disparity map as well. As anticipated, the quality of the disparity map that was produced was poorer when the stereo camera location was further away from the objects being captured. This seemed to be especially apparent in the region of illumination variations.

4.3.3 3D Reconstruction from Disparity Map

The disparity map produced from the proposed SMA can be deployed in a number of computer vision applications. The 3D surface reconstruction was one of the stereo vision platform's most fundamental applications. Based on the Middlebury dataset and real stereo images from UTeMLab-Stereo, this research provided a 3D surface reconstruction application to prove the efficiency and the accuracy of the SMA. Figure 4.42 shows seven different types of disparity maps and 3D surface reconstructions from the Middlebury and UTeMLab-Stereo datasets (Jadeplant, Playroom, Teddy, Australia, Plants, Kotak2, and Cube22). Leveraging the Computer Vision Toolkit for analysing images and 3D models, the final disparity maps were processed for 3D surface reconstruction. Since there were no ground truth images in the Middlebury test and UTeMLab-Stereo datasets, the qualitative evaluation was used to evaluate the performance of the SMA.

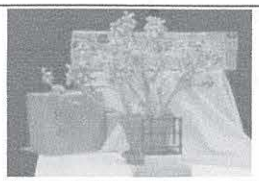
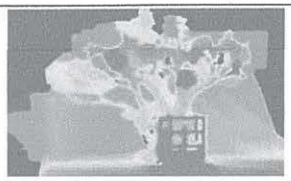
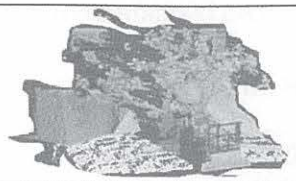
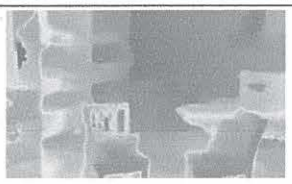
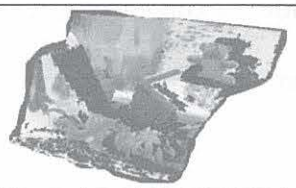
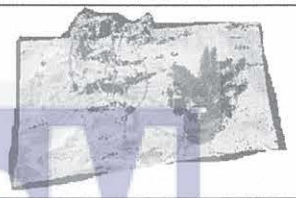
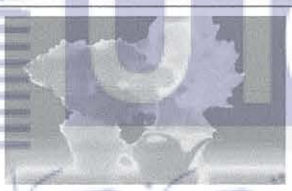
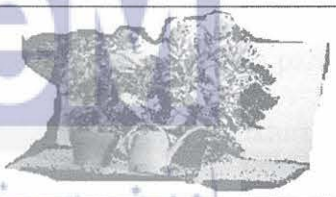
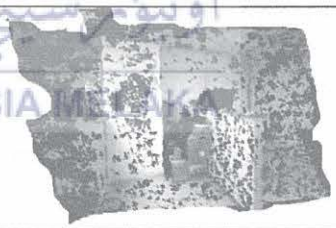
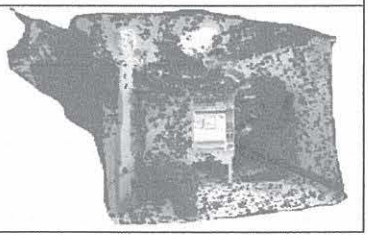
Image name	Left image	Disparity map	3D surface reconstruction
Jadeplant			
Playroom			
Teddy			
Australia			
Plants			
Kotak2			
Cube22			

Figure 4.42: The Disparity Map Results of the 3D Reconstruction Based on Middlebury Dataset and UTeMLab-Stereo

The images from both the test from the UTeMLab-Stereo datasets and the test from the Middlebury datasets were precisely reconstructed and the results exhibited the depth

contours, which showed good outcomes on the object shapes and dimensions. Based on their original dimensions, the plant, vase, and box for the Jadeplant image were mapped precisely. The chair and teddy bear in "Playroom and Teddy" were well reconstructed based on their respective sizes. Moreover, the implementation of the HRA and the SWF successfully preserved the object's edges. The low texture regions were prominent in the Australia and Plants images, particularly in the map and background wall areas. The plant and vase in these images were clearly apparent in three dimensions due to the disparity maps' capability to reconstruct the three-dimensional surface with smooth foreground and background images.

The objects in the Kotak2 and Cube22 images for the UTeMLab-Stereo dataset were well reconstructed based on their shape and dimension, including the boxes, wall divider, waste bin, and human representative. Due to textureless and plain-colored areas, the smaller dark holes contained invalid disparity values, whereas the bigger dark parts were a part of the occlusion. The eRWR and K-means techniques, which reliably restored the textureless and plain surfaces on the disparity map, allowed the SMA to minimise the occlusion and the invalid disparity values. Smooth and well-formed 3D surface reconstructions were effectively generated by the disparity maps produced by the proposed SMA.

4.4 Comparison of Stereo Correspondence Constraints

The performance of the SMA was also evaluated by the stereo correspondence constraints, which examined issues with determining the distance between the two pixels in a stereo image pair. The constraints on stereo correspondence affected the process of extracting depth information from a disparity map and, consequently, the accuracy. Reconstruction of 3D surfaces depended on the level of accuracy and the improvement of these stereo correspondence constraints. Many researchers were focused on these constraints

in recent years for the development of SMA to acquire good depth information. This research compared the radiometric differences, low texture regions, repetitive pattern, depth discontinuities, and occlusion constraints. Middlebury datasets were utilised to develop the disparity map and the average errors which were used to evaluate the proposed SMA with other methods.

Radiometric differences - This term is used to explain the inconsistency between two positions of stereo images that need to be matched and is one of the major challenges that exist when dealing with stereo correspondences. This occurs because, as seen in Figure 4.43, which shows a Middlebury PianoL left reference image with the lamp turned on and vice versa with right image, an image's colours and intensity vary depending on the perspective from which it is perceived. A typical assumption during the matching process is that the corresponding stereo image pixels have a similar colour level. In a real-world environment, a variety of factors, including different camera configurations, lighting-based geometry, and illuminated colour, will prevent the two corresponding pixels from having the same colour depth.

(a) (b)

Figure 4.43: Radiometric Differences Middlebury PianoL (a) Left Image (b) Right Image

In this thesis, the performance of the SMA in stereo images with a radiometric difference condition was evaluated based on the Middlebury PianoL. The PianoL images provided a different colour consistency between the left and right images. Figure 4.44 displays the comparison of disparity map results for PianoL between the proposed SMA,

DAWA-F, ACR-GIF-OW, and MBM methods. The performance comparison of radiometric differences included the ground truth image as the reference image, as well as *all* and *nonocc* errors.

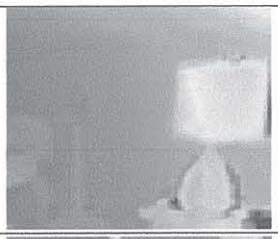
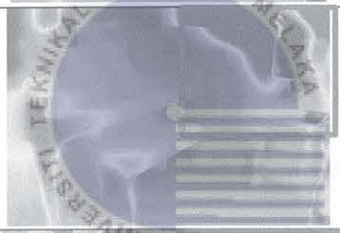
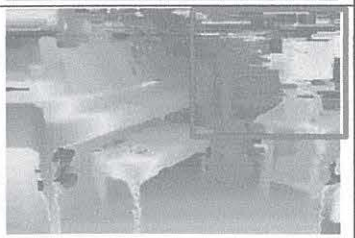
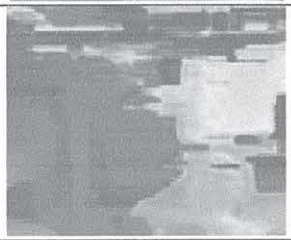
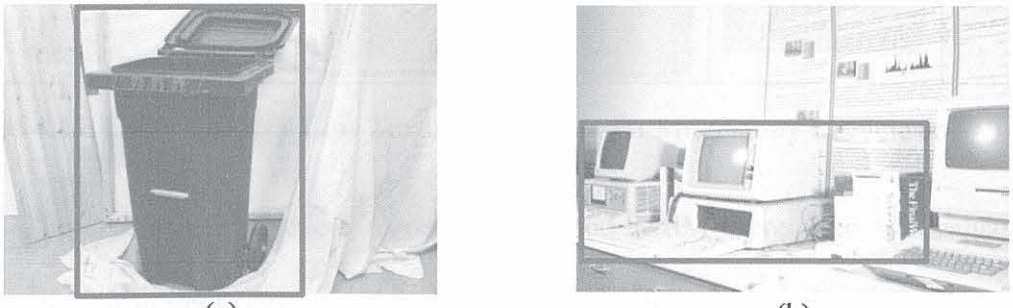
Algorithms	Disparity map (PianoL)	Radiometric differences	Average error	
			nonocc (%)	*all* (%)
Ground Truth			0	0
Proposed SMA			9.28	10.2
DAWA-F			21.3	21.7
ACR-GIF-OW			27.1	27.7
MBM			11.5	12.4

Figure 4.44: Comparison on Radiometric Differences Constraint for Middlebury PianoL

As shown in the red box, the regions affected by the different colour consistency were those near the table lamp. As predicted, the regions near the lamp table generated a significant number of invalid pixels. The proposed SMA produced small areas of invalid pixels and had the lowest average accuracy at 9.28% for *nonocc* error and 10.20% for *all*

error compared with other methods. In conditions of radiometric differences, the ACR-GIF-OW method had larger invalid pixel areas, and the MBM method had significant invalid pixel areas with an additional severe horizontal streak condition. The MBM produced an average accuracy of 11.5% (*nonocc* error) and 12.4% (*all* error), followed by the DAWA-F at 21.3% (*nonocc* error) and 21.7% (*all* error), and the ACR-GIF-OW at 27.1% (*nonocc* error) and 27.7% (*all* error). Therefore, the proposed SMA generated smooth and detailed contours for the disparity map objects, particularly at the lamp table, and successfully reduced the invalid pixels generated by radiometric difference constraints.

Textureless and Low Texture Regions- The areas shown by the red box in Figure 4.45 are the regions that contributed to the mismatching process produced by the plain colour and textureless surface regions, as well as providing a constant luminance in large regions. To establish the algorithm in larger low-texture regions was significantly more difficult and complex due to the similarity of the pixel intensities. The textureless or low texture regions were always regarded as having inadequate texture information, and the colours in these regions had much less contrast with each other, such as the difference between brighter blue and darker blue with the same blue spectrum but distinctive intensities.



(a) (b)

Figure 4.45: Low Texture Middlebury Images (a) Recycle (b) Vintage

Figure 4.46 shows the comparison of disparity maps produced by the proposed SMA and several methods using the Middlebury Recycle and Vintage images to evaluate the

algorithm under low texture performance. In the input images, the objects of interest with low texture regions were the blue trash bin and the desktop screen.

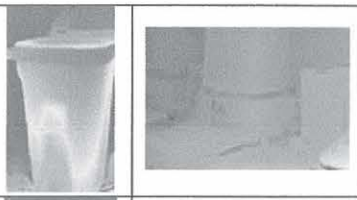
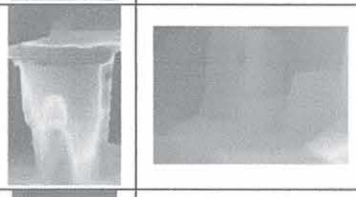
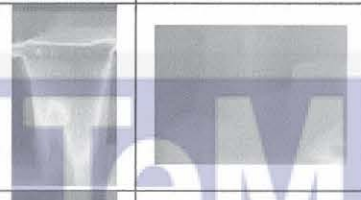
Algorithm	Disparity map (Recycle & Vintage)	Low texture	Average error (Recycle / Vintage)	
			nonocc (%)	*all* (%)
Ground Truth			0	0
Proposed SMA			2.45 / 7.91	3.79 / 9.33
Motion Stereo			3.14 / 14.00	4.64 / 16.00
ADSG			2.47 / 7.17	4.55 / 9.40
MBM			3.02 / 13.3	13.67 / 14.10

Figure 4.46: Comparison on Low Texture Constraint for Middlebury Recycle and Vintage

The methods of Motion Stereo, ADSG and MBM were used to evaluate the performance with the proposed SMA. The blue trash bin and the desktop screen were the foreground object that were separated from the background and consisted of low texture information. Observations from the disparity maps showed the proposed SMA producing smooth and sharp edges in the area of the trash bin with fewer unwanted pixels and almost identical to ground truth when compared with the other methods. Moreover, the Motion

Stereo disparity map produced blurred edges, and the ADSG and MBM experienced edge distortion, which was apparent at the blue trash bin. The condition was similar to the Vintage image in the desktop screen area, which showed the disparity map of the proposed SMA generating clear and detailed contour of the desktop screen compared to the Motion Stereo, ADSG, and MBM. The Motion Stereo and ADSG exhibited significant edge blurring and distortion, whereas the MBM experiences severe horizontal streaks that reduced their average accuracy.

When compared to other methods, the quantitative results showed that the proposed SMA producing the lowest *all* and *nonocc* errors for both images, Recycle and Vintage. The algorithm contributed *all* error accuracy at 3.79% for Recycle image and 9.33% for Vintage images. Meanwhile, for the *nonocc* error the algorithm produced 2.45% for Recycle and 7.91% for Vintage images. The ADSG was ranked second with *all* error at 4.55% and 9.40%, while the *nonocc* error was at 2.47% and 7.17% for Recycle and Vintage images. This was followed by the Motion Stereo at 4.64% and 16.00% for *all* error, and the *nonocc* error at 3.14% and 14.00%. The MBM was ranked last, with *all* error accuracy of 13.67% and 14.10%, and *nonocc* error of 3.02% and 13.3%. The proposed SMA performed effectively well in regions with low texture, which resulted in the reduction in the average accuracy.

Repetitive Patterns- Figure 4.47 displays the areas in the red box as the next common constraint in the stereo correspondences process. The region containing periodic and repeating surface texture was one of the concerns taken into account in the SMA due to a bigger ambiguity in the area for mutual matching points. Typically, man-made and space artefacts possess repeated textures, and this constraint is expected to contribute to mismatching or technical challenges as a consequence of improperly matched coordinates, which offer numerous possible intensity values.

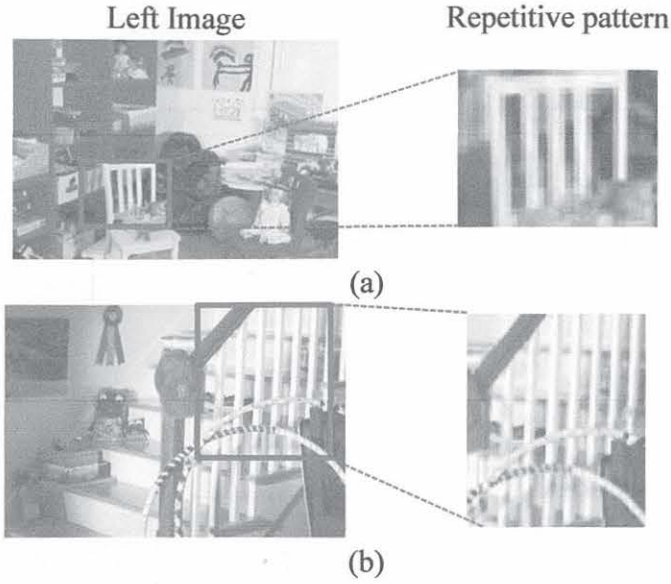


Figure 4.47: Repetitive Pattern of Middlebury Images (a) Playroom (b) Hoops

Figure 4.48 shows the comparison of disparity maps results, the ground truth, *all* and *nonocc* errors between the proposed SMA, ACR-GIF-OW, SGMEPi and MDP. The comparison was made based on Playroom and Hoops images, which contained object with repetitive patterns, backrest of the chair and the stairs. After rigorous examination, the pattern edge of the Playroom's image, especially at the backrest of the chair, was discovered to be sharp and very smooth although the proposed generated small region contained an unwanted invalid pixel. Compared with other methods, the disparity map at the backrest of the chair produced blurry edges and distortions. No artefacts were apparent in the disparity at background of the backrest of the chair, which was a limitation of other methods. Disparity map for all methods, experiencing challenges to produce a good and detailed contour for the box of toys on the chair. Based on the Playroom image, the proposed SMA produced 8.56% *all* error and 4.66% *nonocc* error, which was slightly higher compared with ACR-GIF-OW, which produced *all* error at 8.51% and *nonocc* error at 4.59%. The MDP came next, with 11.0 percent *all* error and 3.46% *nonocc* error. The last method was from the SGMEPi with *all* error at 21.0% and *nonocc* error at 3.48%.

Algorithm	Disparity map (Playroom and Hoops)		Repetitive pattern		Average error (Playroom / Hoops)	
					nonocc (%)	*all* (%)
Ground Truth					0 / 0	0 / 0
Proposed SMA					4.66 / 7.06	8.56 / 18.6
ACR-GIF-OW					4.59 / 13.7	8.51 / 23.1
SGMEPi					3.48 / 5.94	21.00 / 18.6
MDP					3.46 / 8.00	11.0 / 24.7

Figure 4.48: Comparison on Repetitive Pattern Constraint for Middlebury Playroom and Hoops

The performance of the proposed SMA was impressive for the Hoops image, which compared very well but not as well as the SGMEPi. The SGMEPi generated a good disparity map at the area of the staircase, with the repetitive pattern showing sharp and smooth disparity compared with the proposed SMA, which clearly produced a repetitive pattern with edge distortion and edge fattening. A small area of invalid pixels was also detected at the side of the image. Although the qualitative performance of SGMEPi was better than the proposed SMA, the average accuracy was the same for *all* error at 18.6% and slightly lower for *nonocc* error at 5.94% for SGMEPi and 7.06% for the proposed SMA. Other methods, such as the MDP and ACR-GIF-OW, produced significant edge distortion, edge blurring, and edge fattening for the Stairs image. There were also several areas with invalid pixels in the region of staircase. The ACR-GIF-OW ranked third with an average accuracy of 23.1% for *all* error and 13.7% for *nonocc* error. The lowest in the rank was the MDP, with *all* error at 24.7% and *nonocc* error at 8.00%. Hence, the proposed SMA performed as well as, and in some cases, even better than, the more advanced methods in achieving good accuracy when executing images with repetitive pattern constraints.

Depth Discontinuities- The region where pixels are removed and then replaced by the surrounding regions is described as the "depth discontinuities" constraint. If the size of the regions in the stereo pair images differs significantly from one another, this will typically lead to distortion across the depth boundaries and make finding valid corresponding points more difficult. Thin and small objects are frequently impacted because they suffer severe blurring and are subsequently replaced by nearby pixels. Bicycle handles, pipes, pencils, and many more are examples of objects with a thin structure. The problem can be seen clearly at the black pipes as in Middlebury Pipes image and marked in the area of red box as shown in Figure 4.49, which have the sensitive nature towards depth discontinuity.

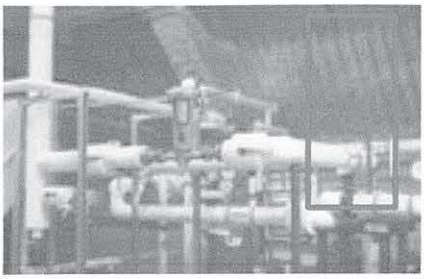


Figure 4.49: Depth Discontinuity of Middlebury Pipes

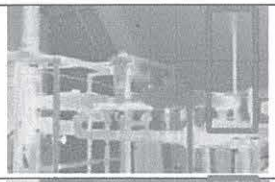
Algorithms	Disparity map (Pipes)	Depth discontinuities	Average error	
			nonocc (%)	*all* (%)
Ground Truth			0	0
Proposed SMA			6.76	13.9
DAWA-F			7.73	15.9
TSGO			4.99	8.88
Motion Stereo			9.66	16.9

Figure 4.50: Comparison on Depth Discontinuities Constraint for Middlebury Pipes

Figure 4.50 shows the proposed SMA and several other methods, including DAWA-F, TSGO, and Motion Stereo addressing the depth discontinuities with the ground truth. A black pipe in the top right of the Pipes image signified the location of the depth discontinuity, which was in the red box. The object's narrow body shape apparently showed the occurrence

of depth discontinuity as expected. However, the black pipe image for the proposed SMA showed a robust structure at the top and bottom of the pipe, but the pipe body was diminished in the middle, based on the qualitative perspective. Thus, the black pipe was still a recognisable structure compared with other methods. The body image of the pipe was the worst for DAWA-F and TSGO, which showed a completely diminished structure and barely recognisable and both bodies of the pipe experiencing depth discontinuity. Meanwhile, Motion Stereo showed the complete body of the pipe, but the area was affected by edge distortion, blurry and experiencing minor depth discontinuity constraint.

Even though the Motion Stereo performed better than the proposed SMA from a qualitative perspective, the proposed SMA outperformed the Motion Stereo in terms of average accuracy by 13.90% for *all* error and 6.76% for *nonocc* error. The Motion Stereo system had an *all* error of 16.90% and a *nonocc* error accuracy of 9.66%. The average accuracy for DAWA-F was at 15.90% for *all* error and 6.76% for *nonocc* error, respectively. The errors for TSGO are 8.88% for *all* error and 4.99% for *nonocc* error, respectively. This was surprising given that the TSGO performed poorly in the depth discontinuity region despite contributing fairly average on the accuracy errors. This analysis provided proof that the proposed SMA was able to provide solution to the depth discontinuity constraint in the stereo correspondence process with good qualitative and quantitative performance.

Occlusion- The occluded regions are a common constraint in stereo matching technologies. Due to geometric displacement, the image match pattern in the target image is not visible in contrast with the reference image. One of the scenes is generated, while another is completely invisible to both cameras. The stereo images cannot be matched if both cameras cannot see something. Low accuracy disparity values are obtained in occlusion regions since they contain unidentifiable objects, shapes, or structures that are challenging to measure. Figure 4.51 shows the occlusion regions occurring in the Middlebury images of

PlaytableP and Teddy. The stereo vision matching approach allowed the pixels from the input images to correspond with each other. Since the left input image was typically pre-set as the reference in this particular instance, disparities based on the left image coordinates were obtained. Middlebury Stereo needed disparity maps with those attributes for accuracy analysis, hence this approach was used. However, the disparity map that was generated was more prone to encounter occlusion on the left side of the disparity map.

Figure 4.51: Occlusion of Middlebury Images (a) PlaytableP (b) Teddy

Figure 4.52 presents the qualitative and quantitative performance of the proposed SMA and several methods in terms of their capability to handle the occlusion. The proposed SMA was compared with ground truth reference, DTS, JMR, and MSMD_ROB methods using Middlebury PlaytableP and Teddy images. The results demonstrated the adequacy of the proposed SMA in handling the occlusion constraint from the disparity map result for both images. No artefacts or invalid pixels were apparent in the disparity map, which was the limitation of other methods. The proposed SMA disparity map showed a sharp and detailed contour of the objects and background structures. Other methods showed apparent occlusion regions consisting of undesired invalid pixels on the left side of the disparity map.

There were occlusion and invalid pixels regions surrounding foreground objects such as the area of the table in PlaytableP and the teddy bear in Teddy image. Larger occlusion regions and unwanted invalid pixels contributed to lower average accuracy for the disparity map. Then, there were also clearly visible regions of horizontal streaks on the disparity map, especially for the DTS and JMR methods. Only the image of Teddy by MSMD_ROB, besides the proposed SMA, showed a smooth disparity map and clear contours of objects.

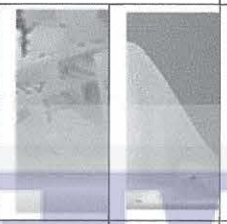
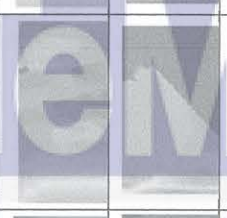
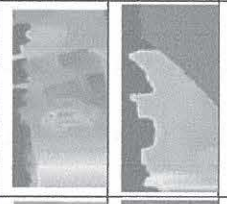
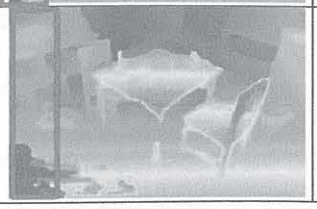
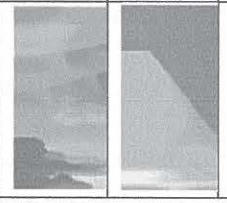
Algorithm	Disparity map (PlaytableP and Teddy)	Occlusion	Average error (PlaytableP / Teddy)	
			nonocc (%)	*all* (%)
Ground Truth			0 / 0	0 / 0
Proposed SMA			2.13 / 2.83	4.51 / 3.46
DTS			2.33 / 1.08	7.69 / 8.71
JMR			1.65 / 1.15	7.44 / 8.98
MSMD_ ROB			6.74 / 2.30	9.32 / 3.02

Figure 4.52: Comparison on Occlusion Constraint for Middlebury PlaytableP and Teddy

In terms of quantitative performance for these methods, the proposed SMA was ranked at the top for PlaytableP and second for Teddy with 4.51% and 3.46% for *all* errors, respectively. *Nonocc* errors contributed 2.13% and 2.83%, respectively. The MSMD_ROB method ranked top for the Teddy image at 3.02% for *all* error, however, the method ranked last for the PlaytableP image at 9.32%. The *nonocc* error for MSMD_ROB was 6.74% and 2.30% for both images. Next, JMR ranked third for the PlaytableP image, where the *nonocc* error was 1.65% and *all* error was 7.66%. The JMR, on the other hand, comes in last in the Teddy image comparison, with a *nonocc* error of 1.15% and an *all* error of 8.98%. Finally, DTS came in third place for PlaytableP and Teddy images, with *nonocc* errors of 2.33% and 1.08%, respectively, and *all* errors of 7.69% and 8.71%. According to the results and performance analysis, the proposed SMAis a potential alternative algorithm to address the occlusion constraint that frequently occurs in the stereo correspondence processes.

4.5 Summary

This chapter presented the outcomes, the qualitative and quantitative performance analysis of the proposed SMA method, and the clarity of the disparity maps. Through each stage of the SMA, each aspect of the research was discussed in relationship to the findings. The first aspect was the determination of optimal parameter settings for the algorithm to generate and obtain a high-quality disparity map while exhibiting fairly low error accuracy. Also, the stage-by-stage error reduction performance was investigated. The adoption of multi-cost matching, which comprised of the TAD, GMC, and MCE in combination with the PPF showed reduced errors for the raw cost volume and sharpen the texture in the matching cost process. Furthermore, the balancing parameter at the multi-cost was able to decrease the error even more. The proposed HRA, which combines the iNLGF and eRWR, enhances the disparity map estimation, particularly at the discontinuity regions, repetitive areas,

occlusions and the object's edges. This stage contributed significantly to error reduction when compared to other stages and outperformed other methods in cost aggregation. In the final stage, the integration of K-means clustering and SWF effectively recovered the occlusion regions, eliminated horizontal artefacts, smoothed the low texture, and enhanced the edge sharpness.

This chapter also presented the performance evaluation of the proposed algorithm using the Middlebury and KITTI datasets as standard benchmarking. The results of the evaluation showed that the proposed algorithm outperformed several established methods with the lowest average accuracy. To further investigate this, the results from the dataset were explored and deployed for 3D surface reconstruction, and the algorithm was executed with real stereo images. The results showed that this approach worked well in all possible scenarios, including real stereo images to 3D surface reconstruction. Results also showed that this algorithm performed better in terms of handling the stereo correspondence constraints, such as the radiometric differences, occlusion, low texture regions, depth discontinuities, and repetitive pattern, with respect to other methods. In summary, this proposed algorithm demonstrated high-quality results in obtaining the disparity map, and the performance outperformed several established literature review's methods such as the local, global, SGM, and ML.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The main conclusions of this research and recommended future works are drawn together and presented in this chapter. The accomplishments of this proposed research that have been demonstrated by the experiment results is concluded in Section 5.1. This section also discusses the relationships regarding the research accomplishments with respect to the objectives of this study. In the last part of this thesis, Section 5.2, several studies that can be extended as future works are proposed.

5.1 Conclusion

This research contributes to the design of new local SMA to produce a disparity map which comprises of four basic stages of taxonomy. The matching cost computation stage (i.e., Stage 1) has been performed using Multi-Cost Pyramid Fusion (MPF) consisting of a pyramid combination of three components which is able to accomplish an accurate pixel cost volume. These three components; the Truncated Absolute Differences (TAD), the Gradient Magnitude CLAHE (GMC), and a new Modified Census Edge (MCE) with the Planar Pyramid Fusion (PPF) are applied to integrate these costs. The Hybrid Random Aggregation (HRA) is a novel hybrid type aggregation that is introduced in the second stage for cost aggregation (i.e., Stage 2). The HRA consists of iterative Non-Local Guided Filter (iNLGF) and extended Random Walk Restart (eRWR) to minimise the ambiguities from the matching process. The minimum cost is determined and used to select the disparity value based on the Winner-Takes-All (WTA) strategy at the disparity selection and the optimisation stage (i.e.,

Stage 3).

The final stage is the disparity refinement (i.e., Stage 4), which is the process to acquire the final disparity map, which performed the hierarchical cluster-edge refinement to further eliminate mismatches produced by occlusion, low texture, and edges preserving. Stage 4 consists of several subsequent steps, including the left-right consistency check process, disparity confidence computation, invalid pixel fill-in based on median interpolation, K-means clustering along with Side Window Filter (SWF).

The first objective is to develop a new computational cost function using the pixel-based stereo matching. The first objective must be accomplished by obtaining data on the implementation of the SMA based on the established taxonomy. This involves a description of the mathematical architecture, a taxonomy of SMA disparity classification, and several benefits and drawbacks of the recent methods as discussed in the literature study. The new matching cost is proposed in Chapter 3, the methodology chapter, which describes the application of the MPF method. The experiments conducted confirm the superiority of the three modified matching cost components especially with the implementation of MCE by producing the lowest average error with 40.9% and 33.3% for *all* and *nonocc* errors, respectively. This method is further analysed by the integration of the PPF which provides a significant performance improvement by reducing the *all* and *nonocc* errors about 3.90% and 4.50%, respectively. The overall performance of the new SMA taxonomy is impressive, with the lowest average errors of *all* and *nonocc*, which are 9.02% and 5.11%, respectively, and significantly improve the boundaries and low texture regions.

The second objective is to design a new stereo matching algorithm that is robust against low texture and occlusion regions with improved edge-preserving properties. A method based on the hybrid random aggregation (HRA) and the hierarchical cluster-edge refinement is proposed to achieve this objective. A new aggregation strategy using the HRA

is introduced in Stage 2. The HRA which combines the pixel cost from iNLGF and the segment cost from eRWR which improved the disparity accuracy significantly and consider the occlusion and depth discontinuities, also accounts for varying illumination and edge preservation. The findings show that such an approach produces good quality results in reducing the *all* and *nonocc* errors by about 21.5% and 22.81%, respectively. Additionally, the K-means clustering, and the SWF are implemented in Stage 4 under the hierarchical cluster-edge refinement to ensure the SMA is improved the disparity performance and robust against the low-textured, repetitive region and preserve the edges. The K-mean clustering process aims to recover the low texture and repetitive regions while the SWF preserves the edges. These improvements in the refinement stage resulted in an impressive performance by reducing the *all* and *nonocc* errors of the Middlebury dataset by 6.48% and 0.88%, respectively. In addition, the KITTI test dataset used on the proposed algorithm also shows the *all* errors to reduce by an average of 1.77% and the *nonocc* errors by an average of 1.73% compared with the established state-of-the-art local SMA.

As stated in the third objective, the performance of the SMA is then validated using real stereo images and benchmarking datasets to compare against other approaches. The triangulation principle is used to establish the depth and the 3D coordinates system in space based on the stereo image projections. Then, the disparity map is produced from the SMA and is applied to generate the 3D reconstruction. This 3D reconstruction can be applied to the stereo vision sensor which functions serves as a passive optical system . In Chapter 4 of the thesis, the performance of the SMA in comparison to other approaches is addressed in Chapter 4 of the thesis. The Middlebury and KITTI test datasets are the two most common standard datasets which are used in this thesis to evaluate the performance of the new SMA. With a 9.02% *all* error table ranking and a 5.11% *nonocc* error table ranking, the Middlebury dataset produces the two highest results. This indicates that the proposed SMA outperformed

the established other local SMA in accuracy by 0.46% and 0.67% for *all* and *nonocc* errors, respectively. The KITTI dataset for the SMA was ranked top three with an *all* error ranking of 7.90% and a *nonocc* error ranking of 7.07%. Generally, the proposed algorithm performs excellently and is competitive with the recently published local, global, and SGM methods found in the literature, outperforming some of their accuracy while also able to decrease the *all* and *nonocc* errors. In summary, this proves that the proposed algorithm in this thesis can be applied as a comprehensive SMA for stereoscopic vision applications.

5.2 Suggestion for Future Work

This section proposes potential future work in relation to SMA development and advancement. The adoption and modification of additional disparity selection and optimisation (Stage 3), such as the global and SGM approaches, may represent a potential development of the system. The features can be studied from this improvement, and performance assessments can be computed. The static scenes serve as the foundation for the input stereo images employed in this thesis. Moving scene images can potentially be used in the subsequent research to actually prove the adaptation by Jeong and Jay Kuo (2019) employed in their multi-view stereo reconstruction of 3D television. As a result, the robustness of the SMA proposed in this thesis can be further investigated in a variety of circumstances.

This thesis do not achieve the real-time implementation of the experimental image execution time. Hence, the deployment of the GPU enables an improved and attainment implementation, according to the literature in Chapter 2. The autonomous vehicle navigation framework, which was developed by Hernandez-Juarez et al. (2016) in their KITTI stereo vision experimental observations, is an example of an application which demands real-time implementation. When processing big blocks of data, the GPU was capable of performing

parallel processing even more effectively than a general-purpose CPU as stated by Chang and Maruyama (2018). Therefore, using a GPU for the SMA proposed in this thesis is a viable strategy for future research that will improve the dependability and efficiency of the proposed SMA for a variety of applications.

REFERENCES

Aboali, M., Manap, N.A., Darsono, A.M., and Yusof, Z.M., 2017. Performance Analysis Between Basic Block Matching and Dynamic Programming of Stereo Matching Algorithm. *Journal of Telecommunication, Electronic and Computer Engineering*, 9(2–13), pp.7–16.

Bae, K.R., and Moon, B., 2017. An Accurate and Cost-Effective Stereo Matching Algorithm and Processor for Real-Time Embedded Multimedia Systems. *Multimedia Tools and Applications*, 76(17), pp.17907–17922.

Bapat, A., and Frahm, J.M., 2019. The Domain Transform Solver. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 6014–6023 2019.

Batsos, K., Cai, C., and Mordohai, P., 2018. CBMV: A Coalesced Bidirectional Matching Volume for Disparity Estimation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2060–2069 2018.

Bebeselea-Sterp, E., Brad, R.R., and Brad, R.R., 2017. A Comparative Study of Stereovision Algorithms. *International Journal of Advanced Computer Science and Applications*, 8(11), pp.359–375.

Bendig, K., Schuster, R., and Stricker, D., 2022. Self-Superflow: Self-Supervised Scene Flow Prediction in Stereo Sequences. *2022 IEEE International Conference on Image Processing (ICIP)*, 481–485 2022.

Bethmann, F., and Luhmann, T., 2015. Semi-Global Matching in Object Space. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 23–30 2015.

Bhalerao, R.H., Gedam, S.S., and Buddhiraju, K.M., 2017. Modified Dual Winner Takes All Approach for Tri-Stereo Image Matching Using Disparity Space Images. *Journal of the*

Indian Society of Remote Sensing, 45, pp.45–54.

Bleyer, M., and Breiteneder, C., 2013. Stereo Matching—State-of-the-Art and Research Challenges. In *Advanced Topics in Computer Vision*, pp. 143–179.

Buades, A., and Facciolo, G., 2015. Reliable Multiscale and Multi-Window Stereo Matching. *SIAM Journal on Imaging Sciences*, 8(2), pp.888–915.

Cabezas, I., Padilla, V., and Trujillo, M., 2011. A Measure for Accuracy Disparity Maps Evaluation. *Iberoamerican Congress on Pattern Recognition*, 223–231 2011. Springer, Berlin, Heidelberg.

Cao, Y.S., Liu, J.G., Wen, T.X., and Bi, X., 2019. Improvement of Stereo Matching Algorithm based on Guided Filtering and Kernel Regression. *Journal of Physics: Conference Series*, 1213(3), pp.1–6.

Chang, Q., and Maruyama, T., 2018. Real-Time Stereo Vision System: A Multi-Block Matching on GPU. *IEEE Access*, 6, pp.42030–42046.

Chang, Q., Zha, A., Wang, W., Liu, X., Onishi, M., Lei, L., Er, M.J., and Maruyama, T., 2022. Efficient Stereo Matching on Embedded GPUs with Zero-Means Cross Correlation. *Journal of Systems Architecture*, 123, p.102366.

Chang, T.A., Lu, X., and Yang, J.F., 2017. Robust Stereo Matching with Trinary Cross Color Census and Triple Image-Based Refinements. *Eurasip Journal on Advances in Signal Processing*, 2017(1)(27), pp.1–13.

Chang, Y.J., and Ho, Y.S., 2019. Adaptive Pixel-wise and Block-wise Stereo Matching in Lighting Condition Changes. *Journal of Signal Processing Systems*, 91, pp.1305–1313.

Chen, H., Wang, L., and Liu, G., 2020. A Survey of Stereo Matching Algorithms. *Chinese High Technology Letters*, 11, pp.82–85.

Cheng, F., Zhang, H., Sun, M., and Yuan, D., 2015. Cross-Trees, Edge and Superpixel Priors-Based Cost Aggregation for Stereo Matching. *Pattern Recognition*, 48(7), pp.2269–

2278.

Çiıla, C., 2015. Recursive Edge-Aware Filters for Stereo Matching. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 27–34 2015.

Dasilva Vieira, G., Alphonsus Soares, F.A.M.N., Laureano, G.T., Parreira, R.T., Ferreira, J.C., Costa, R.M., and Ferreira, C.B.R., 2018. Disparity Refinement Through Grouping Areas and Support Weighted Windows. *Canadian Conference on Electrical and Computer Engineering*, 1–4 2018. IEEE.

Derome, M., Plyer, A., Sanfourche, M., and Le Besnerais, G., 2016. A Prediction-Correction Approach for Real-Time Optical Flow Computation using Stereo. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 365–376 2016. Springer International Publishing.

Diaz, C., Walker, M., Szafir, D.A., and Szafir, D.A., 2017. Designing for Depth Perceptions in Augmented Reality. *Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2017*, 111–122 2017.

Dong, H., Wang, T., Yu, X., and Ren, P., 2018. Stereo Matching via Dual Fusion. *IEEE Signal Processing Letters*, 25(5), pp.615–619.

Dong, Q., and Feng, J., 2018. Adaptive Disparity Computation using Local and Non-Local Cost Aggregations. *Multimedia Tools and Applications*, 77, pp.31647–31663.

Du, X., El-Khamy, M., and Lee, J., 2019. AMNet: Deep Atrous Multiscale Stereo Disparity Estimation Networks. *arXiv preprint*, arXiv:1904, pp.1–25.

Du, Y., and Jia, K., 2019. Neighborhood Correlation and Window Adaptive Stereo Matching Algorithm. *Journal of Information Hiding and Multimedia Signal Processing*, 10(4), pp.509–516.

Emlek, A., Peker, M., and Yalçin, M.K., 2018. Improving the Cost-Volume Based Local Stereo Matching Algorithm. *26th IEEE Signal Processing and Communications*

Applications Conference, SIU 2018, 1–4 2018.

Ende, W., Yalong, Z., Liangyu, P., Yijun, L., and Tianyao, W., 2018. Stereo Matching Algorithm based on the Combination of Matching Costs. *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, CYBER 2017*, 1001–1004 2018.

Fan, R., Ai, X., and Dahnoun, N., 2018. Road Surface 3D Reconstruction Based on Dense Subpixel Disparity Map Estimation. *IEEE Transactions on Image Processing*, 27(6), pp.3025–3035.

Fan, R., Liu, Y., Yang, X., Bocus, M.J., Dahnoun, N., and Tancock, S., 2018. Real-Time Stereo Vision for Road Surface 3-D Reconstruction. *IST 2018 - IEEE International Conference on Imaging Systems and Techniques, Proceedings*, 1–6 2018.

Fu, Y., Chen, W., Lai, K., Zhou, Y., and Tang, J., 2019. Rank-Based Encoding Features for Stereo Matching. *IEEE Multimedia*, 26(4), pp.28–42.

Geiger, A., Lenz, P., Stiller, C., and Urtasun, R., 2020, *The KITTI Vision Benchmark Suite* [Online]. Available at: http://www.cvlibs.net/datasets/kitti/index.php [Accessed: 15 May 2022].

Geiger, A., Roser, M., and Urtasun, R., 2011. Efficient Large-Scale Stereo Matching. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6492 LNCS(PART 1), pp.25–38.

Geng, D.D., and Luo, N., 2017. A Dynamic Programming Global Stereo Matching Algorithm Based on Multiple Neighbors' Nonlinear Diffusion. *Huadong Ligong Daxue Xuebao/Journal of East China University of Science and Technology*, 54(14), pp.876–878.

Gong, Y., Liu, B., Hou, X., and Qiu, G., 2018. Sub-window Box Filter. *2018 IEEE Visual Communications and Image Processing (VCIP)*, 1–4 2018. IEEE.

Guanghan Pan, Sun, T., Weed, T., and Scharstein, D., 2021, *2021 Mobile stereo datasets*

with ground truth [Online]. Available at: https://vision.middlebury.edu/stereo/data/scenes2021/ [Accessed: 10 June 2022].

Hadfield, S., Lebeda, K., and Bowden, R., 2017. Stereo Reconstruction using Top-Down Cues. *Computer Vision and Image Understanding*, 157, pp.206–222.

Hallek, M., Boukamcha, H., Mtibaa, A., and Atri, M., 2022. Dynamic Programming with Adaptive and Self-Adjusting Penalty for Real-Time Accurate Stereo Matching. *Journal of Real-Time Image Processing*, 19(2), pp.233–245.

Hamzah, R.A., and Ibrahim, H., 2018. Improvement of Stereo Matching Algorithm Based on Sum of Gradient Magnitude Differences and Semi-Global method with Refinement Step. *Electronics Letters*, 54(14), pp.876–878.

Hamzah, R.A., Ibrahim, H., and Abu Hassan, A.H., 2017. Stereo Matching Algorithm Based on Per Pixel Difference Adjustment, Iterative Guided Filter and Graph Segmentation. *Journal of Visual Communication and Image Representation*, 42, pp.145–160.

Hamzah, R.A., Wei, M.G.Y., and Anwar, N.S.N., 2020. Development of Stereo Matching Algorithm Based on Sum of Absolute RGB Color Differences and Gradient Matching. *International Journal of Electrical and Computer Engineering*, 10(3), pp.2375–2382.

Hamzah, R.A.R.A., Kadmin, A.F.F., Hamid, M.S.S., Ghani, S.F.A.F.A., and Ibrahim, H., 2018. Improvement of Stereo Matching Algorithm for 3D Surface Reconstruction. *Signal Processing: Image Communication*, 65, pp.165–172.

Han, X., Liu, Y., and Yang, H., 2019. A Stereo Matching Algorithm Guided by Multiple Linear Regression. *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics*, 31(1), pp.84–93.

He, K., Sun, J., and Tang, X., 2013. Guided Image Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), pp.1397–1409.

Hernandez-Juarez, D., Chacon, A., Espinosa, A., Vazquez, D., Moure, J.C., and Lopez, A.M.,

2016. Embedded Real-Time Stereo Estimation via Semi-Global Matching on the GPU. *Procedia Computer Science*, 80, pp.143–153.

Hirschmüller, H., et al., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), pp.328–341.

Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), pp.328–341.

Hirschmüller, H., and Scharstein, D., 2007. Evaluation of Cost Functions for Stereo Matching. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–8 2007.

Hong, G.S., and Kim, B.G., 2017. A Local Stereo Matching Algorithm Based on Weighted Guided Image Filtering for Improving the Generation of Depth Range Images. *Displays*, 49, pp.80–87.

Hong, P.N., and Ahn, C.W., 2020. Unsupervised Learning for Stereo Matching Using Single-View Videos. *IEEE Access*, 8, pp.73804–73815.

Hou, Y., Liu, C., An, B., and Liu, Y., 2022. Stereo Matching Algorithm Based on Improved Census Transform and Texture Filtering. *Optik*, 249((2022) 168186), pp.1–9.

Hu, X., Wu, Y., Zhang, D., Qian, L., and Wu, L., 2018. Research on Improvement of Stereo Matching Algorithm Based on ELAS. *Journal of Computers (Taiwan)*, 29(4), pp.110–121.

Huang, C.H., and Yang, J.F., 2022. Improved Quadruple Sparse Census Transform and Adaptive Multi-Shape Aggregation Algorithms for Precise Stereo Matching. *IET Computer Vision*, 2022(16), pp.159–179.

Huang, H., Huang, B., Lu, H., and Weng, H., 2017. Stereo Matching Using Conditional Adversarial Networks. *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part III 24*, 124–

132 2017. Springer International Publishing.

Huang, J., 2015. Stereo Matching Based on Segmented B-Spline Surface Fitting and Accelerated Region Belief Propagation. *IET Computer Vision*, 9(4), pp.456–466.

Huo, G., and Luo, Y., 2019. An Edge-Constrained Iterative Cost Aggregation Method for Stereo Matching. *IEEE International Conference on Robotics and Biomimetics, ROBIO 2019*, 1783–1790 2019.

Jachalsky, J., Schlosser, M., and Gandolph, D., 2010. Confidence Evaluation for Robust, Fast-Converging Disparity Map Refinement. *2010 IEEE International Conference on Multimedia and Expo, ICME 2010*, 1399–1404 2010.

Jafari Malekabadi, A., Khojastehpour, M., and Emadi, B., 2019. Comparison of Block-Based Stereo and Semi-Global Algorithm and effects of Pre-Processing and Imaging Parameters on Tree Disparity Map. *Scientia Horticulturae*, 247, pp.264–274.

Jang, M., Yoon, H., Lee, Seongmin, Kang, J., and Lee, Sanghoon, 2022. A Comparison and Evaluation of Stereo Matching on Active Stereo Images. *Sensors*, 22(9), pp.1–33.

Jellal, R.A., Lange, M., Wassermann, B., Schilling, A., and Zell, A., 2017. LS-ELAS: Line Segment Based Efficient Large Scale Stereo Matching. *Proceedings - IEEE International Conference on Robotics and Automation (ICRA)*, 146–152 2017.

Jeong, Y.J., and Jay Kuo, C.C., 2019. Stereo Matching with Confidence-Region Decomposition and Processing. *Journal of Electrical Engineering and Technology*, 14(1), pp.463–469.

Ji, S.-W., Kim, S.-W., Lim, D.-P., Jung, S.-W., and Ko, S.-J., 2020. Quaternary Census Transform based on the Human Visual System for Stereo Matching. *IEEE Access*, 8, pp.116501–116514.

Jia, B., Liu, S., and Du, Z., 2016. A Progressive Framework for Dense Stereo Matching. *Pattern Recognition and Image Analysis*, 26(2), pp.294–301.

Jia, X., Chen, W., Li, C., Liang, Z., Wu, M., Tan, Y., and Huang, L., 2021. Multi-Scale Cost Volumes Cascade Network for Stereo Matching. *Proceedings - IEEE International Conference on Robotics and Automation*, 8657–8663 2021.

Jia, Y., Zhu, K., Yu, J., Rong, C., and Li, C., 2020. Stereo Matching Based on Improved Minimum Spanning Tree. *Proceedings - 2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2020 (Vol 2)*, 149–153 2020.

Jin, Y., and Wei, W., 2022. Image Edge Enhancement Detection Method of Human-Computer Interaction Interface Based on Machine Vision Technology. *Mobile Networks and Applications*, 27(2022), pp.775–783.

Joung, S., Kim, S., Park, K., and Sohn, K., 2020. Unsupervised Stereo Matching Using Confidential Correspondence Consistency. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), pp.2190–2203.

Jung, C., Chen, X., Cai, J., Lei, H., Yun, I., and Kim, J., 2015. Boundary-Preserving Stereo Matching with Certain Region Detection and Adaptive Disparity Adjustment. *Journal of Visual Communication and Image Representation*, 33, pp.1–9.

Kavitha, V., and Balakrishnan, G., 2020. A Performance Analysis of Stereo Matching Algorithms for Stereo Vision Applications in Smart Environments. *Electronic Government*, 16(1–2), pp.210–221.

Kerkaou, Z., El Ansari, M., Masmoudi, L., and Lahmyed, R., 2021. Dense Spatio-Temporal Stereo Matching for Intelligent Driving Systems. *IET Image Processing*, 15, pp.715–723.

Khan, A., Khan, M.U.K., and Kyung, C.-M., 2018. Intensity Guided Cost Metric for Fast Stereo Matching Under Radiometric Variations. *Optics Express*, 26(4), pp.1096–4111.

Kim, S., Ham, B., Ryu, S., Kim, S.J., and Sohn, K., 2015. Robust Stereo Matching using Probabilistic Laplacian Surface Propagation. *Asian Conference on Computer Vision*, 368–383 2015. Springer, Cham.

Kim, S.S., Min, D., Kim, S.S., and Sohn, K., 2019. Unified Confidence Estimation Networks for robust Stereo Matching. *IEEE Transactions on Image Processing*, 28(3), pp.1299–1313.

Knöbelreiter, P., Reinbacher, C., Shekhovtsov, A., and Pock, T., 2017. End-to-End Training of Hybrid CNN-CRF Models for Stereo. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2339–2348 2017.

Kok, K.Y., and Rajendran, P., 2019. A Review on Stereo Vision Algorithms: Challenges and Solutions. *ECTI Transactions on Computer and Information Technology*, 13(2), pp.134–151.

Kong, L., Zhu, J., and Ying, S., 2021. Local Stereo Matching Using Adaptive Cross-Region-Based Guided Image Filtering with Orthogonal Weights. *Mathematical Problems in Engineering*, 2021, pp.1–20.

Kordelas, G.A., Alexiadis, D.S., Daras, P., and Izquierdo, E., 2015. Enhanced Disparity Estimation in Stereo Images. *Image and Vision Computing*, 35, pp.31–49.

Kordelas, G.A., Alexiadis, D.S., Daras, P., and Izquierdo, E., 2016. Content-Based Guided Image Filtering, Weighted Semi-Global Optimization, and Efficient Disparity Refinement for Fast and Accurate Disparity Estimation. *IEEE Transactions on Multimedia*, 18(2), pp.155–170.

Lee, J., Jun, D., Eem, C., and Hong, H., 2016. Improved Census Transform for Noise Robust Stereo Matching. *Optical Engineering*, 55(6), pp.063107(1)-063107(10).

Lee, S., Lee, J.H., Lim, J., and Suh, I.H., 2015. Robust Stereo Matching using Adaptive Random Walk with Restart Algorithm. *Image and Vision Computing*, 37(May 2015), pp.1–11.

Lee, Y., Park, M.G., Hwang, Y., Shin, Y., and Kyung, C.M., 2018. Memory-Efficient Parametric Semiglobal Matching. *IEEE Signal Processing Letters*, 25(2), pp.194–198.

Lee, Z., Juang, J., and Nguyen, T.Q., 2013. Local Disparity Estimation with Three-Moded Cross Census and Advanced Support Weight. *IEEE Transactions on Multimedia*, 15(8),

pp.1855–1864.

Lee, Z., Member, S., Juang, J., and Nguyen, T.Q., 2013. Local Disparity Estimation with Three-Moded. *IEEE Transactions on Multimedia*, 15(8), pp.1855–1864.

Li, A., Chen, D., Liu, Y., and Yuan, Z., 2016. Coordinating Multiple Disparity Proposals for Stereo Computation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4022–4030 2016.

Li, A., and Yuan, Z., 2019. Occlusion Aware Stereo Matching via Cooperative Unsupervised Learning. *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part VI*, 97–213 2019.

Li, H., Chen, L., and Li, F., 2019. An Efficient Dense Stereo Matching Method for Planetary Rover. *IEEE Access*, 7, pp.48551–48564.

Li, H., Gao, Y., Huang, Z., and Zhang, Y., 2019. Stereo Matching Based on Multi-Scale Fusion and Multi-Type Support Regions. *Journal of the Optical Society of America A*, 36(9), pp.1523–1533.

Li, H., Zhang, Y., and Gao, Y., 2020. Texture Category-Based Matching Cost and Adaptive Support Window for Local Stereo Matching. *Journal of Electronic Imaging*, 29(2), pp.023026–023026.

Li, J., Zhao, H., Li, Z., Gu, F., Zhao, Z., Ma, Y., and Fang, M., 2018. A Long Baseline Global Stereo Matching Based upon Short Baseline Estimation. *Measurement Science and Technology*, 29(5), p.055201.

Li, L., Zhang, S., Yu, X., and Zhang, L., 2018. PMSC: PatchMatch-Based Superpixel Cut for Accurate Stereo Matching. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(3), pp.679–692.

Li, Q., Duan, Z., Zhang, Y., and Zhu, C., 2020. Stereo Matching Algorithm Based on Cross-scale Random Walk. *Huanan Ligong Daxue Xuebao/Journal of South China University of*

Technology (Natural Science), 2(1), pp.121–126.

Li, Q., Ni, J., Ma, Y., and Xu, J., 2018. Stereo Matching using Census Cost Over Cross Window and Segmentation-Based Disparity Refinement. *Journal of Electronic Imaging*, 27(2), pp.023014–023014.

Li, Yunsong, Hu, Y., Song, R., Rao, P., and Wang, Y., 2018. Coarse-to-Fine PatchMatch for Dense Correspondence. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9), pp.2233–2245.

Li, Yibo, Xiong, Z., Wan, Z., Liu, J., and Yu, Y., 2018. Stereo Matching with Improved ADCensus Transform and Adaptive Window (IEEE/CSAA GNCC). *2018 IEEE CSAA Guidance, Navigation and Control Conference, CGNCC 2018*, 1–5 2018.

Li, Y., Zhang, J., Zhong, Y., and Wang, M., 2019. An Efficient Stereo Matching Based on Fragment Matching. *Visual Computer*, 35(2), pp.257–269.

Liang, Z., Guo, Y., Feng, Y., Chen, W., Qiao, L., Zhou, L., Zhang, J., and Liu, H., 2021. Stereo Matching Using Multi-Level Cost Volume and Multi-Scale Feature Constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), pp.300–315.

Lim, J., and Lee, S., 2019. Patchmatch-Based Robust Stereo Matching under Radiometric Changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5), pp.1203–1212.

Lin, C.H., and Liu, C.W., 2015. Accurate Stereo Matching Algorithm Based on Cost Aggregation with Adaptive Support Weight. *Imaging Science Journal*, 63(8), pp.423–432.

Lin Cao ; Weiwei Yu, 2021. Passive Stereo Matching Algorithms-A Review. *International Core Journal of Engineering*, 7(6), pp.492–501.

Liu, H., Wang, R., Xia, Y., and Zhang, X., 2020. Improved Cost Computation and Adaptive Shape Guided Filter for Local Stereo Matching of Low Texture Stereo Images. *Applied Sciences (Switzerland)*, 10(5), pp.1869(1)-1869(17).

Liu, H., Zhang, H., Nie, X., He, W., Luo, D., Jiao, G., and Chen, W., 2021. Stereo Matching Algorithm Based on Two-Phase Adaptive Optimization of AD-Census and Gradient Fusion. *2021 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2021*, 726–731 2021.

Liu, J., Zhou, Z., Xu, W., and Hu, J., 2019. Adaptive Support-Weight Stereo-Matching Approach with Two Disparity Refinement Steps. *IETE Journal of Research*, 65(3), pp.310–319.

Liu, Y., and Aggarwal, J.K., 2005. Local and Global Stereo Methods. In *Handbook of Image and Video Processing*, pp. 297–308.

Loghman, M., Kim, J., and Choi, K., 2018. Fast Depth Estimation using Semi-Global Matching and Adaptive Stripe-Based Optimization. *Journal of Supercomputing*, 74(8), pp.3666–3684.

Lu, B., Sun, L., Yu, L., and Dong, X., 2021. An Improved Graph Cut Algorithm in Stereo Matching. *Displays*, 69(September 2021), pp.102052(1)-102052(7).

Lu, H., Xu, H., Zhang, L., Ma, Y., and Zhao, Y., 2018. Cascaded Multi-scale and Multi-dimension Convolutional Neural Network for Stereo Matching. *VCIP 2018 - IEEE International Conference on Visual Communications and Image Processing*, 1–4 2018.

Lu, H., Xu, H., Zhang, L., Ma, Y., and Zhao, Y., 2018. Cascaded Multi-scale and Multi-dimension Convolutional Neural Network for Stereo Matching. *VCIP 2018 - IEEE International Conference on Visual Communications and Image Processing*, 1–4 2018.

Lu, J., Li, Y., Yang, H., Min, D., Eng, W., and Do, M.N., 2017. PatchMatch Filter: Edge-Aware Filtering Meets Randomized Search for Visual Correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9), pp.1866–1879.

Lv, C., Li, J., Kou, Q., Zhuang, H., and Tang, S., 2021. Stereo Matching Algorithm Based on HSV Color Space and Improved Census Transform. *Mathematical Problems in*

Engineering, 2021, pp.1857327 (1–17).

Ma, H., Zheng, S., Li, C., Li, Y., Gui, L., and Huang, R., 2017. Cross-Scale Cost Aggregation Integrating Intrascale Smoothness Constraint with Weighted Least Squares in Stereo Matching. *Journal of the Optical Society of America A*, 34(4), pp.648–656.

Ma, H., Zheng, S., Li, Y., Gui, L., Huang, R., and Wei, H., 2018. Confidence-Based Iterative Efficient Large-Scale Stereo Matching. *Cogent Engineering*, 5(1), p.1427676.

Ma, N., Men, Y., Men, C., and Li, X., 2016. Dense Stereo Matching Based on Multiobjective Fitness Function - A Genetic Algorithm Optimization Approach for Stereo Correspondence. *Symmetry*, 8(159), pp.159(1)-159(22).

Mahato, M., Gedam, S., Joglekar, J., and Buddhiraju, K.M., 2019. Dense stereo matching based on multiobjective fitness function - A genetic algorithm optimization approach for stereo correspondence. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6), pp.3341–3353.

Mannan Mondal, A., Haider Ali, M., and Abdul Mannan Mondal, M., 2017. Performance Review of the Stereo Matching Algorithms. *American Journal of Computer Science and Information Engineering*, 4(1), pp.7–15.

Marr, D., and Poggio, T., 1976. Cooperative Computation of Stereo Disparity. *Science*, 194(4262), pp.283–287.

Matsuo, T., Fujita, S., Fukushima, N., and Ishibashi, Y., 2015. Efficient Edge-Awareness Propagation via Single-Map Filtering for Edge-Preserving Stereo Matching. *Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015*, 9393, p.9393OS.

Men, Y., Ma, N., Zhang, G., Li, X., Men, C., and Sun, P., 2015. A Stereo Matching Algorithm Based on Census Transform and Improved Dynamic Programming. *Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology*, 47(3), pp.60–65.

Mohd Saad Hamid, NurulFajar Abd Manap, Rostam Affendi Hamzah, A.F.K., 2020. Stereo Matching Algorithm Based on Deep Learning : A Survey. *Journal of King Saud University - Computer and Information Sciences*, 34(5), pp.1–11.

Mozerov, M.G., and Van De Weijer, J., 2015. Accurate Stereo Matching by Two-Step Energy Minimization. *IEEE Transactions on Image Processing*, 24(3), pp.1153–1163.

Mozerov, M.G., and Van De Weijer, J., 2019. One-View Occlusion Detection for Stereo Matching with a Fully Connected CRF Model. *IEEE Transactions on Image Processing*, 28(6), pp.2936–2947.

Navarro, J., and Buades, A., 2019. Semi-Dense and Robust Image Registration by Shift Adapted Weighted Aggregation and Variational Completion. *Image and Vision Computing*, 89, pp.258–275.

Nguyen, P.H., and Ahn, C.W., 2020. Parameter Selection Framework for Stereo Correspondence. *Machine Vision and Applications*, 31(4), p.27.

Ni, J., Li, Q., Liu, Y., and Zhou, Y., 2018. Second-Order Semi-Global Stereo Matching Algorithm Based on Slanted Plane Iterative Optimization. *IEEE Access*, 6, pp.61735–61747.

Pan, C., Liu, Y., and Huang, D., 2019. Novel Belief Propagation Algorithm for Stereo Matching with a Robust Cost Computation. *IEEE Access*, 7, pp.29699–29708.

Peng, X., Bouzerdoum, A., and Phung, S.L., 2018. Efficient Cost Aggregation for Feature-Vector-Based Wide-Baseline Stereo Matching. *Eurasip Journal on Image and Video Processing*, 2018(1), p.24.

Qi, J., and Liu, L., 2022. The Stereo Matching Algorithm Based on an Improved Adaptive Support Window. *IET Image Processing*, 16(March), pp.2803–2816.

Rathnayaka, P., and Park, S.Y., 2020. IGG-MBS: Iterative Guided-Gaussian Multi-Baseline Stereo Matching. *IEEE Access*, 8, pp.99205–99218.

Ross, P., English, A., Ball, D., Upcroft, B., Wyeth, G., and Corke, P., 2014. Novelty-Based

Visual Obstacle Detection in Agriculture. *Proceedings - IEEE International Conference on Robotics and Automation*, 1699–1705 2014.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., and Westling, P., 2014. High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth. *German conference on pattern recognition*, 31–42 2014. Springer, Cham.

Scharstein, D., and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo. *International Journal of Computer Vision*, 47(1–3), pp.7–42.

Scharstein, D., Taniai, T., and Sinha, S.N., 2018. Semi-Global Stereo Matching with Surface Orientation Priors. *Proceedings - 2017 International Conference on 3D Vision (3DV 2017)*, 215–224 2018.

Schönberger, J.L., Sinha, S.N., and Pollefeys, M., 2018. Learning to Fuse Proposals from Multiple Scanline Optimizations in Semi-Global Matching. *Proceedings of the European Conference on Computer Vision (ECCV)*, 739–755 2018.

Shengxy, 2015, *Image Fusion Using Gaussian Pyramid and Laplacian Pyramid* [Online]. Available at: https://github.com/CSshengxy/Laplacian-Pyramid-Blending [Accessed: 3 March 2022].

Shi, H., Zhu, H., Wang, J., Yu, S.Y., and Fu, Z.F., 2016. Segment-Based Adaptive Window and Multi-Feature Fusion for Stereo Matching. *Journal of Algorithms and Computational Technology*, 10(1), pp.3–11.

Da Silva Vieira, G., Soares, F.A.A.M.N., Laureano, G.T., Parreira, R.T., and Ferreira, J.C., 2018. A Segmented Consistency Check Approach to Disparity Map Refinement. *Canadian Journal of Electrical and Computer Engineering*, 41(4), pp.218–223.

Singh, N.J., Kumar, W.K., and Nongmeikapam, K., 2020. Distance Calculation of an Object in a Stereo Vision System. *dvances in Computational Intelligence, Security and Internet of Things: Second International Conference, ICCISIoT 2019, Agartala, India, December 13–*

14, 2019, Proceedings 2, 407–416 2020.

Song, K., Yan, Y., Niu, M., and Liu, C., 2015. Effective Stereo Matching Method with Equicrural Triangle Census Transform. *Journal of Computational Information Systems*, 8(21), pp.7769–7780.

Song, X., Zhao, X., Fang, L., Hu, H., and Yu, Y., 2020. EdgeStereo: An Effective Multi-task Learning Network for Stereo Matching and Edge Detection. *International Journal of Computer Vision*, 128(2020), pp.910–930.

Suenaga, H., Tran, H.H., Liao, H., Masamune, K., Dohi, T., Hoshi, K., and Takato, T., 2015. Vision-Based Markerless Registration using Stereo Vision and an Augmented Reality Surgical Navigation System: A Pilot Study. *BMC Medical Imaging*, 15(1), pp.1–11.

Sung, C.Y., Tseng, Y.W., and Chen, C.H., 2019. Superpixel Smoothing for Disparity Refinement in Stereo Matching. *Proceedings - 2018 International Symposium on Computer, Consumer and Control, IS3C 2018*, 50–53 2019.

Szeliski, D.S. and R., 2020, *Middlebury Stereo Evaluation - Version 3* [Online]. Available at: https://vision.middlebury.edu/stereo/eval3/ [Accessed: 15 May 2022].

Szeliski, R., 2022. *Computer Vision: Algorithms and Applications*, Springer Science & Business Media.

Tabssum, T., Charles, P., and Patil, A. V., 2017. Evaluation of Disparity Map Computed using Local Stereo Parametric and Non-Parametric Methods. *International Conference on Automatic Control and Dynamic Optimization Techniques, ICACDOT 2016*, 104–109 2017.

Taniai, T., Matsushita, Y., Sato, Y., and Naemura, T., 2018. Continuous 3D Label Stereo Matching Using Local Expansion Moves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11), pp.2725–2739.

Tatar, N., Arefi, H., and Hahn, M., 2021. High-Resolution Satellite Stereo Matching by Object-Based Semiglobal Matching and Iterative Guided Edge-Preserving Filter. *IEEE*

Geoscience and Remote Sensing Letters, 18(10), pp.1841–1845.

Ttofis, C., Kyrkou, C., and Theocharides, T., 2016. A Low-Cost Real-Time Embedded Stereo Vision System for Accurate Disparity Estimation Based on Guided Image Filtering. *IEEE Transactions on Computers*, 65(9), pp.2678–2693.

Valentin, J., et al., 2018. Depth from Motion for Smartphone AR. *ACM Transactions on Graphics (ToG)*, 37(6), pp.1–19.

Wang, G., Liu, Y., Xiong, W., and Li, Y., 2018. An Improved Non-Local Means Filter for Color Image Denoising. *Optik*, (173), pp.157–173.

Wang, W., Yan, J., Xu, N., Wang, Y., and Hsu, F.H., 2015. Real-Time High-Quality Stereo Vision System in FPGA. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(10), pp.1696–1708.

Wang, X., and Liu, Y., 2015. Accurate and Fast Convergent Initial-Value Belief Propagation for Stereo Matching. *PLoS ONE*, 10(9), p.e0137530.

Wang, X., and Xie, B., 2018. Full-Image Guided Method Based on Confidence Coefficient for Fast Stereo Matching. *Journal of Physics: Conference Series*, 012134 2018.

Wang, Z., Zhu, S., Li, Y., and Cui, Z., 2016. Convolutional Neural Network Based Deep Conditional Random Fields for Stereo Matching. *Journal of Visual Communication and Image Representation*, 40, pp.739–750.

Wedel, A., Cremers, D., Wedel, A., and Cremers, D., 2011. *Stereo Scene Flow for 3D Motion Analysis*, Springer Science & Business Media.

Wei, S., Xinyu, W., Minghua, Z., and Qi, H., 2021. Stereo Matching Based on Improved Cost Calculation and a Disparity Candidate Strategy. *Laser and Optoelectronics Progress*, 58(2), pp.0215001-1-0215001–14.

Williem, and Park, I.K., 2018. Deep Self-Guided Cost Aggregation for Stereo Matching. *Pattern Recognition Letters*, 112, pp.168–175.

Wu, W., Zhu, H., Yu, S., and Shi, J., 2019. Stereo Matching with Fusing Adaptive Support Weights. *IEEE Access*, 2019(7), pp.61960–61974.

Wu, W., Zhu, H., and Zhang, Q., 2019. Oriented-Linear-Tree Based Cost Aggregation for Stereo Matching. *Multimedia Tools and Applications*, 78(12), pp.15779–15800.

Xiao, Y., Xu, D., Wang, G., Hu, X., Zhang, Y., Ji, X., and Zhang, L., 2020. Confidence Map Based 3D Cost Aggregation with Multiple Minimum Spanning Trees for Stereo Matching. *Asian Conference on Pattern Recognition*, 355–365 2020. Springer, Cham.

Xu, H., Chen, X., Liang, H., Ren, S., Wang, Y., and Cai, H., 2020. CrossPatch-Based Rolling Label Expansion for Dense Stereo Matching. *IEEE Access*, 8, pp.63470–63481.

Xu, H., and Zhang, J., 2020. AANet: Adaptive Aggregation Network for Efficient Stereo Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1959–1968 2020.

Xu, J., Li, Q., Luo, Y., Zhou, Y., and Wang, J., 2020. State Measurement of Isolating Switch using Cost Fusion and Smoothness Prior Based Stereo Matching. *International Journal of Advanced Robotic Systems*, 17(3), p.1729881420925299.

Xu, S., Zhang, F., He, X., Shen, X., and Zhang, X., 2015. PM-PM: Patchmatch with Potts Model for Object Segmentation and Stereo Matching. *IEEE Transactions on Image Processing*, 24(7), pp.2182–2196.

Xu, Y., Xu, X., and Yu, R., 2019. Disparity Optimization Algorithm for Stereo Matching using Improved Guided Filter. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 23(4), pp.625–633.

Xue, T., Owens, A., Scharstein, D., Goesele, M., and Szeliski, R., 2019. Multi-Frame Stereo Matching with Edges, Planes and Superpixels. *Image and Vision Computing*, 91, p.103771.

Yang, F., Sun, Q., Jin, H., and Zhou, Z., 2020. Superpixel Segmentation with Fully Convolutional Networks. *Proceedings of the IEEE/CVF conference on computer vision and*

pattern recognition, 13964–13973 2020.

Yang, G., Manela, J., Happold, M., and Ramanan, D., 2019. Hierarchical Deep Stereo Matching on High-Resolution Images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 5515–5524 2019.

Yang, M., Liu, Y., and You, Z., 2017. The Euclidean Embedding Learning Based on Convolutional Neural Network for Stereo Matching. *Neurocomputing*, 2017(267), pp.195–200.

Yang, W.-J., Tsai, Z.-S., Chung, P.-C., and Cheng, Y.-T., 2019. An Adaptive Cost Aggregation Method Based on Bilateral Filter and Canny Edge Detector with Segmented Area for Stereo Matching. *Proceedings Volume 11049, International Workshop on Advanced Image Technology (IWAIT) 2019*, 110491J (2019) 2019. SPIE.

Yang, X., Feng, Z., Zhao, Y., Zhang, G., and He, L., 2022. Edge Supervision and Multi-Scale Cost Volume for Stereo Matching. *Image and Vision Computing*, 117(January 2022), pp.104336(1)-104336(10).

Yao, G., Yilmaz, A., Meng, F., and Zhang, L., 2021. Review of Wide-Baseline Stereo Image Matching Based on Deep Learning. *Remote Sensing*, 13(16), p.3247.

Yao, P., and Feng, J., 2021. Ensemble Learning with Advanced Fast Image Filtering Features for Semi-Global Matching. *Machine Vision and Applications*, 32(4), p.83.

Yao, P., Zhang, H., Xue, Y., and Chen, S., 2019. As-Global-as-Possible Stereo Matching with Adaptive Smoothness Prior. *IET Image Processing*, 13(1), pp.98–107.

Yao, S.J., Wang, L.H., Lin, C.L., and Zhang, M., 2018. Real-Time Stereo to Multi-View Conversion System Based on Adaptive Meshing. *Journal of Real-Time Image Processing*, 14(2), pp.481–499.

Ye, X., Gu, Y., Chen, L., Li, J., Wang, H., and Zhang, X., 2017. Order-Based Disparity Refinement Including Occlusion Handling for Stereo Matching. *IEEE Signal Processing*

Letters, 24(10), pp.1483–1487.

Ye, X., Li, J., Wang, H., Huang, H., and Zhang, X., 2017. Efficient Stereo Matching Leveraging Deep Local and Context Information. *IEEE Access*, 5, pp.18745-18755.

Yin, J., Zhu, H., Yuan, D., and Xue, T., 2017. Sparse Representation Over Discriminative Dictionary for Stereo Matching. *Pattern Recognition*, 71, pp.278–289.

Yousif, A.N., Ibrahim, H.M., Alwan, S.J., and Sh, M., 2022. Stereo Vision Development for High Performance on Stereo Systems. *International Journal of Nonlinear Analysis and Applications (IJNAA)*, 13(December 2021), pp.2731–2738.

Yuan, W., Meng, C., Tong, X., and Li, Z., 2021. Efficient Local Stereo Matching Algorithm Based on Fast Gradient Domain Guided Image Filtering. *Signal Processing: Image Communication*, 95, p.116280.

Yue, X., Wang, F., Guo, B., Xu, P., and Shi, J., 2018. Disparity Map Optimization Based on Edge Detection. *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, 3311–3315 2018.

Zbontar, J., and Lecun, Y., 2016. Stereo Matching by Training a Convolutional Neural. *Journal of Machine Learning Research*, 17(1), pp.2287–2318.

Zeglazi, O., Rziza, M., Amine, A., and Demonceaux, C., 2018. A Hierarchical Stereo Matching Algorithm Based on Adaptive Support Region Aggregation Method. *Pattern Recognition Letters*, 112, pp.205–211.

Zeng, L., and Tian, X., 2022. CRAR: Accelerating Stereo Matching with Cascaded Residual Regression and Adaptive Refinement. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(3), pp.1–19.

Zha, D., Jin, X., and Xiang, T., 2016. A Real-Time Global Stereo-Matching on FPGA. *Microprocessors and Microsystems*, 47, pp.419–428.

Zhan, Y., Gu, Y., Huang, K., Zhang, C., and Hu, K., 2016. Accurate Image-Guided Stereo

Matching with Efficient Matching Cost and Disparity Refinement. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(9), pp.1632–1645.

Zhang, C., He, C., Chen, Z., Liu, W., Li, M., and Wu, J., 2019. Edge-Preserving Stereo Matching Using Minimum Spanning Tree. *IEEE Access*, 7, pp.177909–177921.

Zhang, C., Li, Z., Cheng, Y., Cai, R., Chao, H., and Rui, Y., 2015. Meshstereo: A Global Stereo Model with Mesh Alignment Regularization for View Interpolation. *Proceedings of the IEEE International Conference on Computer Vision*, 2057–2065 2015.

Zhang, C., Wu, J., Chen, Z., Liu, W., Li, M., and Jiang, S., 2021. Dense-CNN: Dense Convolutional Neural Network for Stereo Matching using Multiscale Feature Connection. *Signal Processing: Image Communication*, 95, p.116285.

Zhang, F., and Wah, B.W., 2018. Fundamental Principles on Learning New Features for Effective Dense Matching. *IEEE Transactions on Image Processing*, 27(2), pp.822–836.

Zhang, H., Shen, H., Yuan, Q., and Guan, X., 2022. Multispectral and SAR Image Fusion Based on Laplacian Pyramid and Sparse Representation. *Remote Sensing*, 14(4), p.870.

Zhang, J., and Huang, J., 2021. An Improved Algorithm for Disparity Estimation of SGM Stereo Matching Based on Edge Detection. *Journal of Physics: Conference Series*, 2010(012037), pp.1–7.

Zhang, J., Liu, Z., Nezan, J.F., and Zhang, G., 2018. Correspondence Matching Among Stereo Images with Object Flow and Minimum Spanning Tree Aggregation. *International Journal of Advanced Robotic Systems*, 15(2), p.1729881418760986.

Zhang, X.H., Li, G., Li, C. Le, Zhang, H., Zhao, J., and Hou, Z.X., 2015. Stereo Matching Algorithm Based on 2D Delaunay Triangulation. *Mathematical Problems in Engineering*, 2015, pp.1–8.

Zhang, Y., Khamis, S., Rhemann, C., Valentin, J., Kowdle, A., Tankovich, V., Schoenberg, M., Izadi, S., Funkhouser, T., and Fanello, S., 2018. Activestereonet: End-to-End Self-

Supervised Learning for Active Stereo Systems. *Proceedings of the European Conference on Computer Vision (ECCV)*, 784–801 2018.

Zhang, Y., Li, Y., Kong, Y., and Liu, B., 2020. Attention Aggregation Encoder-Decoder Network Framework for Stereo Matching. *IEEE Signal Processing Letters*, 27, pp.760–764.

Zhang, Z., Wang, Y., Huang, T., and Zhan, L., 2020. A Weighting Algorithm Based on the Gravitational Model for Local Stereo Matching. *Signal, Image and Video Processing*, 14(2), pp.315–323.

Zhao, X., Zheng, R., Ye, W., and Liu, Y., 2019. A Robust Stereo Semi-direct SLAM System Based on Hybrid Pyramid. *IEEE International Conference on Intelligent Robots and Systems*, 5376–5382 2019.

Zhou, B., Qin, L., and Gong, W., 2019. Stereo-Matching Algorithm using Weighted Guided Image Filtering Based on Laplacian Of Gaussian Operator. *Laser and Optoelectronics Progress*, 56, pp.226–232.

Zhou, K., Meng, X., and Cheng, B., 2020. Review of Stereo Matching Algorithms Based on Deep Learning. *Computational Intelligence and Neuroscience*, 2020, pp.1–12.

Zhou, Z., Shen, J., Han, P., and Jiang, J., 2020. Stereo Matching Algorithm Based on Census Transformation and Guided Filter. *Journal of Applied Optics*, 41(1), pp.79–85.

Zhou, Z., Wu, D., and Zhu, Z., 2016. Stereo Matching using Dynamic Programming Based on Differential Smoothing. *Optik*, 127(4), pp.2287–2293.

Zhu, C., and Chang, Y.Z., 2019. Efficient Stereo Matching Based on Pervasive Guided Image Filtering. *Mathematical Problems in Engineering*, 2019, pp.1–11.

Zhu, C., and Chang, Y.Z., 2019. Hierarchical Guided-Image-Filtering for Efficient Stereo Matching. *Applied Sciences (Switzerland)*, 9(15), p.3122.

Zhu, R.H., Ge, G.Y., Zhang, G.S., Shen, Z., and Sun, Q., 2019. Semi-Global Stereo Matching Algorithm Based on AD-Census Transform and Multi-Scan Line Optimization.

Guangdianzi Jiguang/Journal of Optoelectronics Laser, 38(5), pp.12–19.

Zhu, S., Xu, H., and Yan, L., 2019. A Stereo Matching and Depth Map Acquisition Algorithm Based on Deep Learning and Improved Winner Takes All-Dynamic Programming. *IEEE Access*, 7, pp.74625–74639.

Zhu, S., and Yan, L., 2017. Local Stereo Matching Algorithm with Efficient Matching Cost and Adaptive Guided Image Filter. *Visual Computer*, 33(9), pp.1087–1102.

Zhu, Z.J., and Dai, Q.Y., 2017. Hybrid Scheme for Accurate Stereo Matching. *Neurocomputing*, 252, pp.24–33.

APPENDIX A

Disparity Selection and Optimization

Table A.1 shows the raw data at each disparity value of the Adirondack image at cost aggregation step (i.e., Stage 2) with the maximum disparity 73. The WTA selects the minimum raw data and choose the designated disparity value.

Table A.1: The Raw Data of Cost Aggregation with Disparity Values at the Coordinates of (709,496) until (718,496).

Pixel Location \ Disparity Range	1	2	3	4	5	6	7		70	71	72	73
709 x 496	0.012956	0.013963	0.016548	0.018109	0.016362	0.016805	0.014095	→	0.01804	0.01926	0.0193	0.0167
710 x 496	0.013856	0.014758	0.016224	0.014947	0.014843	0.018136	0.014674		0.01868	0.01823	0.01867	0.0186
711 x 496	0.014244	0. 014758	0.019024	0.018763	0.018696	0.018693	0.018731		0.01916	0.01926	0.0193	0.01935
712 x 496	0.01424	0.0146	0.0178	0.018763	0.015336	0.015336	0.018731		0.01915	0.01926	0.0193	0.01935
713 x 496	0.014283	0.01476	0.019025	0.016524	0.016457	0.016457	0.01867	→	0.01916	0.01814	0.0193	0.01879
714 x 496	0.014207	0.014742	0.016229	0.015969	0.018702	0.018702	0.018144		0.01728	0.01511	0.01782	0.0149
715 x 496	0.014006	0.014591	0.016141	0.015219	0.016304	0.016304	0.018746		0.01482	0.01505	0.01523	0.01869
716 x 496	0.014447	0.014781	0.019048	0.018187	0.015111	0.015111	0.014645	→	0.01919	0.01928	0.01933	0.01937
717 x 496	0.014231	0.014806	0.015158	0.014904	0.014837	0.014873	0.014873		0.0153	0.01539	0.01543	0.01547
718 x 496	0.013388	0.014854	0.015213	0.014963	0.014896	0.014931	0.014931		0.01536	0.01548	0.01548	0.01551

APPENDIX B

Middlebury Datasets Training and Testing Qualitative Performance

Image	Res (D_{max})	Left image	Ground truth	Result	*all* error	*nonocc* error
Adirondack	718 x 49 (73)					
ArtL	347 x 277 (64)					
Jadeplant	659 x 497 (160)					
Motorcycle	741 x 497 (70)					
MotorcycleE	741 x 497 (70)					
Piano	707 x 481 (65)					
PianoL	707 x 481 (65)					
Pipes	735 x 485 (75)					
Playroom	699 x 476 (83)					
Playtable	699 x 476 (83)					
PlaytableP	699 x 476 (83)					
Recyle	720 x 486 (65)					
Shelves	738 x 497 (60)					
Teddy	450 x 375 (64)					
Vintage	722 x 480 (190)					

Figure B.1: The Result of the Middlebury Training Dataset

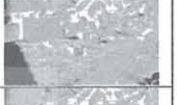
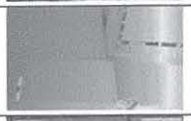
Image	Res (D_{max})	Left image	Ground truth	Result	*all* error	*nonocc* error
Australia	715 x 492 (73)					
AustraliaP	715 x 492 (73)					
Bicycle2	713 x 488 (63)					
Classroom2	750 x 474 (153)					
Classroom2E	750 x 474 (153)					
Computer	322 x 277 (64)					
Crusade	720 x 474 (200)					
CrusadeP	720 x 474 (200)					
Djembe	719 x 494 (80)					
DjembeL	719 x 494 (80)					
Hoops	721 x 498 (103)					
Livingroom	742 x 496 (80)					
Newkuba	701 x 487 (143)					
Plants	710 x 496 (80)					
Stairs	690 x 468 (113)					

Figure B.2: The Result of the Middlebury Test Dataset

Image name	Left image	Ground truth	Disparity map result
artroom1			
artroom2			
bandsaw1			
bandsaw2			
chess1			
chess2			
chess3			
curule1			
curule2			
curule3			
ladder1			
ladder2			
octogons1			
octogons2			
pendulum1			

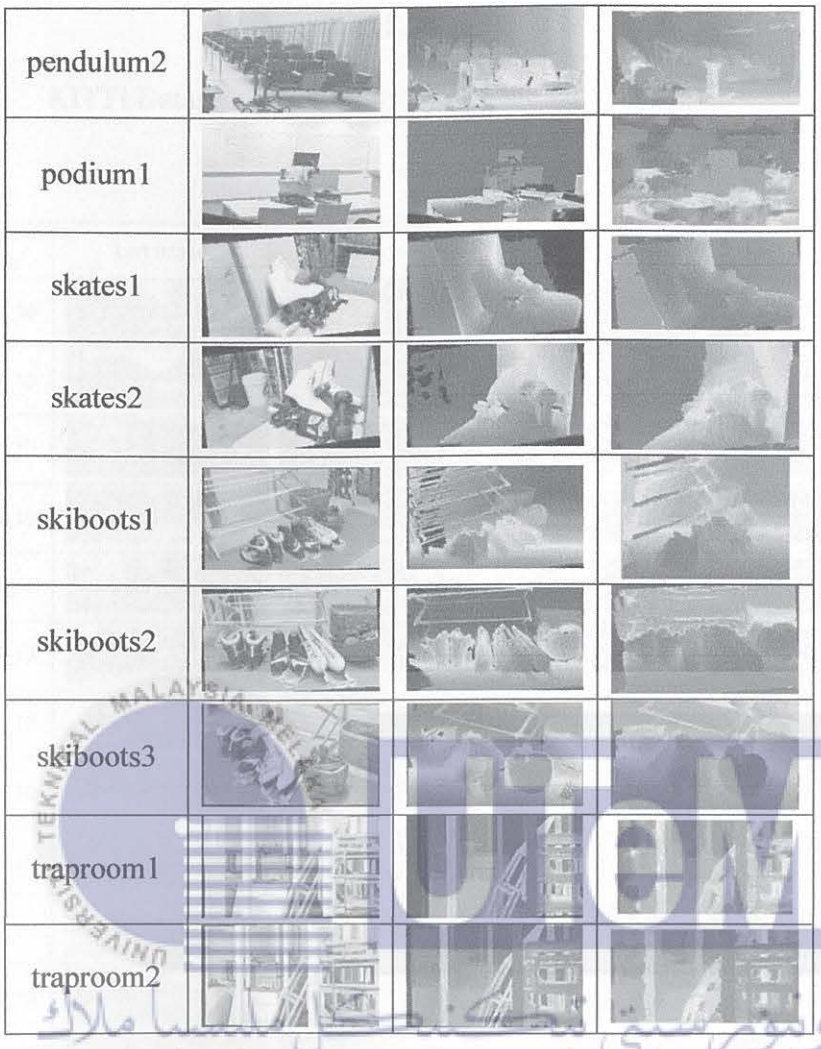


Figure B.3: The Result of the Middlebury Mobile Dataset

APPENDIX C

KITTI Datasets Training and Test Qualitative Performance

Image number	Left image	Ground truth	Result color	Error map
000000_10				
000001_10				
000002_10				
000003_10				
000004_10				
000005_10				
000006_10				
000007_10				
000008_10				
000009_10				
000010_10				
000011_10				
000012_10				
000013_10				
000014_10				
000015_10				
000016_10				
000017_10				
000018_10				
000019_10				

Figure C.1: The Result of KITTI Training Dataset

Image number	Left image	Result	Error map	*all* error (%)	*nonocc* error (%)
000000_10				10.18	9.41
000001_10				5.52	4.88
000002_10				7.73	6.68
000003_10				8.88	8.26
000004_10				9.74	8.48
000005_10				18.06	16.52
000006_10				14.55	13.75
000007_10				7.13	6.52
000008_10				5.68	5.68
000009_10				6.71	6.14
000010_10				5.36	4.81
000011_10				3.76	3.21
000012_10				2.38	1.70
000013_10				2.56	2.03
000014_10				3.05	2.44
000015_10				5.88	4.87
000016_10				6.97	6.31
000017_10				3.23	2.32
000018_10				11.06	10.58
000019_10				3.76	2.90

Figure C.2: The Result of KITTI Testing Dataset