UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEPTH MAP RECONSTRUCTION FROM STEREO IMAGE USING SUM OF ABSOLUTE DIFFERENCES

This report is submitted in accordance with the requirement of the Universiti Teknikal Malaysia Melaka (UTeM) for the Bachelor of Electronic Engineering Technology (Telecomunication) with Honours.

by

EFFY SHAZRIN BIN SABTUHE

FACULTY OF ELECTRICAL AND ELECTRONIC ENGINEERING TECHNOLOGY

2021

DECLARATION

I hereby, declared this report entitled Depth Map Reconstruction From Stereo Images Using Sum Of Absolute Differences is the results of my own research except as cited in references.

Signature: ....

Author : EFFY SHAZRIN BIN SABTUHE

Date: 29/03/2021

APPROVAL

This report is submitted to the Faculty of Electrical and Electronic Engineering Technology of Universiti Teknikal Malaysia Melaka (UTeM) as a partial fulfilment of the requirements for the degree of Bachelor of Electronics Engineering Technology (Telecommunications) with Honours. The member of the supervisory is as follow:

Signature: ....

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

ABSTRACT

Projek ini menyediakan kaedah untuk menyelesaikan masalah korespondensi ketika memadankan gambar stereo menggunakan algoritma Sum of Absolute Differences (SAD). Perisian MATLAB menyediakan alat tersebut. Pemetaan perbezaan dihasilkan melalui algoritma pemadanan blok Sum of Absolute Differences (SAD). Konvensional SAD biasanya diimbas di seluruh gambar stereo untuk mengetahui perbezaan garis antara gambar yang ditangkap di kiri dan kanan, dan kemudian memperoleh peta perbezaan yang setara, yang boleh mengakibatkan masa berlalu yang tinggi. Terdapat empat langkah asas dalam kaedah penglihatan stereo untuk mencitrakan pembinaan semula. Biasanya terdapat empat fasa ini. Penyelewengan gambar yang diambil dari lensa kamera diekstrak terlebih dahulu pada langkah penyimpangan. Langkah seterusnya adalah menyesuaikan jarak dan sudut ketinggian antara kedua gambar kamera untuk menentukan panjang fokus dan sumbu epipolar, seperti yang diduga di luar. Perbandingan antara gambar kiri dan kanan akan dihitung dalam tahap korespondensi dan digunakan untuk mengukur peta perbezaan. Kaedah ini sering dikenali sebagai saling beroperasi. Projek ini dapat mengembangkan algoritma pemadanan stereo menggunakan jumlah perbezaan mutlak untuk melakukan penyusunan peta mendalam dari algoritma yang dicadangkan dan algoritma stereo sederhana mengira hasil yang setanding dengan canggih terkini di penanda aras Middlebury. Kaedah projek ini menggunakan antara muka pengguna grafik yang dikembangkan dari persekitaran MATLAB yang menghitung pemetaan perbezaan dengan menggunakan jumlah perbezaan mutlak. Projek ini telah memenuhi semua objektif saya dan hasil yang saya dapat untuk mencapai penanda aras yang ditetapkan. Dalam pelbagai domain, seperti penjejakan, navigasi robot, dll., Algoritma ini dapat digunakan dengan lebih baik. Tidak ada petunjuk kedalaman yang lebih baik atau sangat diperlukan daripada petunjuk kedalaman yang lain. Petunjuk mempunyai kelebihan dan kekurangan tersendiri.

ABSTRACT

This project provides a method for solving the issue of correspondence when matching the stereo image using the algorithm Sum of Absolute Differences (SAD). MATLAB software provides the tool. The mapping of disparities is produced through the block matching algorithm Sum of Absolute Differences (SAD). The SAD conventional is usually scanned in entire stereo images to figure out the discrepancy in lines between the captured images on the left and right, and then obtains the equivalent map of disparity, which can result in a high elapse time. There are four basic steps in a stereo vision method for imaging the reconstruction. There are typically four phases of this The distortion of the captured images from the camera lens is extracted first in the un distortion step. The next step is to adjust the distance and the elevation angle between the two camera images to determine the focal length and the epipolar axis, as presumed beyond. The comparisons inter the left and the right image will be calculated in the correspondence stage and used to measure the map of disparities. This method is often known as being interoperable. This project was able to develop stereo matching algorithm using sum of absolute differences to do the depth map reconstruction from the proposed algorithm and a simple stereo algorithm compute results comparable to current state-of-the-art on Middlebury benchmark. The method of this project is using a graphical user interface that developed from MATLAB environment that compute the disparity mapping by using sum of absolute differences. This project has answered all my objectives and the results I got to achieve the set benchmark. In various domains, such as tracking, robot navigation, etc., these algorithms can be better used. There is no better or indispensable depth cue than another depth cue. A cue has its own perks and disadvantages.

DEDICATION

I would like to dedicated and special thanks to

My beloved father and mother,

To my beloved family, my respected lecturer and fellow friends

And for the rest, might Allah have blessed you

Thanks for all the guided and support

ACKNOWLEDGEMENTS

First and foremost, I am very grateful to the almighty ALLAH S.W.T for letting me to finish my Final Year Project 1 and give me strength that I need in order to fulfill my duty as an Electronic student.

Special thanks to my supervisor Ts. Dr. Rostam Affendi Bin Hamzah for the guidelines and teach me during completing my final year project for duration 1 year. May Allah always blessed you and I never forget your advice and guidance during I make a mistake.

Thanks also to all of the kind lecturers in Bachelor of Electronic Engineering Technology (Telecommunication) with Honours section for their accommodation, suggestion and opinion during the project progress in university. In particular, I would like to thank all the staff and technicians, for their cooperation, indirect or directly contribution upon completing my project.

Most importantly, special thanks also to my family for their external support when I told them I wanted to continue my education and especially my friends for encouraging me to finish this project. Without their support the ideas could not have been realized.

TABLE OF CONTENTS

TABLE OF CONTENTS	PAGES
CHAPTER 1 : INTRODUCTION	1
1.0 Introduction	1
1.1 Background	1 -2
1.2 Problem Statement	3
1.3 Objective	4
1.4 Project Scope	4
1.5 Project Outlines	5
CHAPTER 2 : LITERATURE REVIEW	6
2.0 Introduction	6
2.1 Related Work	6
2.1.1 Depth Estimation in Disparity Mapping	6-8
2.1.2 Stereo Matching Based On Sad	9-12
2.1.3 3D Reconstruction	13-14
2.2 Previous Work	15
2.2.1 Stereo Image and Depth Map Generation for Images with Different Views and Resolutions	15-16
2.2.2 Improve Accuracy of Disparity Map for Stereo Images Using Sift and Weighted Color Model	17
2.2.3 Creation of A Depth Map from Stereo Images of Faces for 3d Model Reconstruction	18 -19
2.2.4 Computer Vision Based Distance Measurement System using Stereo Camera View	19-120

2.2.5 Disparity map estimation with deep learning in stereo vision 21-22
2.2.6 Efficient Binocular Stereo Matching Based On Sad and Improved Census Transformation 23
2.2.7 3D Distance Measurement Accuracy On Low – Cost Strereo Camera 24
CHAPTER 3: METHODOLOGY 25
3.1 Introduction 25
3.2 Project Overview 25
3.3 Flowchart Represents Process Of Project 26
3.4 The Report Flowchart 27
3.5 Block Diagram Of Project 28
3.6 Program Development 28
3.7 Software Implementation 29
3.8 Stereo Vision Principle Depth Map Generation 30-31
3.9 Sum of Absolute Differences (Sad) 32
3.91 The Examples of Sad in Matlab 33
CHAPTER 4: RESULT AND DISCUSSION 34
4.1 Introduction 34
4.2 Test platform 34
4.3 Test Data 35-36
4.4 A GUI demonstrates an efficient streo matching algorithm 37-39
4.5 Depth mapping propagation 40-45
4.6 Summary 46
CHAPTER 5: CONCLUSION AND FUTURE WORK
5.1 Conclusion 47
5.2 Recommendation for future work 48

REFERENCES 49-52

APPENDIX 53

LIST OF FIGURE

FIGURE	TITLE	PAGES
Figure 2.1	(a) combined images; (b) depth map; (c) 3D model of the depth map.	8
Figure 2.2	Comparisons of stereo matching	11
Figure 2.3	Standard images for benchmarking provided in Middlebury Stereo Vision databases in (Szeliski, 2015).	12
Figure 2.4:	Signal flow of a new system for the formation of stereo images and depth maps.	15
Figure 2.5	Procedures in stereo image formation	16
Figure 2.6	Circle indicates bounded area suggested by SIFT. Patch window (M). Square indicates "disparity max" area.	17
Figure 2.7	Camera calibration using a chessboard.	18
Figure 2.8	Calculating the depth map. (T – Stereo base, Z – distance)	19
Figure 2.9	Block diagram of stereo vision system established	20
Figure 2.10	Computing inequality between two points on disparity maps	20
Figure 2.11	Components of a typical convolutional network	21
Figure 2.12	Architecture of the convolutional network	22
Figure 2.13	Flowchart of stereo matching.	23
Figure 2.14	Callibaration pattern: (a) Chessboard (b) Feature Descriptor	24
Figure 3.1	Project development flowchart	26

Figure 3.2	Shows the report flowchart	27
Figure 3.3	Block diagram of project.	28
Figure 3.4	MATLAB	29
Figure 3.5	Two cameras' location and their projection planes image.	30
Figure 3.6	Relationship between disparity and depth.	31
Figure 3.7	Tsukaba stereo image pair with true disparity map (a), (b), and (c).	33
Figure 4.1	a) Tsukuba b) Venus c) Cone d) Teddy	36
Figure 4.2	Graphical user interface for background streo vision	37
Figure 4.3	View images from the GUI	38
Figure 4.4	a) Streo matching b) Depth mapping reconstruction	39
Figure 4.5	Result of the algorithm	41
Figure 4.6	Comparison Disparity with difference window size	43
Figure 4.7	3D reconstruction	45

CHAPTER 1

INTRODUCTION

1.0 Introduction

This chapter provides a summary of using sum of absolute differences (SAD) in determining the disparity mapping. The problem background and problem statement are described next. This is followed by the objective and the focuses of this research which involves the study of depth map reconstruction from stereo images utilise sum of absolute differences (SAD).

1.1 Background

Matching stereo is an issue in finding correspondence between two input pictures. This is one of the fundamental issues of computer vision for a wide variety of applications and has thus been thoroughly studied in the area of computer vision for many years. For each point in the left image, stereo matching consists of finding its corresponding in the right one. The dissimilarity between those points' horizontal distances is the disparity. A map of disparity consists of all attribute outcomes of disparity within an image. Such a map is essentially a representation of the perceived scene profundity. Hence, the maps of inequalities were used to resolve issues effectively.

There are three specific groups of techniques to match the stereo: area-based, phase-based, and feature-based techniques. The most beneficial area-based techniques in real-time stereo vision are SAD-based implementations because they can be executed directly in hardware. The computations needed in tenure of design units are readily, as there are only summaries and absolutism values. Simultaneous design units may be used to manage different ranges of disparities and to reduce the required computational time.

In general, it is essential to have at least two images for reconstruction of an object, please consider the relevant points on these images and to construct a map of depth. The quality of the acquired images can be determined by several factors: lighting, object shape, camera angle and camera focal length. Many stereo matching algorithms are based on the constraints of similarities, epipolarity, consistency, continuity and ordering and taking the basic process: matching cost calculation; matching cost aggregation; variance calculation; refinement of differences. For the images on the left and right, the matching expense is determined by the difference in the gray values of the corresponding pixels.

Traditional stereo matching methods, such as Sum of Square Differences (SSD) and Sum of Absolute Differences (SAD), They are designed for simplistic scenes and cannot handle less textured images, and are prone to light and noise variations. On the other hand, normalized cross-correlation (NCC) can avoid the noise better, and light variants would not be impacted, but it is a computation-intensive approach. Matching stereo consists of finding the corresponding in the right image for every juncture in the left image. The difference between all points in horizontal intervals is the variance. A map of disparity consists of all the probable values of disparity within an image. A rather map is essentially a depiction of the perceived scene size. Although this technique is successful in images without texture, it varies depending on centralized sample pixel, and therefore the corresponding results can be reduced if the noise affects the central pixel.

1.2 Problem Statement

Computer vision is actually a vital field of study. This requires techniques for the acquisition, processing, interpretation and comprehension of images. Computer vision approaches seek to use different mathematical methods to model a complex visual environment. Some of the objectives of computer vision are to describe the world we see on the basis with one or even more images and to restructure its properties, including the distribution of light, colour and shape. Stereo vision is a computer vision field that tackles a major research issue: Tri-dimensional point reconstruction maps the depth estimate.

A stereo vision system is made up of a stereo camera, including two cameras located horizontally (one on the left and one on the right). Then, the two captured images simultaneously by such cameras are analyzed to retrieve visual depth information. The challenge is to identify the correct method for estimating the discrepancies between the views displayed in the two images to map (e.g. plot) the environment's correspondence (e.g. disparity). Intuitively, a map of disparities represents the equivalent pixels, which are horizontally moved between the left and right images. Every year, new methods and techniques are developed to solve this problem, and show a trend towards improving accuracy and time consumption.

1.3 Objective

The objective of this project comes from some short of research and problem statement. This objective showed the reason and the outcomes of this project. The objective had state below:

- To develop stereo matching algorithm using Sum of Absolute Differences (SAD).
- To reconstruct depth map from the proposed algorithm.
- To analyze the performance of the proposed algorithm using Standard benchmarking evaluation system.

1.4 Project Scope

The aim of stereo matching is to draw distinctions from two very similar perspectives. Current stereo matching methods can induce incorrect matching when there are loads of disruptions in image noise and disparity. This project to achieve disparity (difference) maps is a set of extraordinary issues which remain to be addressed in stereo vision systems. Various techniques are used to provide maps of disparities.

Commonly preferred is the method that is based on stage geometry that is given by matching pixels between two images. The SAD algorithm attempts to modify the manner in which the disparity is measured by using the edge information derived from the captured stereo images. While less computational loads are needed by the edge operator used in the proposed process, the disparity map may be depicted with a low elapsing period while retaining the equal efficacy of the reconfiguration map. This tool is provided by MATLAB software.

1.5 Project Outlines

In chapter 1, it will be explaining briefly about the possibility of the project. Tasks background will be discussed in this part. This part will concentrate on the outlines of the outline of the undertaking, specifying the objective, the issue articulation, and the scope of the project.

In chapter 2, this section is about the idea, hypothesis, and some characteristic of equipment that utilized as a part of this task. This part also contains a meaning of term used as a part of this undertaking and furthermore discusses about the idea of the research and how it identified with the theory.

In chapter 3, this section will explain about the methodology. Methodology chapter is a step that need to be follow and detailed reports of studies that need to be complete to achieve the objective. This chapter describes the methodology required to complete the project and discusses the creation of the project.

In chapter 4, this section is about the result and discussion. It will explain the data that has been collected after the project has been completed. The comparison of the data will be discussed. The testing, implementing and troubleshooting throughout this project will be explained in this chapter.

In chapter 5, this section is about conclusion. It will explain the completed project either the objective is achieved or not. Recommendation and future improvement also can be suggested and discussed in this chapter.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

The chapter will address the literature review where information on the relevant project is based. It will cover both the concept based on the previous project and the theory to achieve this project. This study focuses on the algorithm matching the Stereo and the method used for this related project.

2.1 Related Work

2.1.1 Depth Estimation in Disparity Mapping

In recent decades' numerous algorithms and applications have been presented to estimate disparity maps. According to (A.Qayyum, 2015) in method for monitoring trees and vegetation is proposed, based on satellite images. The stereo matching algorithms are calculated based on stereo satellite images to gage disparity maps. The estimate to depth map of the level of the trees and plants at the base points is inversely rate able to the level of disparity. Dynamic programming (DP) and block matching with energy minimization was proposed for the calculation of the depth map.

The pedestrian detection method is proposed in (Dongyoung, 2017) focused on a complex map of inequalities for Smarter Vehicles. The dense map of disparities is used to make pedestrian identification more effective. The method consists of several phases, the identification of hazard areas using information on road and column identification characteristics, the classification of pedestrian zones using segmentation based on complex maps with disparities and classification with pedestrians using the ideal criteria.

The work of (A. K. Wassim, 2016) is an object tracking investigation where a stereoscopic camera detects objects, make it a low cost solution for target detection. Its goal is to verify and track objects in a video frame throughout the video without previous knowledge of the objects. Calculate the map of disparities using a couple of stereo images. Instead, to detect object blobs, the map of disparities undergoes a profound segmentation and the corresponding area in rectified stereo image is the object of interest. Furthermore, in (Okae J. D., 2017) An effective optimization is given to increase the efficiency of the stereoscopic vision system and to precisely compare the measured map of disparities with the actual depth data.

With modern technology and artificial knowledge in a methodology (Lobaton Q. G., 2017) in a value multiples stereo disparity approach is proposed for the identification of sturdy obstacles for autonomous driving applications in outdoor scenes. The calculation of disparity is affected greatly by reflections, loss of texture and repeated patterns of objects. This can lead to misleading estimates, that can lead to some bias in approaches to obstacle detection that make use of the map of disparities. In order to address this problem, a new research is suggested instead of calculating the disparity of a single attribute, which uses the variety of candidates for each point of the picture. Based on a mathematical analysis distinguished by the success of various metrics, these are selected: the number and distance between the candidates according to the actual value of the disparity. It continues to create a location map that provides an estimate of the obstacles.

However, in (Z. Liang Y. F., 2018) explain the different phases in which the depth estimation is carried out: first, the extraction of functions, initial measurement and final optimization of the estimated depth. In (G. Song H. Z., 2017) suggest a new method of optimizations using tight smoothing constraints obtained in a neural network. The aim is to provide rigorous softening of the output disparity map. In this study, the first move was to question the architecture of CNN, called DD-CNN, to determine whether the disparities are discontinuous. Practice of this method was performed using real data from stereo data from Middlebury.

In the next stage, they identify an objective functions that consists of a given term using the (Zbontar J. a., 2015) method and a term designed to penalize disparities. These studies have found that the subnetwork framework produces a depth map through an architects that encode and decodes information based on DispNetCorr1D (N. Mayer E. l., 2016). Finally, a dual-structured network is built in (W. Luo, A. G. Schwing and R. Urtasun, 2016). The system requires an input image that moves through a finite range of layers followed by standardization and a linear rectified construct. For each layer, various filters were tested in their experiments, and the parameters were exchanged between both the two structures.

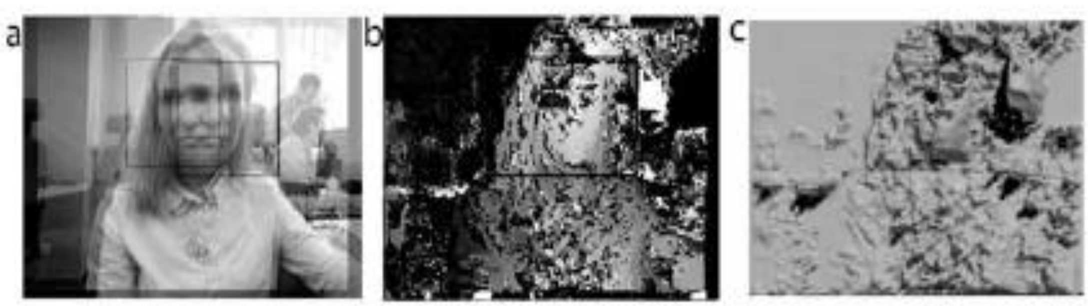


Figure 2.1: (a) combined images; (b) depth map; (c) 3D model of the depth map.

2.1.2 Stereo matching based on SAD

The stereo vision is the most important field of computer vision and it provides various algorithms for computing the map of disparity. Through using two stereo images, the scene Depth can be obtained from two separate points with some displaced values. The correlation values of the left image compared with the right image are the result of stereo matching. The photos depth map is determined using the disparity map or stereo matching. The stereo matching function for computing the exact depth map is very difficult. According to (P, 2017) proposed a cost aggregation approach based on the segment-tree for non-local stereo matching, leading to improvements in both the precision of disparities and the processing speed. In placing more emphasis, (K. Zhang Y. F., 2017) a cross-scale design was suggested to enhance the cost averaging for effective stereo matching. (Cigla, 2015) Recursive edge-aware (REAF) filters provided for precise and effective stereo matching.

In the case of the global stereo matching, the disparities are calculated by decreasing the global energy feature. An algorithm to identifying depth discontinuities from pairs of stereo images. Their approach can handle dynamic programming acceleration. A new stereo matching based on segmentation using graph cuts, which is used by assigning disparity planes to each segment to achieve the optimal solution. (Weijer, 2015) Suggested integration of cost allocation-filtering approaches and global energy minimization approaches to encourage increased stereo matching using a two-stage energy minimization algorithm using MRF modelling. Its solution can be used to successfully address the problem of stereo matching in occlusion areas.

In several practical applications (C. Kwan, 2018), It is essential that the images are synchronized with different views. According to various perspectives, the parallax poses severe problems in the synchronization of images, as most capture algorithms only refer to pictures with flattened material. Problem of registration for images having different content depths. This can be seen that the communication of features can either be done in the foreground or in the background, not both.

Its technique could be used to effectively address the issue of stereo matching in occluded regions. (LeCun, 2015) Used the Neural Network Convolution approach to predict matching image patches and calculated stereo matching costs, further optimized by cross-based expense agglomeration and sub-global matching. However, (W. Luo, 2016)proposed a deep learning network to produce accurate results efficiently on GPU. Use of the matching approach semi-global, the author (Pollefeys, 2017) presented an estimation method for learning based penalties to predict detailed estimates of dense disparity map.

While previous approaches can effectively yield precise disparities in stereo matching, It isn't easy to enforce them, and complex scenes may fail to implement them. In addition, the learning-based methods are not reliable, but are dependent on the training data. In this project, they suggest a stable and efficient SAD-based algorithm. My method is simple to apply and my results are somewhat close to those produced by the use of state-of-the-art models in the public data collection.

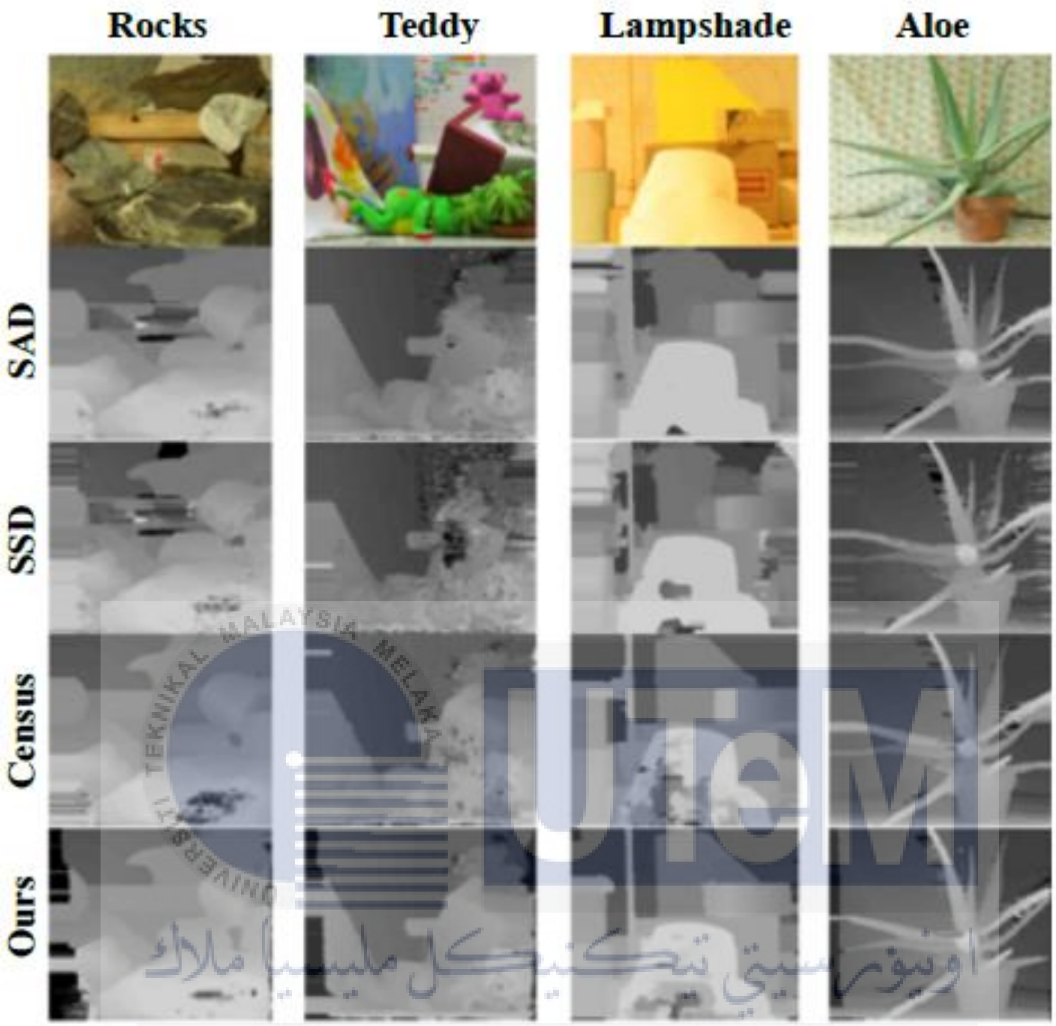


Figure 2.2: Comparisons of stereo matching

Figure 2.3: Standard benchmarking images provided in the Middlebury Stereo Vision datasets in (D. Scharstein and R. Szeliski, 2015).

2.1.3 3D Reconstruction

Single-view 3D Reconstruction Recovering 3D structure from single-view images is an ill-posed problem theoretically. Many attempts have been made to tackle this problem, such as ShapeFromX (Malik, 2015) (S. Savarese, 2007), where X may reflect silhouettes (E. Dibra, 2017), shading (Roth., 2015), and texture. However, in real-world situations, these methods are rarely applicable because they all involve strong presumptions and abundant skills in natural imagery (Y. Zhang, 2019).

3D-Vale-GAN (J. Wu, 2016) adopts GAN and VAE to create 3D structures with the performance of generative adversarial networks (GANs) (I. J. Goodfellow, 2014) and variational autoencoders (VAEs) by taking a single-view image as an input. For restoration, however, 3D-VAE-GAN includes class marks. By estimating depth, surface normal, and silhouettes of 2D images, MarrNet (J. Wu Y. W., 2017) reconstructs 3D objects, which is difficult and typically leads to significant distortion (Tulsiani., 2018). Octree is used by OGN (M. Tatarchenko, 2017) and O-CNN to represent higher resolution 3D volumetric objects with a limited memory budget. However, owing to the minimal relations between points, the points have a significant degree of independence in the point cloud representation. Consequently, these techniques do not reliably restore 3D quantities.

SfM and SLAM methods of multi-view 3D Reconstruction are successful in handling many situations. These methods match characteristics between images and estimate the pose of the camera for each image. However, when multiple viewpoints are separated by a large margin, the matching process becomes hard. In addition, it is sometimes impossible to scan all the surfaces of an object before reconstruction, leading to incomplete 3D shapes with occluded or hollowed-out areas (B. Yang, 2018).

Deep-learning based methods for 3D reconstruction have been suggested, powered by large-scale datasets of 3D CAD models. RNNs are used by both 3D-R2N2 and LSM to deduce 3D form from single or multiple input images and achieve impressive results. RNNs, even so, are time-consuming and permutation-variant, producing inconsistent results for reconstruction. To aggregate the features from multiple images, 3DensiNet (M. Wang, 2017) utilises max pooling. Max pooling, however, only extracts maximum values from features that can ignore other valuable characteristics that are useful for 3D reconstruction.

2.2 Previous Work

2.2.1 Stereo Image and Depth Map Generation for Images with Different Views and Resolutions

According to (Ayhan, 2018), some early findings in developing stereo images and depth maps of different resolution and view angles using uncalibrated images. Their purpose is to examine whether or not images can be combined with various resolutions and angles of view in order to produce high-quality stereo images and depth maps. Firstly, using a high-resolution image to improve the low-resolution picture (assuming it is the wrong image with no lack of generality) (assuming it is the right one). And merge the modified left image with the previous, high-resolution right image to generate stereo and depth maps of the same quality.

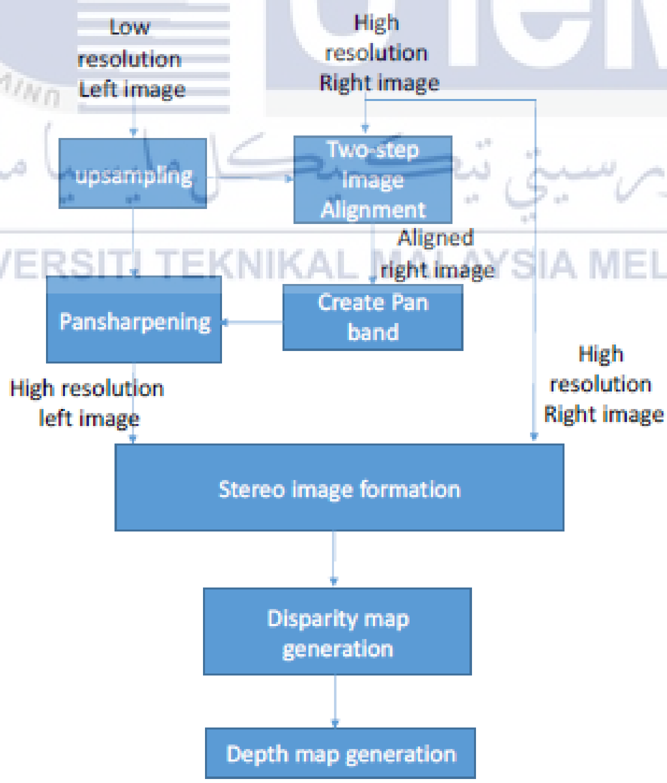


Figure 2.4: Signal flow of a new stereo image formation and depth map generation system.

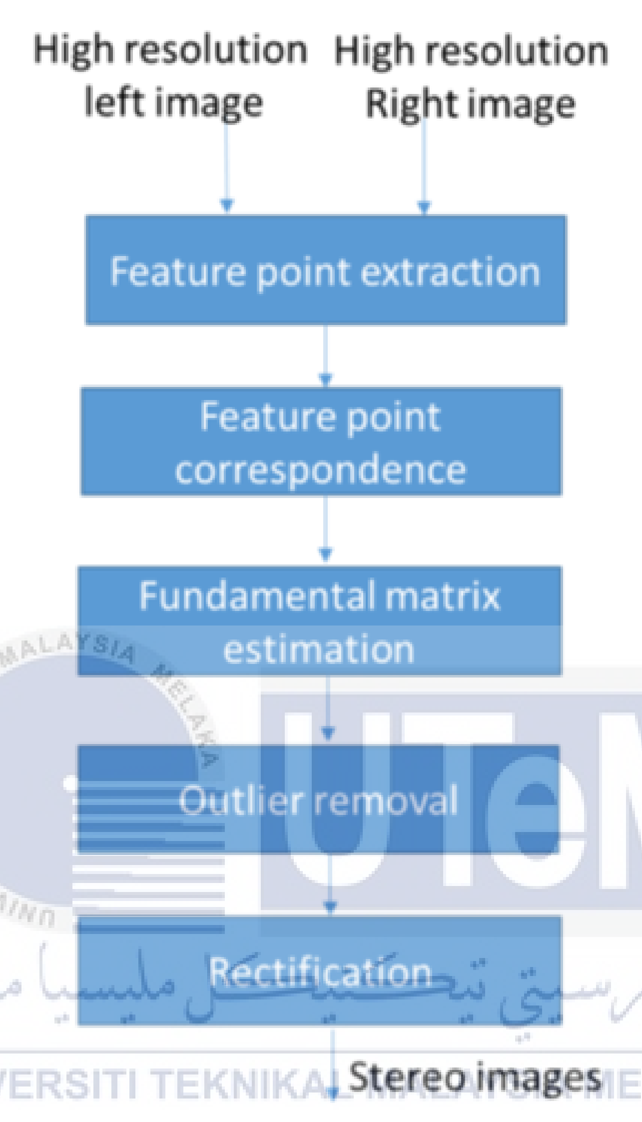


Figure 2.5: Procedures in stereo image formation.

2.2.2 Improve Accuracy of Disparity Map for Stereo Images using SIFT and Weighted Color Model

The author says (Piamsa-nga, 2015) utilize SIFT algorithms and weighted YCbCr to reduce computation time and increase accuracy to produce a map of disparity from stereo images. SIFT is used to bind the search area in the left and right images to match the objects. Weighted YCbCr is used to increase the accuracy of matching. Therefore, the speed and accuracy of the estimation of the depth depend on the analysis of disparity maps.

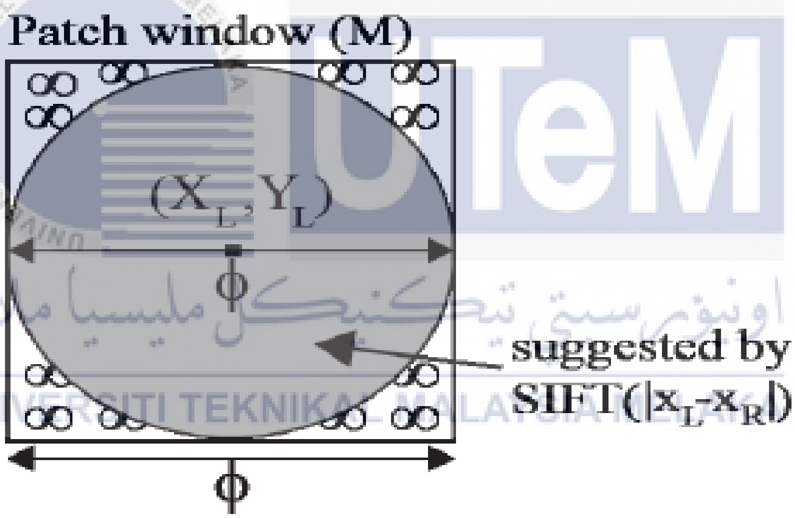


Figure 2.6: Circle indicates bounded area suggested by SIFT. Patch window (M). Square indicates "disparity max" area.

2.2.3 Creation of a Depth Map from Stereo Images of Faces for 3D Model Reconstruction

Stereo cameras are used in one of the available methods of reconstruction of 3D face models. The process of reconstruction usually involves several steps in (Olga Krutikovaa, 2017). A "Autodesk 3Ds Max" 3D editor was used to create a virtual scene that includes stereo cameras and a human head. The suggested approach has already been tested using two cameras named "VISAR," and a microcontroller called "Arduino Micro." "Arduino" program enables cameras to be combined by using the "Arduino Micro" microcontroller. The images are captured from the cameras using a program called "FlyCap." Since the initial images contain distortions, the algorithm's first step is camera calibrations. Similar dots are placed on all stereo images for calibration. Afterwards, Those points are being used to measure the degree of disruption and the images are corrected when necessary. The rectified images are being used for calculating the depth map. From the front half-tone images the depth map of the faces is built.

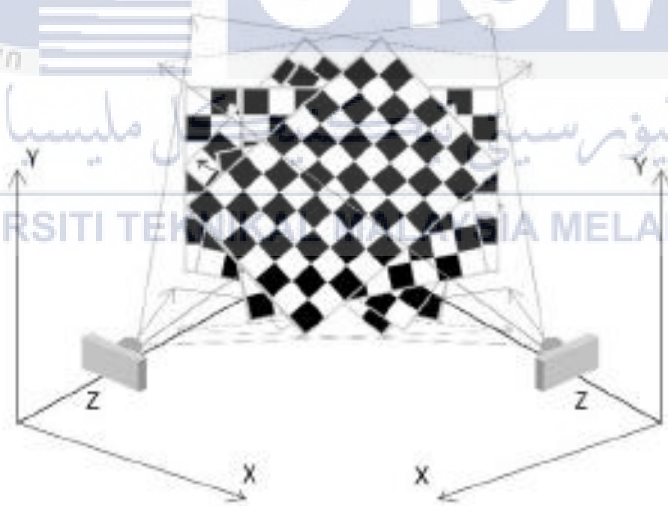


Figure 2.7: Camera calibration using a chessboard.

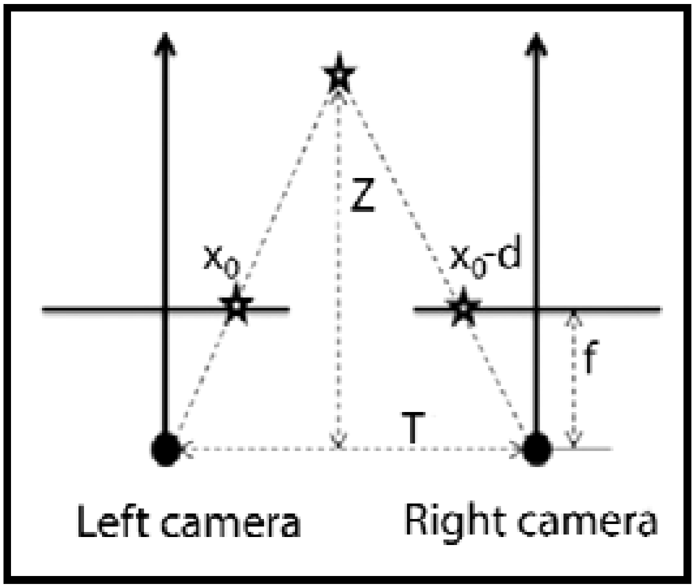


Figure 2.8: Calculating the depth map. (T – Stereo base, Z – distance)

2.2.4 Computer Vision Based Distance Measurement System using Stereo Camera View

Only Stereo cameras can create computer vision similar to human vision. Using the stereo camera system, a computer vision system is built to measure distances of objects. This can be supported by (ÇEVİK, 2019), Measure the distance from the stereo camera to the screen of the face images obtained. The maps of disparities are first derived in measuring the distances of the facial images to be evaluated, and the region of the face is observed. The range measurements are then conducted on the photographs collected in the stereo camera system, as the differences are measured between the frames. In the final stage the performance of the tool was checked by comparing a calculated Euclidean distance and the actual distance values for the computer vision system proposed.

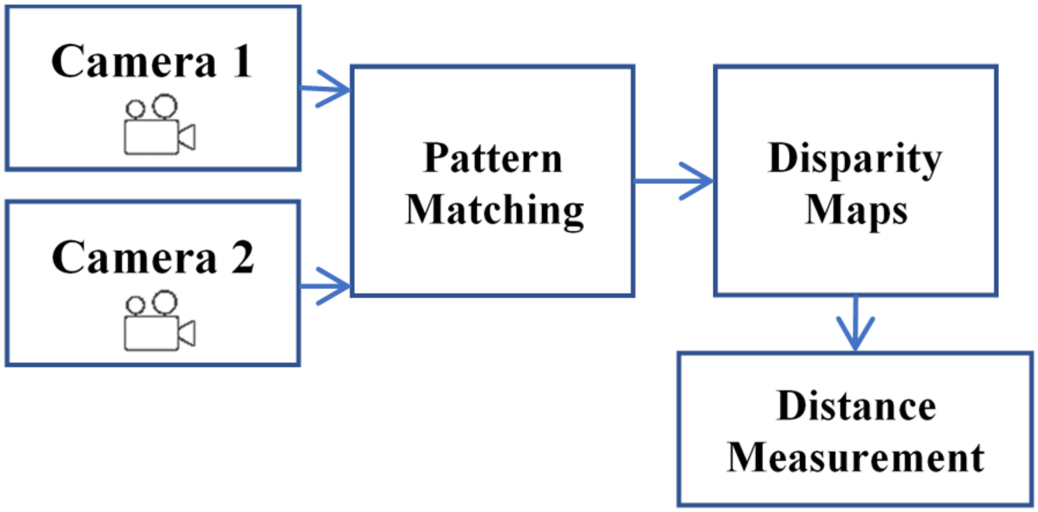


Figure 2.9: Block diagram of stereo vision system established

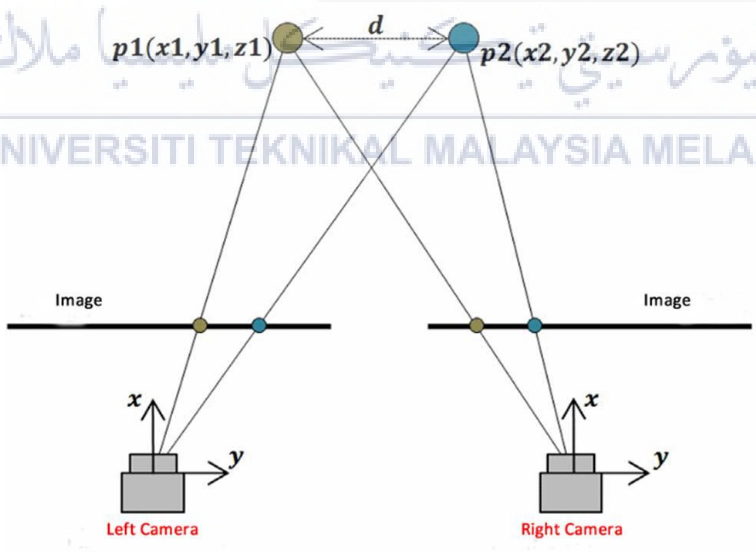


Figure 2.10: Computing inequality between two points on disparity maps

2.2.5 Disparity map estimation with deep learning in stereo vision

How a new fully network convolutional architecture can estimate the map of disparity between stereo images has been demonstrated. With the data obtained, it is possible to note how post-processing of the source images allows to discern the edges in the images, that seems to have been the key challenge to be tackled as a next step in the inquiry to yield more detailed findings in (Mejia, 2019) . Method for calculating the disparity map from the corrected stereo image pair. Their proposal, a new convolution layer-based neural network architecture to anticipate the stereo depth images. The Middlebury databases have been used to equate predicted map error with a known map of disparities to train the network.

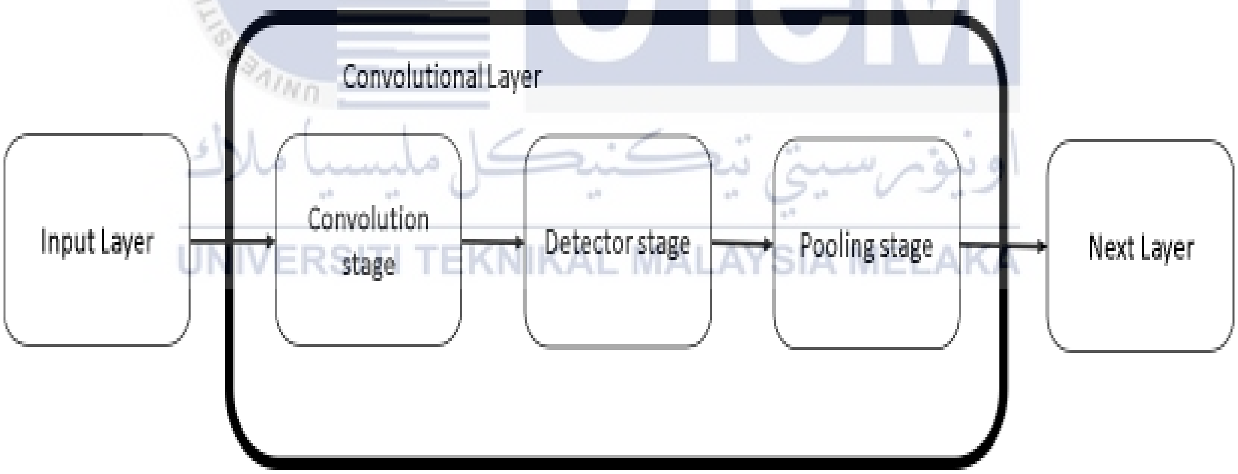


Figure 2.11: Components of a typical convolutional network

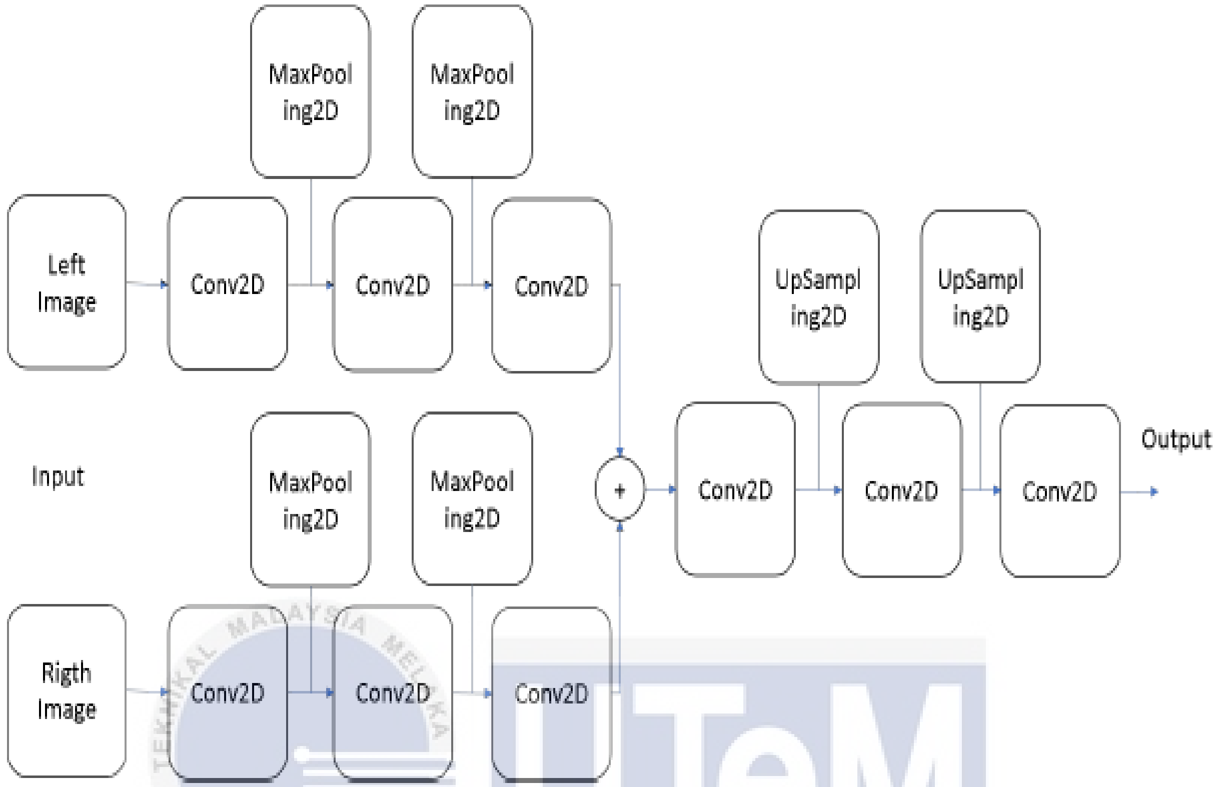


Figure 2.12: Architecture of the convolutional network.

2.2.6 Efficient Binocular Stereo Matching Based On Sad and Improved Census Transformation

In (YUN ZHANG, 2017) a new SAD-based, stereo matching algorithm and improved census transformation were proposed. In order to minimize noise and differences, they have also introduced better bilateral and selective filters to improve stereo matching performance. First, they conduct enhanced census transformation by integrating SAD and better census transformation, and then get equivalent costs. Finally, they combine the corresponding costs and measure the disparities. They as well propose enhanced selective filters and bilateral to increase the quality of disparities in order to create greater disparities.

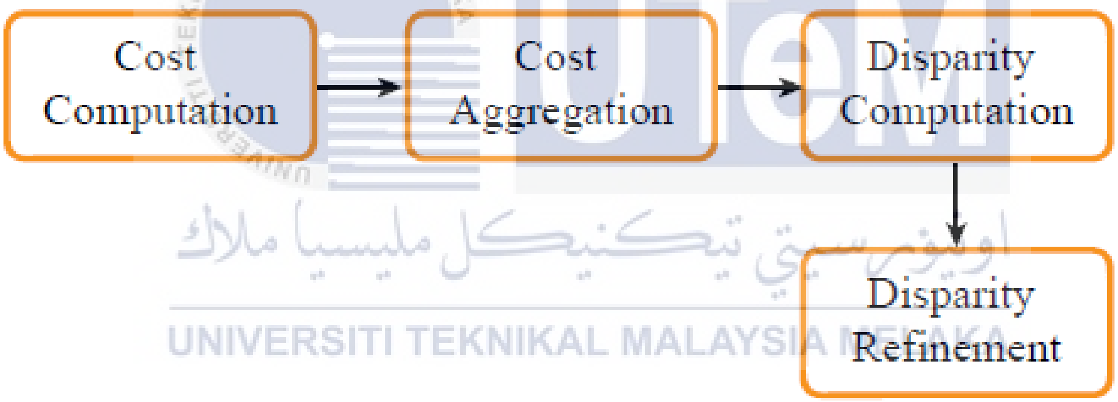
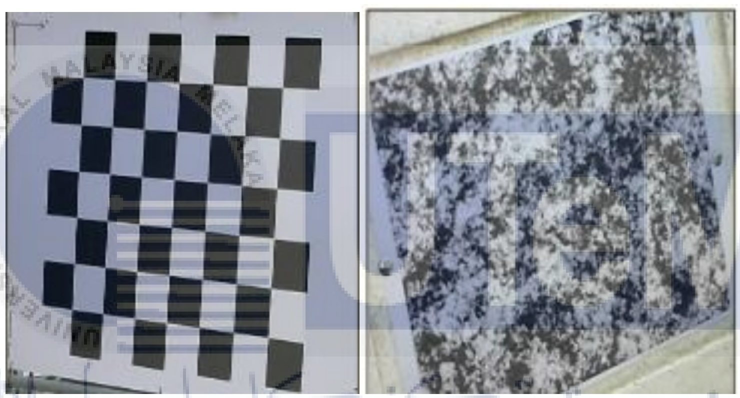


Figure 2.13: Flowchart of stereo matching.

2.2.7 3D DISTANCE MEASUREMENT ACCURACY ON LOW-COST STEREO CAMERA

In (M.F. Abu Hassan, 2017) the Logitech HD Webcam C270.0 offers a low-cost 3D sensor using a pair of mid-range webcams. Two calibration methods were tested, applied on the custom 3D sensor and assessed for their performance. The first method that was taken into account, which used the pattern of the checker board, outperformed the second method that used the special pattern. The former has the smallest MRE of 0.1272 pixels and generates a good recreation of 3D images reflecting the true physical 3D scene.



(a) (b)

Figure 2.14 Calibration pattern: (a) Chessboard (b) Feature Descriptor

CHAPTER 3

METHODOLOGY

3.1 Introduction

To achieve the objective of this project, this part will describe about the flow of this project and the path needed. It includes the concept design of the experiment will be discussed in this chapter too. It will cover the software development that will be used to accomplish the goal of this project. The aspect for each part in this steps in important because it will be determined whether the result is successful at the end of the project. This part also covered about the flowchart and block diagram to explain more detail about the process.

3.2 Project overview

This chapter will contribute to the methodology and process of this project. This methodology involves the process used for the research in attaining the objectives and achieve the best result. It consists of the introduction of the experiment, the procedure of experiment and the project flow chart. This chapter is an important part before proceed to the next chapter which in finding and analyzing will be implemented.

3.3 Flowchart represents process of project

Every project that developed, it must have a flowchart to understand the operation of the project using a visual diagram that aid to reader easy understanding. By using flowchart, it can make better understanding on how the process is done and also it used for improvement process. In this project, first, the stereo images are taken from left and right, then converted to grayscale images and then used to calculate the disparity map using SAD method.

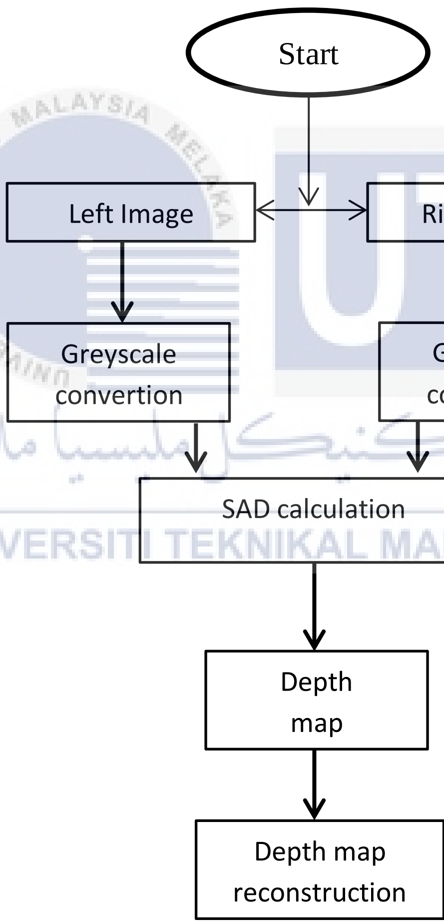


Figure 3.1: Project development flowchart

3.4 The Report Flowchart

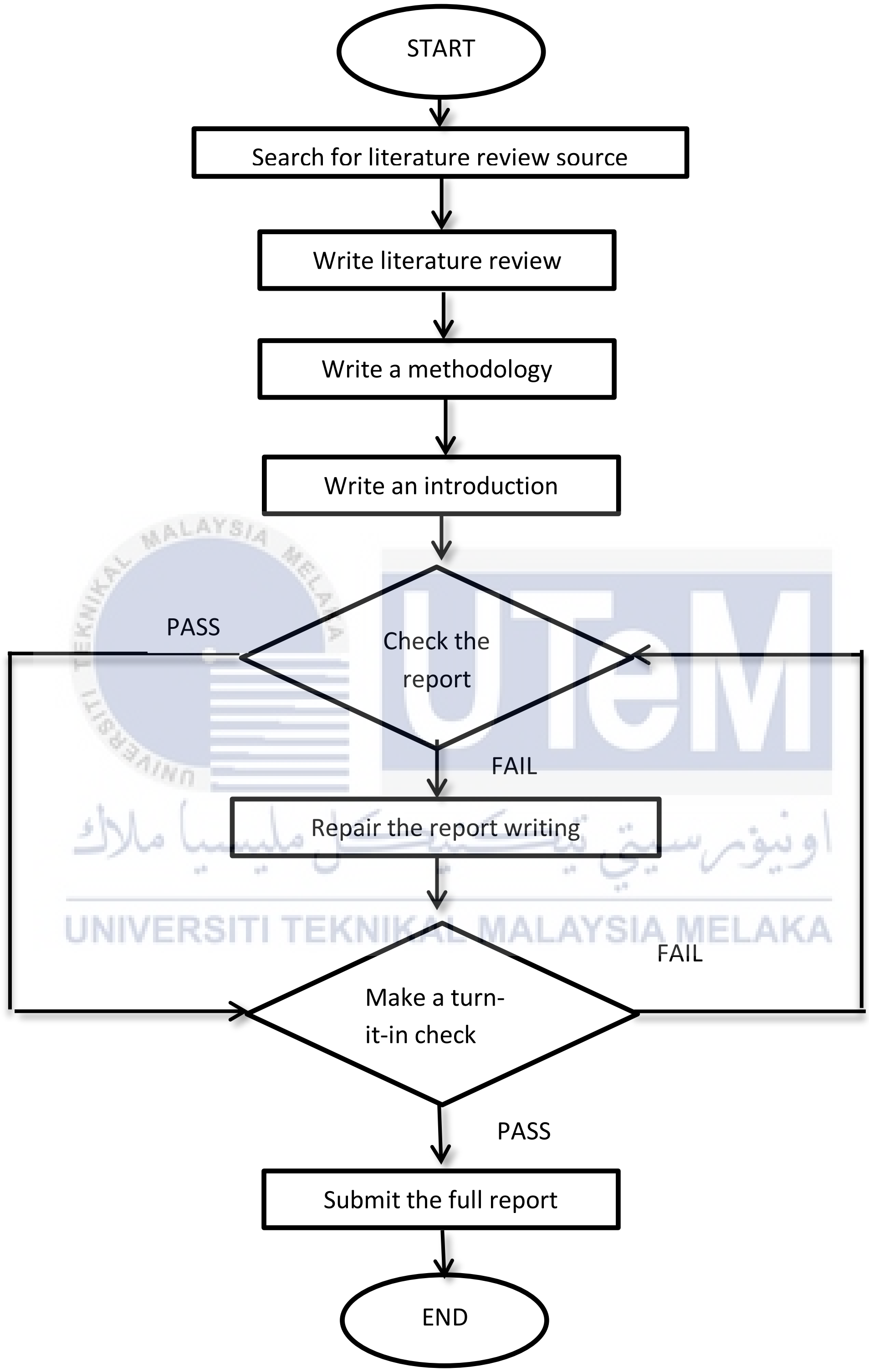


Figure 3.2: Shows the report flowchart

3.5 Block Diagram of Project

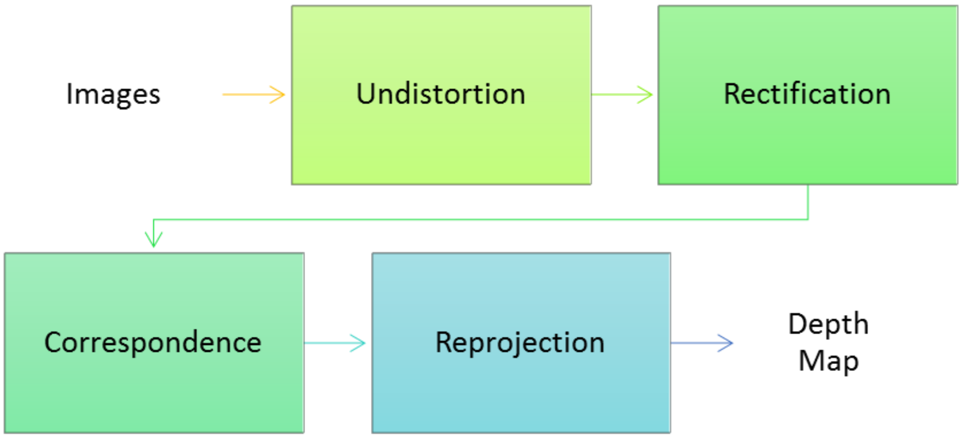


Figure 3.3: Block diagram of project.

As described in Figure 3.3 the reconstruction of a stereo vision system is usually mapped in four basic stages. There are usually four stages. The distortion of the captured images from the camera lens is first removed from the non-distortion stage. The elevation angle and the distance between the two camera images are corrected in the next step to establish the focal length and the epipolar axis, as assumed above.

The resemblance between the left and the right image would be calculated in the corresponding step, and used to calculate the map of disparity. This method is often known as being interoperable. Finally, the map of disparities can be reconstructed using the triangle property, and this procedure is termed reprojection.

3.6 Program Development

For program development, the coding will be based on the information of the standard measurement. References from the internet, article and journal will affect the development of the coding. Plus, the calculation model will be developed using MATLAB software. After that the automatizing measurement coding will be inserted in the program.

3.7 Software Implementation

MATLAB or known as Matrix Laboratory is a multi-numerical computing system and a patented programming language that are developed by Math works. This software can use to operates in many mathematical forms such as matrix operations, function plotting, data implementations of algorithms and combining the programs that are written in other languages.

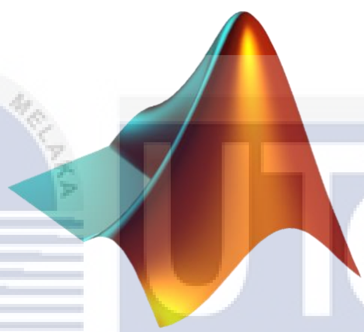


Figure 3.4: MATLAB

Image Processing toolbox will be used in this project. This toolbox provides many features such as visualization, analysis, images processing and algorithms development. This toolbox also allows segmentation method and enhancement, noise reduction, 3D processing image, image registration and sculptural transformation to be performed.

Furthermore, Image Acquisition Toolbox also is implied in this project. This toolbox will provide functions and let the users to connect and configure with external hardware properties. The toolbox also can generate the code to automatizing the user's acquisitions.

3.8 Stereo Vision Principle Depth Map Generation

In stereo vision, principal directions displaced cameras are used to capture the same view to get two different views from two different angles on a scene. The two images captured have many similarities, and fewer differences. In the case of human perception, the brain integrates the two images taken together to create a 3D model for the objects being observed. In computer vision, the 3D model for the observed objects is produced by finding correlations between the stereo images and by using projective geometry to analyze these images. Reconstruction problems using stereo are similar to the stereo pair.

The computer compares the images to find the corresponding bits, while moving the two images over each other. The magnitude of the difference between the two corresponding pixel resolution be called "disparity," which is related to the distance of the object. The displacement set between the matched pixels is generally indicated as a "disparity map". The greater disparity of image pixels means the object is nearer to the cameras and did appear sharper in the map of disparity, while the lesser differences means the thing appears darker away from cameras.

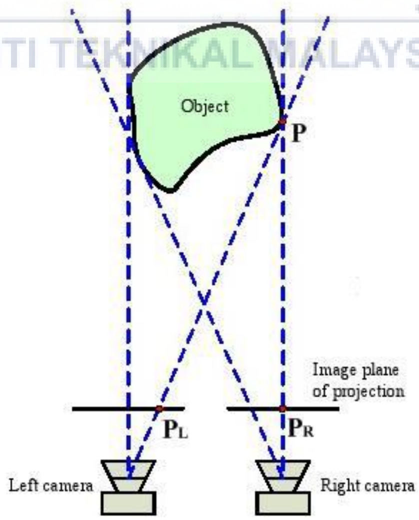


Figure 3.5: Two cameras' location and their projection planes image.

Figure 3.5 uses two similar cameras to represent the geometrical basis for stereoscopic images. Such cameras are mounted on the very same plane, and are switched in the same direction. Image planes are provided in front of cameras for ease of projection modeling.

Disparity is the difference in stereo image pair between two pixels which corresponds to the same physical point. After making the stereo images, a process of correspondence of features is required to specify the pixels that relate to the same physical point. The map of disparities is determined for every pixel in the image, based on the results of the correspondence feature. Figure 3.6 shows the depth-disparity relation.

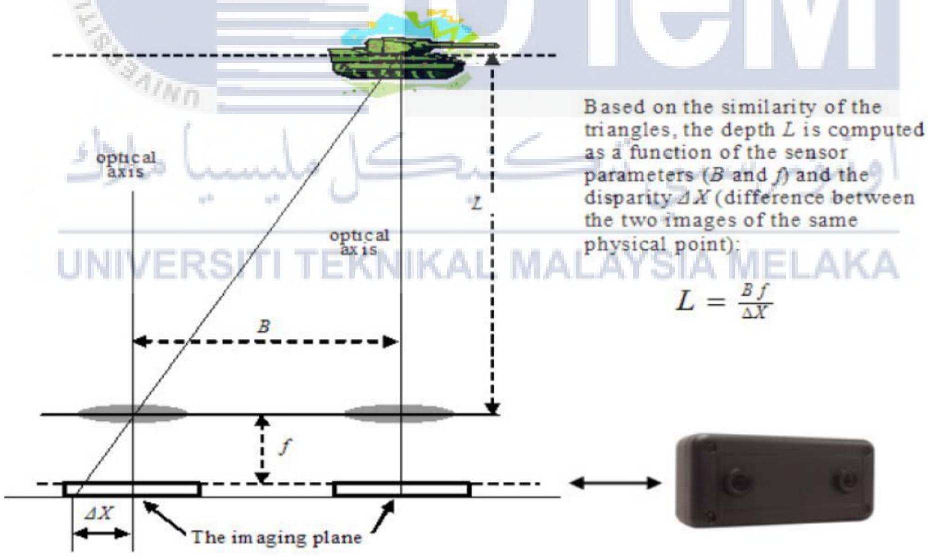


Figure 3.6: Relationship between disparity and depth.

3.9 Sum of Absolute Differences (SAD)

The map of disparities is obtained by moving the source window in the left pixel-by - pixel picture-pixel over the aim windows in the right image within a different search scope called disparity range and calculate (SAD) for each moving phase, with the overall SAD computed among two windows representing the similarity of both windows. The amount of difference between matched windows reflects the value of disparity (d) of the reference image's central pixel. To calculate their differential values, this process is repeated for whole pixels in the reference image. Sort those values to create a map of disparities. The SAD is calculated on the basis of the equation:

$$\mathrm{SAD} = \Sigma(\mathrm{i},\mathrm{j}) \in \mathrm{W}(\mathrm{x},\mathrm{y})\,|Il\,(\mathrm{i},\mathrm{j}) - \mathrm{I}r(\mathrm{i},\mathrm{j}+\mathrm{d})| \quad (1)$$

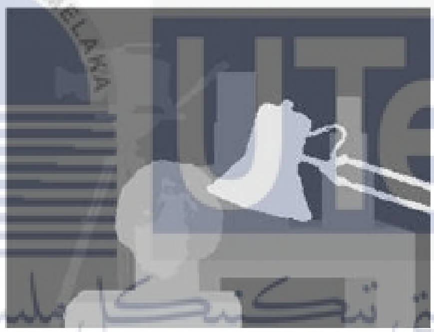
$$SAD : \Sigma(Il\,(x,y), Ir\,(x+d,y)) = \Sigma|Il\,(x,y) - Ir\,(x+d,y)| \quad (2)$$

Where: W (x, y) Is a window surrounding the location (x , y), d is the value of the disparities, Il and Ir are the intensity values in the left and right images respectively. This method has been developed with MATLAB program and executed at different window sizes with various stereo image pairs to depict its results. One of the Stereo Pairs (Tsukaba) with a scale of 384×288.

3.9.1 The examples of SAD in MATLAB

(a) Left image.

(b) Right image.



(c) true disparity map.

Figure 3.7: Tsukaba stereo image pair with true disparity map (a), (b), and (c).

The outcome of the matching algorithm is a differential diagram that can be displayed as an image, representing the 2D picture 3D (left and right) as seen in Figure 3.7. The disparity is directly proportionate to the concentration, the target can be seen closing to the stereo cameras and emerging in a light gray color. The farthest point, on the other hand, has a lower variance value and appears in dark gray color.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Introduction

In this chapter, it will cover about the finding result and the analysis from the collected data. The summary and the analysis for the project will be discussed at the last part of this chapter.

4.2 Test platform

The experiment are run on a AMD E1-1200 APU with system on a Radeon™ HD Graphic 1.4GHz with 4.00 GB memory(RAM). Experiment is implemented on MATLAB R2020b.

4.3 Test Data

Stereo images carefully selected from the Middleburry dataset are test data. Such pictures, which are shown in Figure 4.1, are called Tsukuba, Venus, Cone and Teddy. In its metadata, ground truths of the depth of such images are available and each image defines the maximum disparity to search for correctness. Their image characteristics are defined as follows:

- **Tsukuba** – disparity max is 15, repeat texture, texture less, and thin structure.
- **Venus** – disparity max is 19, detail in an image that the shape of a plane stacked.
- **Cone** – disparity max is 59, high-resolution with complex geometry.
- **Teddy** – disparity max is 59, complex geometry and resembles the texture is similar surface without obvious characteristic.

	Left	Right
a)		

b)		
c)		
d)		

Figure 4.1 a) Tsukuba b) Venus c) Cone d) Teddy

4.4 A GUI demonstrates an efficient stereo matching algorithm.

The stereo vision system was test by Middleburry dataset. The kinematics algorithms were implement in MATLAB® R2020b. A graphical user interface is developed in the MATLAB environment as show in Figure 4.2. The GUI reads the images obtained from the data base. The depth map of the scene is generated by the SAD algorithm.

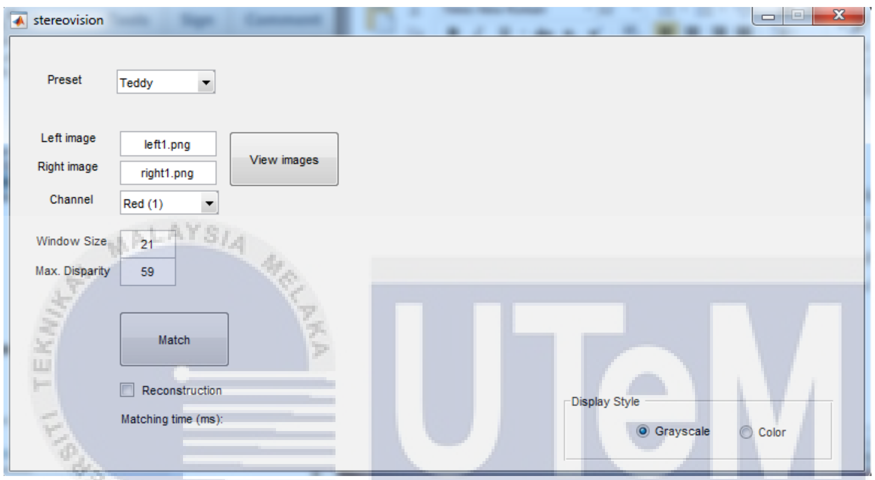


Figure 4.2 Graphical user interface for background stereo vision.

Background Figure 4.2 display the outline and the GUI of the depth mapping reconstruction process. The method of the disparity via the stereo imaging is based on the detection of analogous images information around corresponding point in left (reference) and right (target) stereo images. Provided that enough and fairly distributed points are detected in both images this method obtains similar results as the target calibration.

In the resulting rectified images, the correspondence problem is reduced to a 1-dimensional search for correspondences. The relative positions of corresponding points, where point coordinates are given with respect to the principal points of left and right images, denote the depth information (disparity D) of the respective pixels. I implemented a block matching method, which makes use of the properties of the surrounding pixels and which uses a couple of statistical measures, e.g. the correlation C of grey values, where corresponding pixels are given by the maximum correlation of blocks.

At the top left of the GUI, preset function is to select the images that I already set in database at the coding of MATLAB. The view images block to view the images we choose for example, I choose the Teddy pictures then it will show the left and right images that I choose as show in Figure 4.3. For window size and Max disparity, the data already set at the coding in MATLAB. I already set up the data of the value of Max disparity and the window size. When we select the preset (picture), the value of the max disparity and window size will produce automatically at the GUI.

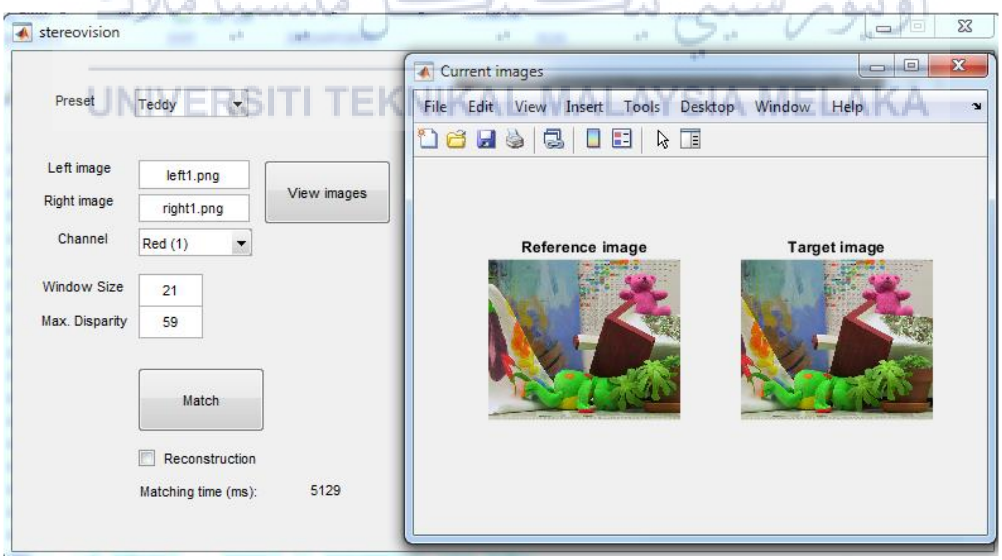
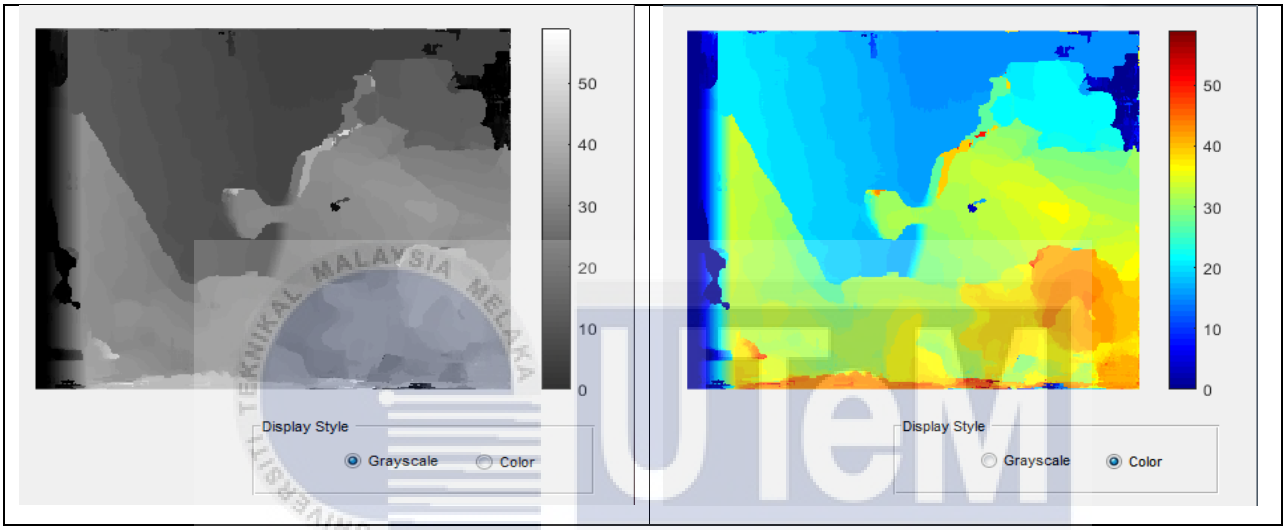


Figure 4.3 View images from the GUI

The button *Match* at the GUI background for do the stereo matching and tick the *Reconstruction* to see the depth mapping reconstruction for the pictures was select at the stereovision GUI. For the display result for stereo matching can be choose either Greyscale or Color disparity display. Figure 4.4 show the result of the stereo matching and the disparity map reconstruction.



a)

b)

Figure 4.4 a) Stereo matching b) Depth mapping reconstruction

4.5 Depth mapping propagation

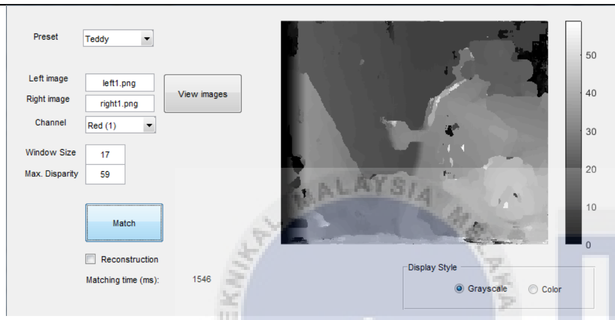
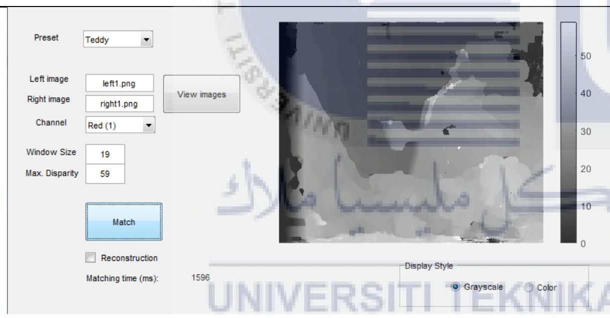
As described in the preceding subsection, the correlation routine can be performed by calculating disparity values for each pixel on the range of rectified stereo images. As shown in Figure 4.5, the disparity map is composed of the computed depth value for each pixel in the 2D image. The darker intensities reflect objects far from the camera or sensor, while the lighter intensities reflect objects closest to the camera.

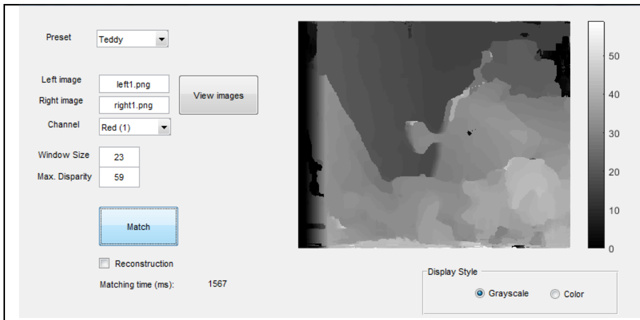
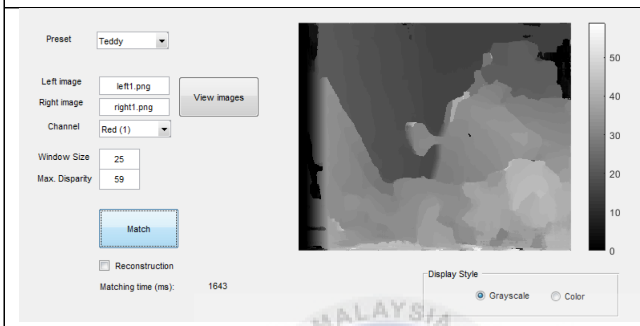
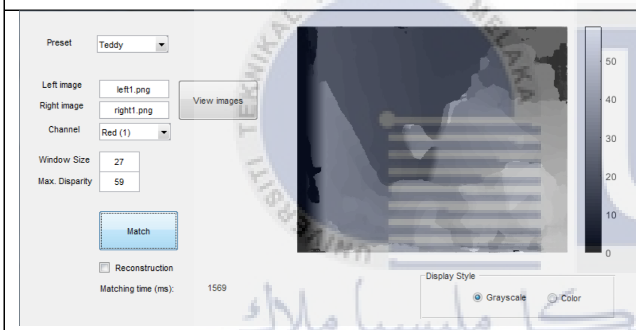
For example, Tsukuba picture, the table lamp object is whiter or brighter than the object that is far from the lamp. Next, the Venus picture disparity map show that the paper brighter than the order object and have darker area at the left of picture where disparity was unable to be calculated. The Cones picture also same like Venus picture, there are some areas that have black or darker area that cannot calculate by the disparity. Last but not lease, the teddy picture show that frog doll is lighter than the teddy bear. That means the teddy bear far from the camera or sensor than the frog doll.

I change the size of the support window for the richly textured regions in the stage of cross-based support window construction. I increased the value of the window size from 15, 17, 19, 21, 23, 25 and 27. As seen in Figure 4.6, these parameters influence the small shift in the accuracy disparity of teddy images.

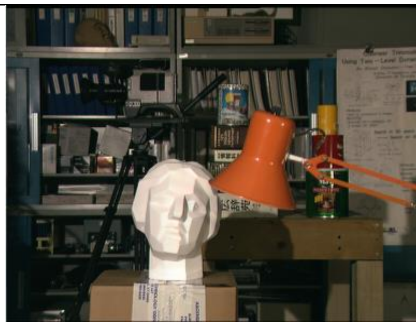
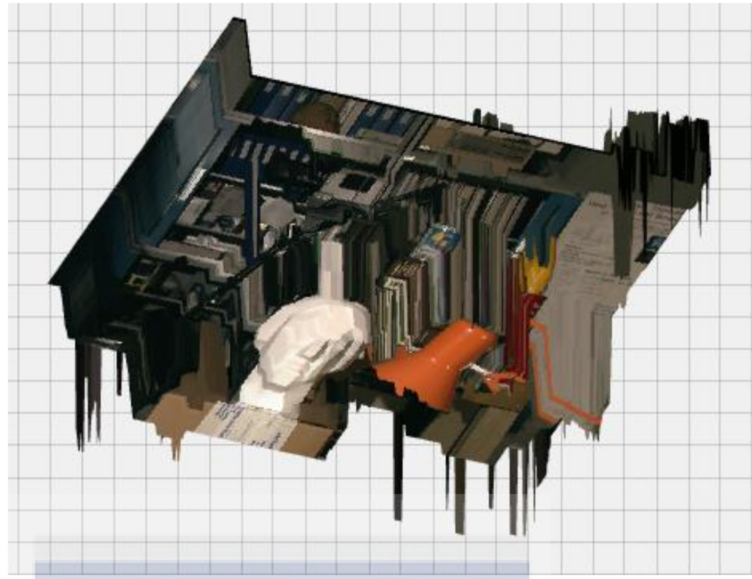
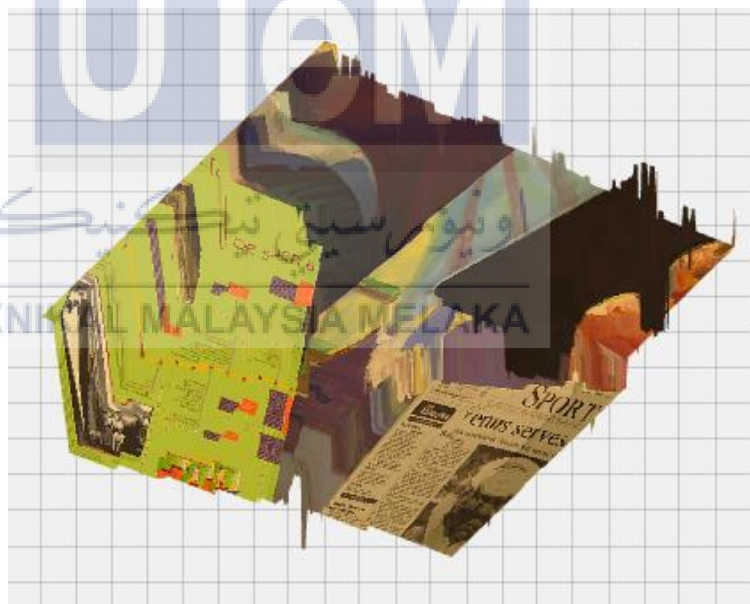
Each pixel corresponding to the (x, y, z) co-ordinates representing its location in the 3D space can be re-projected from the disparity map onto a set of data points in the 3D point cloud coordinate system. The 3D scenes captured from the stereo images are represented in Figure 4.7. The results obtained are realistically precise, describing the scenario depicted very well and interpreting even intense viewpoints correctly. In Figure 4.7 represent a couple of high object positioned against the wall. For example, the cones image we can clearly see the distance and position of each cone, so we can know which object is closer to the camera or sensor. This same level of precision is achieved in all the image from Middlebury dataset and the results are quite satisfying.

Figure 4.5 Result of the algorithm

Teddy	Max. Disparity	Window size	Time(ms)
	59	15	941
	59	17	1546
	59	19	1596
	59	21	1571

	59	23	1567
	59	25	1643
	59	27	1569

4.6 Comparison Disparity with difference window size

<table>
<tr><td></td><td></td><td>3D Reconstruction</td></tr>
<tr><td>Left</td><td></td><td rowspan="2"></td></tr>
<tr><td>Right</td><td></td></tr>
<tr><td>Left</td><td></td><td rowspan="2"></td></tr>
<tr><td>Right</td><td></td></tr>
</table>

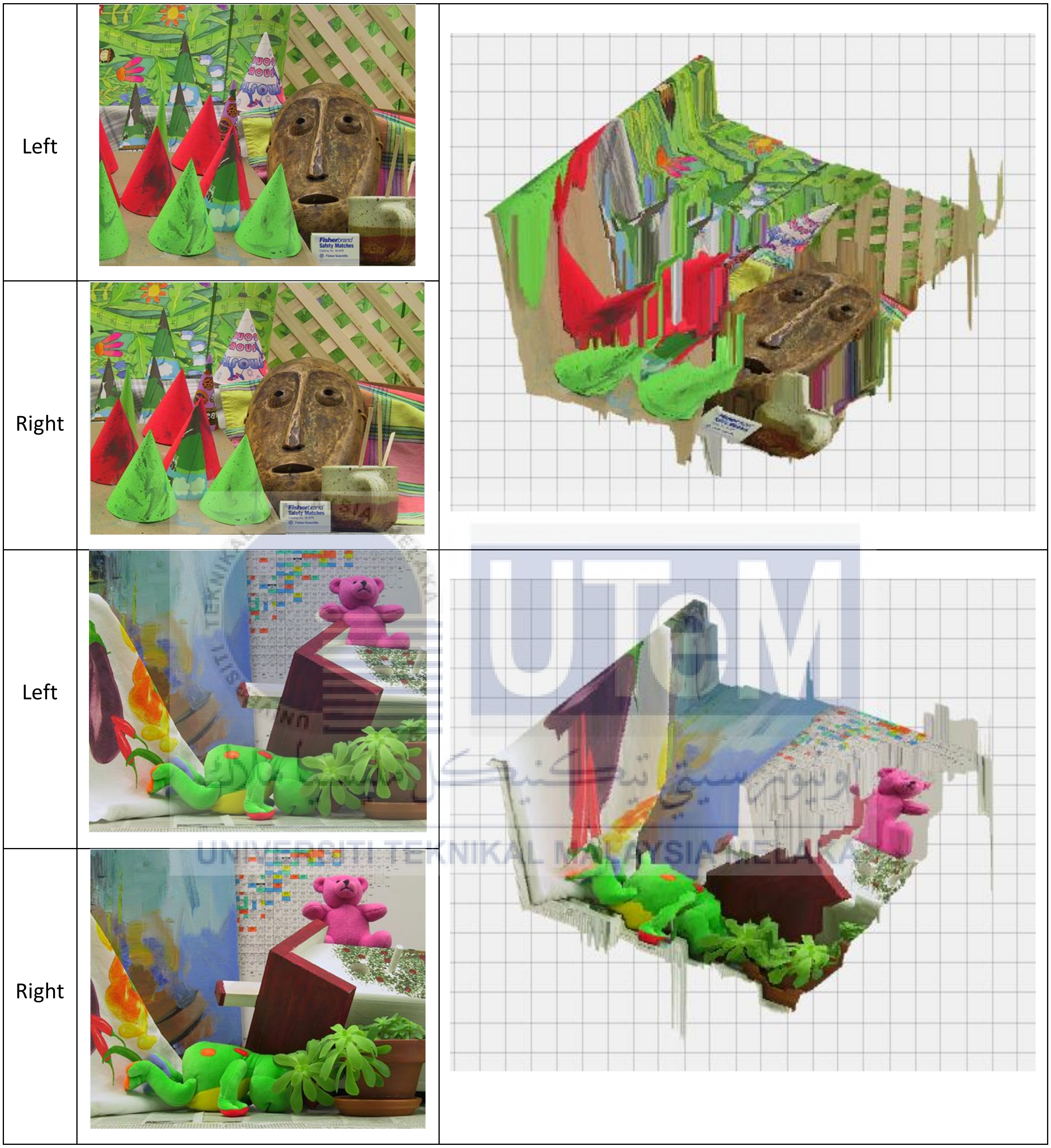


Figure 4.7 3D reconstruction

4.6 Summary

This chapter presents the result of how the project is done. The experiment was implemented on MATLAB R2020b and this project using graphical user interface that developed in MATLAB environment. The test data was selected carefully from Middleburry dataset which is Tsukuba, Cones, Venus and teddy. A graphical user interface (GUI) that develop in MATLAB will show the depth map reconstruction after select the picture that we want to simulate. After select the picture from the GUI background, the GUI will show the result of disparity mapping algorithm. A grayscales disparity map reveals the disparities for each pixel of the corresponding intensity is the result of the stereo matching process. The lighter areas are closer to the camera, the darker ones are further away. Black areas are points that could not be calculated by disparity. In summary, MATLAB is used in each of the images to locate matched points, significantly assisted and accelerated by the epipolar constraint concepts. It continues to calculate the disparity between the respective lines, thus giving me the depth of the object in question. In the stereo images that give my final solution to 3D reconstruction, this method can be repeated for every pixel.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

This paper proposes a stereo matching algorithm based on disparity propagation using sum of absolute differences. By extracting disparity subsets for reliable pixels and define a new cost volume accordingly, mismatches in the initial disparity map are corrected through disparity propagation from nearby reliable points. The objectives of this project "Depth Mapping Reconstruction from Stereo Images Using Sum of Absolute Differences" has been achieved successfully. Developed by making the stereo matching algorithm using Sum of Absolute Differences (SAD) by using MATLAB. Next, a simple stereo algorithm compute results comparable to current state-of-the-art on Middlebury benchmark. The accuracy of the propagation can be guaranteed with the pixelwise line segments, which shows good performance in both speed and accuracy. This project has answered all my objectives and the results I got to achieve the set benchmark. The result that has been displayed in Figure 4.5 can be supported by (D. Scharstein and R. Szeliski, 2015) (XingzhengWang, 2016). Each one of these algorithms has its own specifications. In various domains, such as tracking, robot navigation, etc., these algorithms can be better used. There is no better or indispensable depth cue than another depth cue. A cue has its own perks and disadvantages.

5.2 Recommendation for Future Work

Some suggestions are presented as a reference for following researcher proceeding with this project. As a future work, I would like to apply my algorithm to real world problems and I plan to port the algorithm to graphics hardware. Next, the proposed SAD algorithm will be optimized and implemented in a hardware framework for an indoor robot navigation system.

REFERENCES

A. K. Wassim, R. P. (2016). 3 D Object Tracking Using Disparity Map., (pp. 108-111).

A.Qayyum, A. M. (2015). Disparity Map Estimation Based on Optimization Algorithms using Satellite Stereo. (pp. 127-132). IEEE International Conference on Signal and Image Processing Applications (ICSIPA).

Ayhan, C. K. (2018). Stereo Image and Depth Map Generation for Images with Different Views and Resolutions.

B. Yang, S. R. (2018). Dense 3D object reconstruction from a single depth view. *TPAMI*. DOI: 10.1109/TPAMI.2018.2868195.

C. Kwan, B. C. (2018). Enhancing Stereo Image Formation and Disparity Map Estimation for Mastcam Images. IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference.

ÇEVİK, E. D. (2019). Computer Vision Based Distance Measurement System using Stereo Camera View.

Cigla, C. (2015). Recursive edge-aware filters for stereo matching. (pp. 27–34). Boston, MA, USA: IEEE Conference on Computer Vision and Pattern Recognition Workshops.

D. Scharstein and R. Szeliski. (2015). *Middlebury Stereo Evaluation.*

Dongyoung, L. C.-H. (2017). Dense Disparity Map-based Pedestrian Detection. (pp. 1015-1018). IEEE International Conference on Intelligent Trans-.

E. Dibra, H. J. (2017). Human shape from silhouettes using generative HKS descriptors and cross-modal neural networks.

G. Song, H. Z. (2017). Training a Convolutional Neural Network for Disparity Optimization in Strereo Matching . (pp. 48-52). CM, New York, NY, USA: International Conference on Computational Biology and Bioinformatics.

I. J. Goodfellow, J. P.-A.-F. (2014). Generative adversarial nets.

J. Wu, C. Z. (2016). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling.

J. Wu, Y. W. (2017). MarrNet: 3D shape reconstruction via 2.5D sketches.

K. Zhang, Y. F. (2017). Cross- scale cost aggregaration for stereo matching . (pp. 965-975). IEEE Trans, Circuit Syst. Video Techn.

LeCun, J. Z. (2015). Computing the stereo matching cost with a convolutional neural network. (pp. 1592–1599). IEEE Conference on Computer Vision and Pattern Recognition.

Lobaton, Q. G. (2017). Obstacle Detection in Outdoor Scenes Based on Multi Valued Stereo Disparity Maps. (pp. 1-8). IEEE Symposium Series On Computation Intelligence (SSCI).

M. Tatarchenko, A. D. (2017). . Octree generating networks: Efficient convolutional architectures for high-resolution 3D outputs. *ICCV*.

M. Wang, L. W. (2017). 3DensiNet: A robust neural network architecture towards 3D volumetric object prediction from 2D image. ACM MM 2017.

M.F. Abu Hassan, A. H. (2017). D DISTANCE MEASUREMENT ACCURACY ON LOW-COST STEREO CAMERA. *Computer Assisted Intelligent Event Processor.*

Malik, J. T. (2015). Shape, illumination, and reflectance from shading. pp. , 37(8):1670–1687.

Mejia, J. (2019). Disparity map estimation with deep learning in stereo vision., (pp. 27 - 40).

N. Mayer, E. l. (2016). A large dataset to train convolutional networks for disparity, optical. (pp. 4040-4048). IEEE Conference on Compputer Vision and Pattern Recognition.

Okae, J. D. (2017). Optimization of Stereo Vision Depth Estimation using Edge - Based Disparity Map. (pp. 1171 - 1175). 10th International Conference on Electrical Engineering .

Olga Krutikovaa, A. S. (2017). Creation of a Depth Map from Stereo Images of Faces for 3D Model Reconstruction. (pp. 452 – 459). Riga, Latvia: Procedia Computer Science 104.

P, Y. H. (2017). Segment - tree based cost aggregation for Stereo matching with enhanced segmentation advantages. (pp. 2027-2011). LA, USA 2027- 2031: IEEE International Conference on Acoustics, Speech and Signal Proceesing.

Piamsa-nga, O. P. (2015, february). Improve Accuracy of Disparity Map for Stereo Images using SIFT and Weighted Color Model. pp. 109 - 114.

Pollefeys, A. S. (2017). Sgm-nets: Semi-global matching with neural networks. (pp. 6640–6649). IEEE Conference on Computer Vision and Pattern Recognition.

Roth., S. R. (2015). Discriminative shape from shading in uncalibrated illumination.

S. Savarese, M. A. (2007). 3D reconstruction by shadow carving:. pp. 71(3):305–336.

Tulsiani., S. (2018). Learning Single-view 3D Reconstruction of Objects and Scenes. *PhD thesis*. UC Berkeley.

W. Luo, A. G. (2016). Efficient deep learning for stereo matching. (pp. 5695–5703). Las Vegas, NV, USA: IEEE Conference on Computer Vision and Pattern Recognition.

W. Luo, A. G. Schwing and R. Urtasun. (2016). (pp. 5695-5703). Las Vegas: IEEE Conference on Computer Vision and Pattern Recognition.

Weijer, M. G. (2015). Accurate stereo matching by two-step energy minimization. (pp. 1153–1163). IEEE Trans.Image Processing.

XingzhengWang, Y. T. (2016, september 14). Stereo Matching by Filtering-Based Disparity Propagation. p. PLoS ONE 11(9): e0162939.

Y. Zhang, Z. L. (2019). An efficient generation network for 3d object reconstruction from a single image. *IEEE Access*, pp. 7:57539–57549.

YUN ZHANG, W. C. (2017). EFFICIENT BINOCULAR STEREO MATCHING BASED ON SAD AND IMPROVED CENSU TRANSFORMATION.

Z. Liang, Y. F. (2018). Learning For Disparity Estimation through Feature Constancy. CVPR.

Zbontar, J. a. (2015). Stereo matching by training a convolutional neural network to compare image patchers.

APPENDIX

Activity	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Confirm the title														
Study journal														
Drafting literature review														
Update literature review														
Methodology														
Introduction														
Full report and turnitin														
Preparation slide														
Presentation PSM 1														

APPENDIX 1: The Gantt Chart of Progression project 1