Faculty of Electrical Engineering

DISPARITY MAP ALGORITHM FOR STEREO MATCHING PROCESS USING LOCAL BASED METHOD

Melvin Gan Yeou Wei

Master of Science in Mechatronic Engineering

2021

DISPARITY MAP ALGORITHM FOR STEREO MATCHING PROCESS USING LOCAL BASED METHOD

MELVIN GAN YEOU WEI

**A thesis submitted
in fulfilment of the requirements for the degree of Master of Science
in Mechatronic Engineering**

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

DECLARATION

I declare that this thesis entitles "Disparity Map Algorithm for Stereo Matching Process Using Local Based Method" 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 : MELVIN GAN YEOU WEI

Date : 12/12/2021

APPROVAL

I hereby declare that I have read through this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Science in Mechatronic Engineering.

Signature : ....

Supervisor's Name : TS. DR. ROSTAM AFFENDI BIN HAMZAH

Date : 12/12/2021

DEDICATION

To my late father and beloved mother

ABSTRACT

The aim of Stereo Vision Disparity Map (SVDM) algorithm is to obtain the disparity map from two images. These input images have different viewpoints that corresponds with each other, forming a two-dimensional mapping of matching pixels and is known as disparity map. The SVDM algorithm can be categorized into local, semi global and global methods. Global method performs a matching process using global energy or a probability function over the whole image. This method involves high computational complexity and slow implementation. Therefore, it is not suitable for real-time applications. However, local method solves the matching problem via a local analysis and aggregating matching costs over a support region at each pixel in the images. The local method delivers fast execution and low computational requirement. Semi global method is the combination of both methods which make this method is more complex. However, the computation for the development of SVDM algorithm is more challenging especially for the images with complex scenes. There are several factors such as low texture region, repetitive patterns, illumination different, discontinuity, and occlusion. Hence, this thesis proposes a local-based SVDM algorithm that increases the accuracy on the complex scenes. The proposed SVDM algorithm involves four stages which starts from matching cost computation. At this stage, the proposed work uses the combination of Absolute Difference (AD) and Gradient Matching (GM) to produce the initial disparity map. Second stage, the Minimum Spanning Tree (MST) is utilised to remove noise from the initial disparity map. Then, the optimization stage uses a Winner-Take-All (WTA) strategy. The WTA strategy absorbs the minimal aggregated corresponding value for each valid pixel in disparity map. At the final stage, Bilateral Filter (BF) with Histogram Equalization and Weighted Median (WM) filter are proposed. These filters are capable to increase the accuracy and preserve the object edges. In this research, two standard online benchmarking database systems are used to measure the accuracy of the proposed algorithm. These systems are from the Middlebury Stereo for quantitative measurement and Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) for qualitative measurement. Once satisfactory results are obtained, there is demonstration on 3D surface reconstruction using disparity maps produced by the proposed SVDM algorithm. In conclusion, the proposed SVDM algorithm produces accurate results from the validation process. The final results are 7.55% for avg nonocc error and 10.6% for avg all error which are competitive with other established methods when compared in the standard benchmarking evaluation system from the Middlebury Stereo. The disparity map results on the complex scenes are also improved and fine quality of 3D surface reconstruction have been produced for both of the Middlebury and KITTI images.

ALGORITMA PETA PERBEZAAN UNTUK PROSES PADANAN STEREO MENGGUNAKAN KAEDAH SETEMPAT

ABSTRAK

Tujuan algoritma Peta Perbezaan Penglihatan Stereo (PPPS) adalah untuk mendapatkan peta perbezaan dari dua imej. Imej-imej ini mempunyai sudut pandang berbeza yang saling berkaitan, membentuk pemetaan piksel dua dimensi yang sepadan dikenali sebagai peta perbezaan. Algoritma PPPS boleh dikategorikan kepada kaedah tempatan, global dan separa global. Kaedah global melakukan proses padanan menggunakan tenaga global atau fungsi kebarangkalian pada keseluruhan imej. Kaedah ini melibatkan kekompleksan pengkomputeran yang tinggi dan pelaksanaan yang perlahan. Oleh itu, ia tidak sesuai untuk aplikasi-aplikasi masa nyata. Walau bagaimanapun, kaedah tempatan menyelesaikan masalah pemadanan melalui analisis tempatan dan mengagregatkan kos pemadanan di kawasan sokongan pada setiap piksel dalam imej. Kaedah tempatan memberikan pelaksanaan cepat dan keperluan pengkomputeran yang rendah. Kaedah separa global adalah gabungan kedua-dua kaedah yang menjadikan kaedah ini lebih kompleks. Walau bagaimanapun, pengiraan untuk pembangunan algoritma PPPS lebih mencabar terutamanya untuk imej dengan pemandangan yang kompleks. Terdapat beberapa faktor seperti kawasan bertekstur rendah, corak berulang, pencahayaan berbeza, ketakselanjaran, dan oklusi. Oleh itu, tesis ini mencadangkan algoritma PPPS berasaskan tempatan yang meningkatkan ketepatan pada pemandangan kompleks. Algoritma PPPS yang dicadangkan melibatkan empat peringkat yang bermula dari pengiraan kos yang sepadan. Pada tahap ini, algoritma yang dicadangkan menggunakan kombinasi Perbezaan Mutlak (PM) dan Padanan Gradient (PG) untuk menghasilkan peta perbezaan awal. Tahap kedua, Minimum Spanning Tree (MST) digunakan untuk menghilangkan hingar dari peta perbezaan awal. Kemudian, tahap pengoptimuman menggunakan strategi Winner-Take-All (WTA). Strategi WTA menyerap nilai sepadan minimum terkumpul untuk setiap piksel yang sah dalam peta perbezaan. Pada peringkat akhir, penapis Bilateral, pemerataan histogram dan penapis Pemberat Pertengahan (PP) dicadangkan. Penapis ini mampu meningkatkan ketepatan dan memelihara tepi objek. Dalam penyelidikan ini, dua sistem pangkalan data penandaarasan standard digunakan untuk mengukur ketepatan algoritma yang dicadangkan. Sistem ini adalah daripada Stereo Middlebury untuk pengukuran kuantitatif dan Institut Teknologi Karlsruhe dan Institut Teknologi Toyota (KITTI) untuk pengukuran kualitatif. Setelah keputusan yang memuaskan diperolehi, terdapat demonstrasi pada pembinaan semula permukaan 3D menggunakan peta perbezaan yang dihasilkan oleh algoritma PPPS yang dicadangkan. Kesimpulannya, algoritma PPPS yang dicadangkan menghasilkan keputusan yang jitu daripada proses validasi. Keputusan akhir ialah 7.55% untuk purata ralat bukan "nonocc" dan 10.6% untuk purata ralat "all" di mana ianya berdaya saing dengan kaedah-kaedah yang sudah teruji jika dibandingkan dalam sistem standard penilaian penanda aras daripada Middlebury Stereo. Keputusan peta perbezaan pada pemandangan yang kompleks juga bertambah baik dan pembinaan semula permukaan 3D dengan kualiti yang jelas dihasilkan untuk kedua-dua imej Middlebury dan KITTI.

ACKNOWLEDGEMENTS

First, I would like to express my sincere acknowledgment and gratitude toward my main supervisor, Ts. Dr. Rostam Affendi bin Hamzah, senior lecturer in Fakulti Teknologi Kejuruteraan Elektrik & Elektronik (FTKEE) of Universiti Teknikal Malaysia Melaka (UTeM), for his guidance, patience, encouragement, time, knowledge and wisdom. With all these, I was able to enrich my knowledge, especially on Stereo Vision Disparity Map (SVDM) technology, which also help me in completing this thesis. Next, I would also like to express my gratitude towards my co-supervisor, Dr. Nik Syahrim bin Nik Anwar, senior lecturer from Fakulti Kejuruteraan Elektrik (FKE) for his guidance, encouragement, knowledge and time, as it also helps me to completing this thesis. Then, I would also like to express my gratitude towards the panels during my proposal defence, Prof. Madya Ir. Dr. Abdul Rahim bin Abdullah, Prof. Madya Ts. Dr. Muhammad Fahmi bin Miskon and Prof. Madya Dr. Herman bin Jamaluddin for providing me the insight in improving my research. Finally, to my late father, my beloved mother, my sibling; my younger sister and younger brother, family and friends, I would like to acknowledge their continuous support throughout the project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

1D	-	One Dimensional
2D	-	Two Dimensional
3D	-	Three Dimensional
AD	-	Absolute Differences
AR	-	Augmented Reality
AMF	-	Adaptive Median Filter
ASW	-	Adaptive Support Weight
ASMF	-	Adaptive Switching Median Filter
AW	-	Adaptive Window
BF	-	Bilateral Filter
BP	-	Belief Propagation
CLAHE	-	Contrast Limited Adaptive Histogram Equalization
CPU	-	Central Processing Unit
CN	-	Census Transform
CNN	-	Convolutional Neural Network
DAMF	-	Different Applied Median Filter
DP	-	Dynamic Programing
FM	-	Feature Matching
FW	-	Fixed Window
GC	-	Graph Cut

GF	-	Guided Filter
GM	-	Gradient Matching
GPU	-	Graphical Processing Unit
LPS	-	Local Plane Sweep
LR	-	Left Right
LRC	-	Left Right Consistency
MF	-	Median Filter
MRF	-	Markov Random Field
MSDF	-	Modified Switching Median Filter
MST	-	Minimum Spanning Tree
NCC	-	Normalised Cross Correlation
PP	-	Post Processing
PSMF	-	Progressive Switching Median Filter
RANSAC	-	Random Sample Consensus
RT	-	Rank Transform
SAD	-	Sum Absolute Differences
SCL	-	Scattered Control Landmarks
SD	-	Squared Differences
SGM	-	Semi Global Method
SLIC	-	Simple Linear Iterative Clustering
SPP	-	Spatial Pyramid Pooling
SSD	-	Sum Squared Differences
SVDM	-	Stereo Vision Disparity map
TAD	-	Truncated Absolute Differences
WTA	-	Winner Take All

WGIF	-	Weighted Guided Image Filtering
ZNCC	-	Zero Normalised Cross Correlation

LIST OF SYMBOLS

b	-	Baseline of stereo camera
d	-	Disparity
f	-	Focal length of stereo camera
m	-	Gradient magnitude
p	-	Targeted Pixel
q	-	Neighbouring Pixels
w	-	Window size
z	-	Information of depth
E	-	Edge connects the vertices
I	-	Image
V	-	A set of vertices
(i, j)	-	Coordinates for neighbouring pixels
(x, y)	-	Coordinate of targeted pixel
$\|w\|$	-	number of pixels in support winndow
m_l	-	Gradient magnitude on left image
m_r	-	Gradient magnitude on right image
w_k	-	size for the support window
G_x	-	Gradient in horizontal direction
G_y	-	Gradient in vertical direction
I_l	-	Left image

I_p	-	pixel of interest at reference image
I_q	-	neighbouring pixels at reference image
I_r	-	Right image
μ_k	-	Mean value
σ_c	-	Colour similarity parameters
σ_s	-	Spatial distance
σ_k	-	Variance value

LIST OF PUBLICATIONS

Journal (Scopus Index)

1. Hamzah, R. A., Wei, M. 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.
2. Hamzah, R. A., Wei, M. G. Y., and Anwar, N. S. N., 2019. Stereo Matching based on Absolute Differences for Multiple Objects Detection. *Telkomnika*, *17*(1), pp.261 - 267.

Proceeding (Scopus Index)

1. Hamzah, R. A., Wei, M. G. Y. and Anwar, N. S. N., 2020. Depth Estimation Based on Stereo Image Using Passive Sensor. In *Advances in Electronics Engineering*, Springer, pp.127 - 136.
2. Gan, Y., Hamzah, R. A., & Anwar, N. N. (2018). Local Stereo Matching Algorithm based on Pixel Difference Adjustment, Minimum Spanning Tree and Weighted Median Filter. In *2018 IEEE Conference on Systems, Process and Control (ICSPC)*, pp. 39 - 43.

Journal (Non-Scopus Index)

1. Gan, Y., Hamzah, R. A., and Anwar, N. N., 2019. A New Structure of Stereo Algorithm Using Pixel Based Differences and Weighted Median Filter. *International Journal of Human and Technology Interaction (IJHaTI)*, *3*(2), pp. 23 - 28.

CHAPTER 1

INTRODUCTION

This chapter provides a brief introduction of stereo vision technology and is divided into six sections. Here, section 1.1 introduces the background of stereo vision technology. Next, section 1.2 briefly explains some applications for stereo vision technology. Then, section 1.3 explains the problem statements and section 1.4 presents the objectives of this research. After that, section 1.5 describes the scope of this research and finally, section 1.6 provides the overall outline of the thesis.

1.1 Background

With a pair of eyes, humans are able to estimate depth and identify obstacles. Then, animals with binocular vision, which are normally hunters, uses their eyes to estimate depth for hunting. For living organism, the ability to estimate depth is naturally implemented by the brain but not for a computer. However, the ability to recognize the depth can be modelled mathematically (Bhatti, 2012) using the stereo vision system. Stereo vision technology uses two cameras, where they are aligned horizontally, mimicking the human eyes, the left side and the right side, which are presented in Figure 1.1, obtained from Hariyama et al. (2008). When a scene is captured, two input images are produced, the left image and the right image. Then these input images also provided two different viewpoints from a single scene. In stereo vision, a minimum of two viewpoints are need and these inputs images (left image and right image), represents those viewpoints. By using the mathematical models and computer computation, the input images will correspond with each other and produce disparity map. In addition, the disparity map is later used for three dimensional (3D) reconstruction to obtain depth map. Hence, a good accuracy disparity map is always preferable.

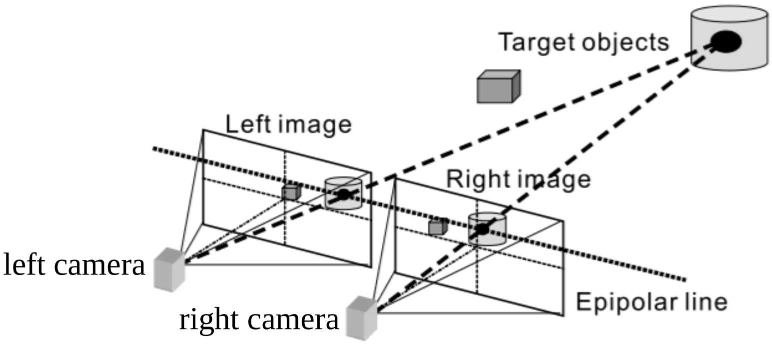

Figure 1.1: Stereo vision system

The basic of a stereo vision is presented in Figure 1.2, obtained from Fahmy et al. (2013). *P(x,y,z)* represents Scene Point, *P* where is then observed at points $P_l(x_l, y_l)$ and $P_r(x_r, y_r)$ at the left image and right image, respectively. Then, *f* represents focal length and *b* represent baseline; the distance between the two cameras.

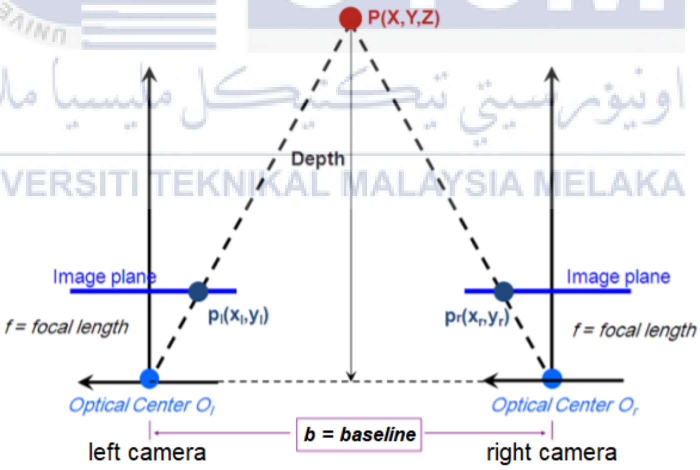

Figure 1.2: Stereo vision basics

The difference in position of the corresponding points for the respective images is defined as disparity, *d* and the respective images refer to the input images, left image and right image. The equation of disparity, *d* obtained from Adi and Widodo (2017), is presented in equation (1.1),

$$d = x_l - x_r \tag{1.1}$$