UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DISPARITY MAP ALGORITHM USING HIERARCHICAL OF BITWISE PIXEL DIFFERENCES AND SEGMENT-TREE FROM STEREO IMAGE

MUHAMMAD NAZMI BIN ZAINAL AZALI

MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

2024

Faculty of Electronics and Computer Technology and Engineering

DISPARITY MAP ALGORITHM USING HIERARCHICAL OF BITWISE PIXEL DIFFERENCES AND SEGMENT-TREE FROM STEREO IMAGE

Muhammad Nazmi bin Zainal Azali

Master of Science in Electronic Engineering

2024

DISPARITY MAP ALGORITHM USING HIERARCHICAL OF BITWISE PIXEL DIFFERENCES AND SEGMENT-TREE FROM STEREO IMAGE

MUHAMMAD NAZMI BIN ZAINAL AZALI

A thesis submitted
in fulfillment of the requirements for the degree of
Master of Science in Electronic Engineering

Faculty of Electronics and Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2024

DEDICATION

I dedicate this entire study to Allah Almighty, the source of wisdom, patience, and inspiration. With His grace, I have successfully completed this research. My heartfelt gratitude goes to my dear mother and father, who have been a constant source of blessing and encouragement throughout the journey of conceptualizing and completing this research. Special thanks to my wife and family for their unwavering support during this research endeavor. I extend my dedication to my supervisor, Ts. Dr. Rostam Affendi bin Hamzah, and co-supervisor, Dr. Zarina binti Mohd Noh, for their valuable collaboration and guidance that contributed to the successful completion of this research work. Lastly, I express my gratitude to God for the gift of nature, and it brings immense joy to share that during this research, my son, Muhammad Naaz Mikhael bin Muhammad Nazmi, was born.

ABSTRACT

In computer vision technology, stereo matching algorithm plays an important role in generating disparity map or depth map through a correspondence process from stereo images. The algorithm development can be categorized into local, global, and semi-global methods. Global method produces high computational complexity and slow implementation, deferring its suitability for real-time application. Local methods excel in solving matching problems through local-based analysis with fast execution and low computational demands. Combining attributes from both, the semi-global method introduces more complex structure and high computational complexity. This thesis presents a local-based stereo matching algorithm to increase the accuracy on complex regions. These regions are low texture, repetitive patterns, illumination differences, discontinuity, and occlusion. The proposed algorithm has four stages that start with a novel bitwise pixel-based differences at matching cost computation. This stage utilizes XOR gate to produce the initial disparity map. The next stage involves the utilization of Segment Tree (ST) to eliminate the noise at aggregation step. Then, an optimization stage employs Winner-Take-All (WTA) strategy. The final step of the proposed algorithm framework is refinement stage. At this stage, Bilateral filter (BF) and Weighted Median (WM) filter are utilized. These filters not only increase the accuracy but are also capable of preserving the object's edges. Then, hierarchical Gaussian pyramid is applied at each stage to further enhance the final disparity map. The performance evaluation of the proposed algorithm is conducted using two standard online benchmarking databases, which are the Middlebury Stereo for quantitative metrics and Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) for qualitative assessments. The adaptability of the algorithm is demonstrated through a 3D surface reconstruction using a final disparity map. In conclusion, the proposed algorithm displays significant efficiency, yielding an average non-occlusion error of 5.61% and an all-error rate of 8.92%. Hence, the proposed algorithm is competitive with other existing methods, especially when incorporating the pyramid method over non-pyramid approaches.

ALGORITMA PETA KETAKSAMAAN MENGGUNAKAN HIERARKI PERBEZAAN BIT PIKSEL DAN SEGMEN-PEPOHON DARIPADA IMEJ STEREO

ABSTRAK

Dalam teknologi penglihatan komputer, algoritma pemadanan stereo memainkan peranan penting dalam menghasilkan peta ketaksamaan atau peta kedalaman melalui proses kesepadanan daripada imej stereo. Pembangunan algoritma boleh dikategorikan kepada kaedah tempatan, global dan separa global. Kaedah global menghasilkan kekompleksan pengiraan yang tinggi dan pelaksanaan yang perlahan, mengurangkan kesesuaiannya untuk aplikasi masa nyata. Kaedah tempatan baik dalam menyelesaikan masalah padanan melalui analisis berasaskan tempatan dengan pelaksanaan pantas dan pengiraan komputasi yang rendah. Menggabungkan atribut daripada keduanya, kaedah separa global memperkenalkan struktur yang lebih kompleks dan kerumitan pengiraan yang tinggi. Tesis ini membentangkan algoritma pemadanan stereo berasaskan tempatan bagi meningkatkan ketepatan pada kawasan kompleks. Kawasan ini adalah tekstur rendah, corak berulang, perbezaan pencahayaan, ketakselanjaran dan oklusi. Algoritma yang dicadangkan mempunyai empat peringkat bermula dengan perbezaan berasaskan bit piksel baharu pada pengiraan kos sepadan. Peringkat ini menggunakan get XOR untuk menghasilkan peta ketaksamaan awal. Peringkat seterusnya melibatkan penggunaan Segmen-Pepohon (SP) bagi menghapuskan hingar pada langkah pengagregatan. Kemudian, peringkat pengoptimuman menggunakan strategi Winner-Take-All (WTA). Langkah terakhir kerangka algoritma yang dicadangkan ialah peringkat perbaikan. Pada peringkat ini, penapis Dwisisi (PD) dan penapis Pemberat Median (PM) digunakan. Penapis-penapis ini bukan sahaja meningkatkan ketepatan tetapi juga mampu mengekalkan sisi objek. Kemudian, piramid Gaussian berhierarki digunakan pada setiap peringkat untuk menambahbaik peta ketaksamaan akhir. Penilaian prestasi algoritma yang dicadangkan menggunakan dua piawaian pangkalan data penandaarasan dalam talian, iaitu Stereo Middlebury untuk metrik kuantitatif dan Institut Teknologi Karlsruhe dan Institut Teknologi Toyota (KITTI) untuk penilaian kualitatif. Kebolehsuaian algoritma ditunjukkan melalui pembinaan semula permukaan 3D menggunakan peta ketaksamaan akhir. Kesimpulannya, algoritma yang dicadangkan memaparkan kecekapan yang ketara, menghasilkan purata ralat bukan oklusi sebanyak 5.61% dan kadar ralat semua sebanyak 8.92%. Oleh itu, algoritma yang dicadangkan berdaya saing dengan kaedah sedia ada yang lain, terutamanya apabila menggabungkan kaedah piramid berbanding pendekatan bukan piramid.

ACKNOWLEDGEMENT

In the Name of Allah, the Most Gracious, the Most Merciful. First and foremost, I would like to take this opportunity to express my sincere acknowledgement to my supervisor Ts. Dr. Rostam Affendi bin Hamzah from Faculty of Electronics and Computer Technology and Engineering (FTKEK), Universiti Teknikal Malaysia Melaka (UTeM) for his essential supervision, support, guidance, patience, and encouragement towards the completion of this thesis. I would also like to express my greatest gratitude to Dr. Zarina binti Mohd Noh from Faculty of Electronics and Computer Technology and Engineering (FTKEK), co-supervisor of this research for her evaluation guidance and ideas, as well as her assistance in completing this thesis. Special thanks to UTeM FRGS grant funding for the financial support throughout this project. Particularly, I would also like to express my gratitude towards the panels during my proposal defenses, Prof. Madya Dr. Nurulfajar bin Abd Manap, Dr. Suhaila binti Mohd Najib and Ts. Dr. Muhammad Noorazlan Shah Zainudin for providing me the insight in improving my research. Special thanks to all my peers, beloved family, and my wife for their moral support in completing this degree. In addition, thank you to everyone who has been to the crucial parts of realization of this project.

Last but not least, I want to thank me for believing in me. I want to thank me for doing all the hard work. I want to thank me for having no days off. I want to thank me for never quitting. I want to thank me for always being a giver and try to give more than I receive. I want to thank me for trying to do more right than wrong. I want to thank me for just being me at all times.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	-	Two Dimensional
3D	-	Three Dimensional
ACR-GIF-OW	-	Adaptive Cross-Region-Based Guided Image Filtering
AD	-	Absolute Differences
ADSM	-	Adaptive Deconvolution-Based Stereo Matching
AMF	-	Adaptive Median Filter
AR	-	Augmented Reality
ASW	-	Adaptive Support Weights
Avg	-	Average
AW	-	Adaptive Window
BF	-	Bilateral Filter
CA	-	Cost Aggregation
CLAHE	-	Contrast Limited Adaptive Histogram Equalization
CM	-	Cost Matrix
CMOS	-	Complementary Metal-Oxide-Semiconductor
CNN	-	Convolutional Neural Network
CPU	-	Central Processing Units
CT / CEN	-	Census Transform
DAMF	-	Different Applied Median Filter
DoG	-	Difference-of-Gaussians
DR	-	Disparity Refinement
DS	-	Disparity Selection
GB	-	Gigabyte

GBM	-	Global Matching
GC	-	Graph Cuts
GCPs	-	Ground Control Point
GF	-	Guided Filter
GHz	-	Gigahertz
GIF	-	Guided Image Filtering
GM	-	Gradient Matching
GMM	-	Gaussian Mixture Model
GPU	-	Graphics Processing Unit
INV	-	Illumination Normal Vector
JEM	-	Joint Energy Minimization
JWGF	-	Joint Weighted Guided Filter
LCA	-	Locally Competitive Algorithm
LR	-	Left-Right
MCC	-	Matching Cost Computation
MF	-	Median Filter
MPSoC	-	Multiprocessor System on a Chip
MRF	-	Markov Random Field
MSDF	-	Modified Switching Median Filter
MST	-	Minimum Spanning Tree
MW	-	Multi-Window
NCC	-	Normalized-Cross-Correlation
PSMF	-	Progressive Switching Median Filter
RAM	-	Random-Access Memory
RGB	-	Red, Green, and Blue
RT	-	Rank Transform

SAD	-	Sum of Absolute Differences
SD	-	Squared Differences
SGBM	-	Semi-Global Block Matching
SGM	-	Semi Global Method
SLAM	-	Simultaneous Localization and Mapping
SNR	-	Signal to Noise Ratio
SSD	-	Sum of Squared Differences
ST	-	Segment-Tree
TADc	-	Truncated Absolute Difference Based on Color Channels
TADg	-	Truncated Absolute Difference Based on Image Gradient
WM / WMF	-	Weighted Median Filter
WTA	-	Winner-Take-All
ZNCC	-	Zero-Mean-Normalize-Cross-Correlation

LIST OF SYMBOLS

b	-	Baseline of Stereo Camera
d	-	Disparity
E	-	The Edge Connects the Vertices
f	-	Focal Length of The Stereo Camera
I	-	Image
k	-	Constant Parameter
l	-	Left Images
p	-	Pixel of Interest
q	-	Neighboring Pixels
r	-	Right Image
s	-	The Scale Levels
S_i	-	The Scale of Change at The Pyramid Level
V	-	Set of Vertices That Corresponds to The Data Set
w	-	Window Size
W_c	-	A Set of Displacement Windows
ξ	-	The Transform
E_{smooth}	-	Energy Smoothness
e_j	-	Each Edge
β	-	The Weighting Factor
μ_k	-	Mean Value
σ_k	-	Variance Value
(i, j)	-	Coordinates for Neighboring Pixels
(x, y)	-	Coordinate of Targeted Pixel

^	-	XOR Logic Gate
$\|w\|$	-	Number of Pixels in Support Window
$\|p - q\|$	-	Spatial Euclidean
$\otimes$	-	Bit String Set

LIST OF APPENDICES

LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

Journal (Scopus Index)

Azali, M. N. Z., Hamzah, R. A., Noh, Z. M., Abidin, I. Z., and Wook, T. M. F. T., 2023. Stereo matching algorithm using census transform and segment tree for depth estimation. *Telkomnika (Telecommunication Computing Electronics and Control)*, *21*(1), 150–158.

Hamzah, R. A., Azali, M. N. Z., Noh, Z. M., Zahari, M., and Herman, A. I., 2022. Development of depth map from stereo images using sum of absolute differences and edge filters. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(2), 875–883.

Nazmi, Z. A. M., Hamzah, R. A., Zarina, M. N., and Madiha, Z., 2022. Disparity Map from Stereo Images for Three-dimensional Surface Reconstruction. *Engineered Science*, *19*, 1–8.

Proceeding (Scopus Index)

Azali, M. N. Z., Hamzah, R. A., and Noh, Z. M., 2022. Disparity Map Algorithm Using Census Transform and Hierarchical Segment-Tree from Stereo Image. *IET Conference Proceedings*, *2022*(22), 244–249.

CHAPTER 1

INTRODUCTION

1.1 Background

There are seven subsections in this chapter, which provides a complete summary of this thesis and stereo vision system introduction. Throughout section 1.2, the development of stereo vision system is described. The applications of stereo vision systems are briefly explained in section 1.3, and the challenges discovered during the research for this thesis are discussed in section 1.4. The problem statements are then described in section 1.5, and the research's objectives are then laid out in section 1.6. The research's scope is then discussed in part 1.7, and section 1.8 concludes by outlining the thesis overall structure.

1.2 Introduction

The stereo vision is a part of the important field in computer vision, which provides numerous algorithms for computing various image processing-related areas of research. Fundamentally, a stereo image will go through the stereo matching algorithm, which is based on a specific technique to generate a depth map. In instance, the depth map is sometimes referred to as a disparity map in Geiger et al. (2012); Hamzah et al. (2021) arguments. The process begins with a pair of stereo images denoted as left and right images, from which the scene depth can be retrieved from two separate points with certain baseline displaced values. Stereo matching produces the correlation values of the left image compared to the right image. Figure 1.1 provides a clear illustration of the general process and viewpoint of stereo vision. According to Budiharto et al. (2011); Winarno et al. (2016), the stereo matching process using the disparity map as a base generates the depth map by altering the intensity of the pixel values on the map.

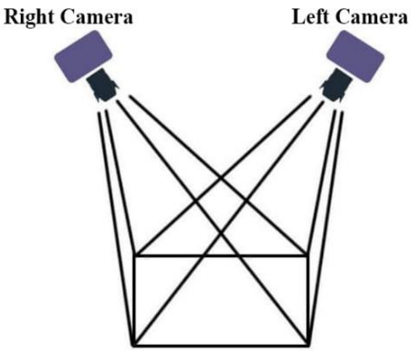

Figure 1.1 Stereo Vision angle and viewpoint

Local-based matching, global matching, semi-global matching, and deep learning are the methods that can be used for stereo matching process, although they vary in performance and methodology. Additionally, comprehensive descriptions of each approach are explained as follows:

a. Local-based matching: The depth of two images is estimated using a simple and fast pixel-to-pixel or block matching technique, which compares small, fixed-size blocks of pixels in each image. The algorithm first divides the image into manageable blocks, which match the blocks in the left and right images. The block with the best match in the other image is selected, and the disparity value is determined using the displacement of the blocks. Although pixel-to-pixel and block matching are fast approaches, errors caused by occlusions, texture less region, and repetitive patterns heavily can occur. Block Matching has demonstrated its effectiveness as shown by the Janeczek et al. (2017) analysis. Based on their findings, the local-based method demonstrates high noise, but it is capable of operating both dependably and in real-time due to fast execution process.

b. Global matching (GBM): GBM is a more complex technique that estimates depth using an energy of global optimization technique. GBM examines the energy of entire images instead of discrete blocks of pixels Zeglazi et al. (2018). Compared to the block matching method, GBM is more accurate but has a high processing time. In contrast, according to research by Yang et al. (2019), the global method uses energy minimization approach from the Markov Random Field (MRF) technique. The aggregation stage is not included in the framework for this method.

c. Semi-global matching: Generally, this approach is complex, and the structure employs the combination of local and global based framework. To increase the precision of the disparity map values, this technique employs a cost aggregation phase. Based on the intensity discrepancies between each pixel in the left image and its corresponding pixels in the right image, the algorithm determines a cost for each pixel. Once the costs have been combined, then the cost volume will be produced from all potential differences. While slower than block matching, SGBM is more accurate. Semi-Global Block matching (SGBM) is a precise stereo algorithm method that can be managed significant displacements and be used in real-time, according to Hirschmüller (2008).

d. Deep Learning: Deep learning is another complex method that teaches neural networks how the left and right images are correlated. In order to anticipate the disparity map, Convolutional Neural Network (CNN) is frequently employed to extract information or features from the input stereo images. Deep learning-based techniques have a high level of accuracy, but they need a lot of labelled data and computing power to train. Complex structure of