IMPROVEMENT OF LOCAL-BASED STEREO VISION DISPARITY MAP ESTIMATION ALGORITHM

ROSTAM AFFENDI HAMZAH

UNIVERSITI SAINS MALAYSIA

2017

IMPROVEMENT OF LOCAL-BASED STEREO VISION DISPARITY MAP ESTIMATION ALGORITHM

by

ROSTAM AFFENDI HAMZAH

**Thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy**

June 2017

ACKNOWLEDGEMENT

First of all, I express my gratitude to the Almighty Allah SWT, who is the ultimate source of guidance in all our endeavors. Next, I would like to express my sincere gratitude to my supervisor; Associate Professor Dr Haidi Ibrahim and my co-supervisor; Dr Anwar Hasni Abu Hassan for giving me the opportunity to work under their supervision. I would like to convey my thanks for their insightful guidance and encouragement throughout the research. Thanks for their advices, ideas and suggestions in accomplishing this research work. I have learned a lot from them about doing research and presenting the results.

I would like to thank the School of Electrical & Electronic Engineering, Universiti Sains Malaysia (USM) to provide the research platform during all these years. I would also like to thank the Institute of Postgraduate Studies (IPS), USM which provides a sponsorship to one of my oral presentation in international conference through IPS fund. I am grateful for all the supports and helps from my colleagues; Nik Shahrim Nik Anwar, Mohd Rahmat Arifin, Sumariamah Mohd Radzi and Low Wei Zeng, assistant engineers; Khairul Anuar Ab. Razak and senior lab assistant; Nor Azhar Zabidin, who were always there to support me in any needs. Special thanks to my sponsorship from Ministry of Higher Education under Skim Latihan Akademik Bumiputra (SLAB) and Universiti Teknikal Malaysia Melaka (UTeM). Lastly, I could never finish this thesis without the support from my family. I wish to express my love and gratitude to my wife, my parents and my kids for all their supports, and always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
AD	Absolute Differences
ALD	Arm Length Differences
AR	Augmented Reality
ASW	Adaptive Support Weight
AW	Adaptive Window
BF	Bilateral Filter
BFV	Bitwise Fast Voting
BP	Belief Propagation
BXF	Box Filter
CPU	Cental Processing Unit
CN	Census Transform
CSCN	Center Symmetric Census Transform
DoG	Difference of Gaussian
DP	Dynamic Programming
FM	Feature Matching
FPGA	Field Programable Gate Array
FW	Fixed Window
GC	Graph Cut
GCP	Ground Control Points
GCSF	Growing Scene Flow
GF	Guided Filter
GM	Gradient Matching
GPU	Graphical Processing Unit
HBDS	Hierarchical Bilateral Disparity Structure
JBF	Joint Bilateral Filter

LPF	Low Pass Filter
LPS	Local Plane Sweep
LR	Left-Right
LS	Least Square
MeanShiftSeg	Mean Shift Segmentation
MF	Median Filter
MorSeg	Morphological Segment
MRF	Markov Random Field
MST	Minimum Spanning Tree
MW	Multiple Window
NCC	Normalised Cross Correlation
PCL	Point Cloud Library
RAM	Random Access Memory
RANSAC	Random Sample Consensus
RT	Rank Transform
SAD	Sum of Absolute Differences
SCL	Scattered Control Landmarks
SD	Square Differences
SGM	Semi Global Method
SIFT	Scale Invariant Feature Transform
SSD	Sum of Square Differences
ST	Spanning Tree
SVDM	Stereo Vision Disparity Map
WBF	Weighted Bilateral Filter
WTA	Winner Takes All
ZNCC	Zero Normalised Cross Correlation

LIST OF SYMBOLS

a	Constant parameter in plane fitting
A	Pixel size of camera
b	Baseline
C	Component of a segment
d	disparity
e	Epipolar line
f	Focal length
G_x	Horizontal direction
G_y	Vertical direction
h	Kernel bandwidth
I	Guidance image
I_l	Image left
I_r	Image right
k	Constant parameter in segmentation
K	Kernel density
m	Pixel coordinates in a segment
m_l	Magnitude value of left image
m_r	Magnitude value of right image
n	Iteration number
N	Maximum disparity value
p	Coordinates pixel of interest
q	Neighbouring pixels
R	Range value
S	Segment
v_p	Vertice of point p
v_q	Vertice of point q
w	Window support
w_c	Window support of cost aggregation

w_{CN}	Window size of CN
w_g	Support window of guidance image
w_p	Support window of BF
x_l	Position of left plane projection
x_r	Position of right plane projection
Z	Depth
z_c	Size of a component
μ	Mean value
σ	Variance value
$\otimes$	Bitwise catenation
$\circledast$	Convolution sum operation
ε	Constant value of smoothness term
β	Constant value of per-pixel difference
α	Constant value to balance color and gradient terms
τ_{AD}	Truncated value of AD
τ_{GM}	Truncated value of GM
τ_{LR}	Constant value of disparity map validation
τ_{plane}	Threshold value of plane fitting
τ_{seg}	Threshold value of segmentation
σ_s	Spatial adjustment parameter
σ_c	Disparity similarity parameter
ω_{seg}	Weight difference of a segment
Δ	Internal difference of a segment
δ	Average distance

PENAMBAHBAIKAN ALGORITMA PENGANGGARAN PETA PERBEZAAN PENGLIHATAN STEREO SECARA TEMPATAN

ABSTRAK

Anggaran Peta Perbezaan Penglihatan Stereo (PPPS) adalah satu topik penyelidikan yang aktif dalam penglihatan komputer. Untuk meningkatkan ketepatan PPPS adalah sukar dan mencabar. Ketepatan dipengaruhi oleh rantau dari sisi tak selanjar, bertutup, corak berulang dan bertekstur rendah. Oleh itu, tesis ini mencadangkan algoritma untuk pengendalian yang lebih cekap bagi cabaran ini. Pertama, algoritma PPPS yang dicadangkan menggabungkan tiga ciri pengiraan kos padanan berasaskan perbezaan setiap piksel. Gabungan ciri Perbezaan Mutlak (PM) dan Padanan Kecerunan (PK) mengurangkan herotan radiometrik. Kemudian, kedua-dua perbezaan digabungkan dengan Transformasi Banci (TB) untuk mengurangkan kesan perbezaan pencahayaan. Kedua, tesis ini membentangkan teknik baru pengendalian sisi tak selanjar yang dinamakan Penapis Berpandu Lelaran (PBL). Teknik ini diperkenalkan untuk memelihara dan menambah baik sempadan objek. Akhirnya, proses-proses pengisian perbezaan tak sah, peruasan graf tak berarah dan pemadanan satah digunakan di peringkat terakhir untuk memulihkan rantau bertutup, corak berulang dan bertekstur rendah pada PPPS. Berdasarkan keputusan eksperimen data penandaarasan piawai dari Middlebury, algoritma yang dicadangkan ini dapat mengurangkan masing-masing 17.17 % dan 18.11 % daripada ralat *semua* dan $tidakbertutup$, berbanding dengan tanpa rangka kerja yang dicadangkan. Tambahan lagi, rangka kerja yang dicadangkan mengatasi sebahagian daripada algoritma terkini dalam literatur.

IMPROVEMENT OF LOCAL-BASED STEREO VISION DISPARITY MAP ESTIMATION ALGORITHM

ABSTRACT

Stereo Vision Disparity Map (SVDM) estimation is one of the active research topics in computer vision. To improve the accuracy of SVDM is difficult and challenging. The accuracy is affected by the regions of edge discontinuities, occluded, repetitive pattern and low texture. Therefore, this thesis proposes an algorithm to handle more efficiently these challenges. Firstly, the proposed SVDM algorithm combines three matching cost features based on per pixel differences. The combination of Absolute Differences (AD) and Gradient Matching (GM) features reduces the radiometric distortions. Then, both differences are combined with Census Transform (CN) feature to reduce the effect of illumination variations. Secondly, this thesis also presents a new method of edge discontinuities handling which is known as iterative Guided Filter (iGF). This method is introduced to preserve and improve the object boundaries. Finally, the fill-in invalid disparity, undirected graph segmentation and plane fitting processes are utilized at the last stage in order to recover the occluded, repetitive and low texture regions on the SVDM. Based on the experimental results of standard benchmarking dataset from the Middlebury, the proposed algorithm is able to reduce 17.17% and 18.11% of *all* and *nonocc* errors, respectively, as compared to the algorithm without the proposed framework. Moreover, the proposed framework outperformed some of the state-of-the-arts algorithms in the literature.

CHAPTER ONE

INTRODUCTION

This chapter is divided into seven sections. Section 1.1 introduces the background of stereo vision system. The introduction consists of basic fundamental explanation based on mathematical models. Then, Section 1.2 gives examples of stereo vision applications. Section 1.3 provides the research challenges and Section 1.4 describes the problem statement. Section 1.5 presents the objectives of this thesis. After that, Section 1.6 and 1.7 explain about the scope and structures of this thesis, respectively.

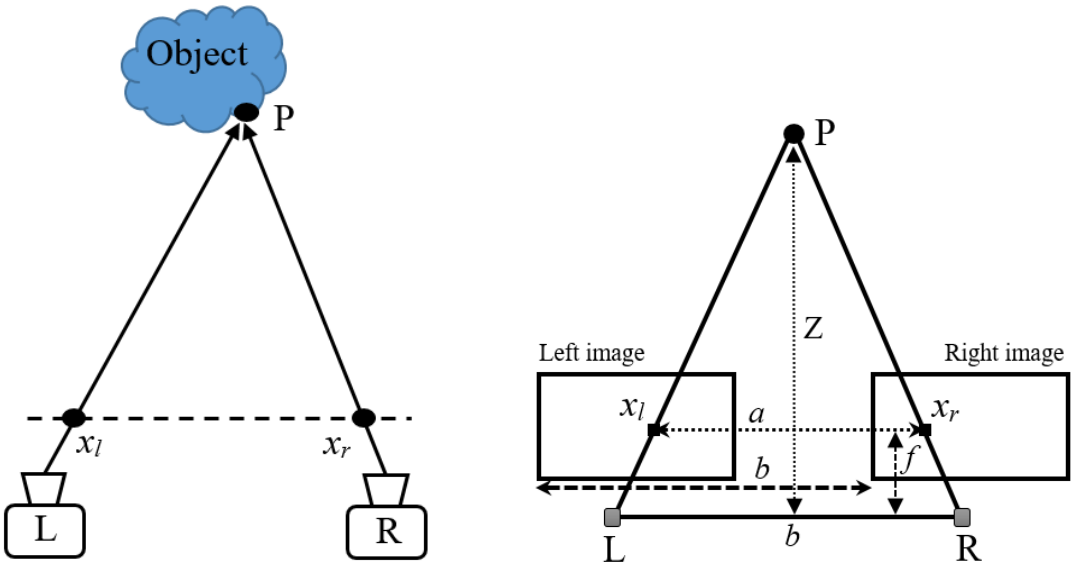
1.1 Background of Stereo Vision

Human vision is capable to recognize the depth easily through the stereoscopic fusion from the eyes. This job is automatically implemented by the human brain. The depth of a scene from stereoscopic fusion is also can be modeled mathematically (Bhatti, 2012). This model is called as stereo vision system which is one of the most active and important research areas in computer vision. Stereo vision consists of two cameras (i.e., left and right) which perceives one scene from two different viewpoints. These two viewpoints are processed permitting the visual depth data to be recovered. The process involves in computation of three-dimensional (3D) information of the scene from two-dimensional (2D) input images. The depth information of stereo images can be acquired by shifted together to discover the parts or pixels that match each other. The shifted value is named as disparity (Xu & Zhang, 2013). The higher disparity value means the object is closer to the cameras. The disparity value is nearly zero if the object is far from the cameras. This indicates the same pixel location of the left and right images.

Figure 1.1 shows a basic concept of stereo vision system and its translation of mathematical models (Ma et al., 2012). Figure 1.1(a) shows stereo sensor (i.e., L=left camera, R=right camera) detects an object at point P with the same viewpoint. The horizontal dotted line is the plane projection of a stereo system which image P at left and right cameras are placed at pixel locations of x_l and x_r respectively. Figure 1.1(b) shows the translation of stereo vision geometry. At the plane projection views, the left camera produces left image (i.e., Left image) which the matching point is located at x_l coordinate. The right camera produces right image (i.e., Right image) which the matching pixel is located at x_r. The distance between L and R is baseline range b and a is the distance of matching pixel coordinates (i.e., between x_l and x_r). Fundamentally, based on the triangulation principle, the angle of $(\angle L,P,R)$ and $(\angle x_l,P,x_r)$ is similar which enables to compute the depth based on Equation (1.1):

$$\frac{b}{Z} = \frac{a}{Z-f} = \frac{(b-x_l)+x_r}{Z-f} \tag{1.1}$$

where b denotes the baseline of stereo camera sensor and Z is the depth or distance. The



(a) Stereo vision sensor with an object detection at point P.

(b) Translation of stereo vision geometry.

Figure 1.1: A stereo vision system which contains a point detection and its translation model.

x_l and x_r are the coordinates of plane projections on matching pixel and f represents the stereo camera focal length.

After further calculation, the final depth estimation is given by Equation (1.2):

$$Z = \frac{bf}{x_l - x_r} = \frac{bf}{d} \tag{1.2}$$

where $d = x_l - x_x$ is the disparity value. This value can be plotted into 2D map which is known as disparity map. This map is important and contains of information for stereo vision applications. The process or algorithm of estimating the Stereo Vision Disparity Map (SVDM) is based on the taxonomy which was developed by Scharstein and Szeliski (2002). They categorized three major methods in SVDM development (i.e., global, Semi Global (SGM) and local methods). The framework of SVDM consists of four main steps (i.e., Step 1: matching cost computation; Step 2: cost aggregation; Step 3: disparity selection and optimization; Step 4: disparity map refinement). The mentioned steps will be described extensively in Chapter 2.

1.2 Application of Stereo Vision

The stereo vision system covers a wide range of applications such as:

(i). Augmented Reality (AR): Stereo vision information is an important element of AR systems which depends on the accurate depth estimation of a scene. This is to put an accurate position of computer created objects with real life video which was implemented by Markovic et al. (2014), Suenaga et al. (2015).

(ii). Robotic and automotive applications : Industrial robotic inspection and autonomous robot navigation involves in static and dynamic environments. It requires the infor-

mation of realistic motion and depth estimation. Stereo vision can be used efficiently to estimate the depth which was implemented by Dinham and Fang (2013), Di Fulvio et al. (2014), Philipsen et al. (2015).

(iii). 3D surface reconstruction: The analysis of 3D surface reconstruction is important to determine the status and conditions of an object or environment for example in archaeological artifact observation by Dellepiane et al. (2013) and 3D terrain reconstruction by Correal et al. (2014).

1.3 Research Challenges

The accuracy of SVDM algorithm might be affected by several factors. These factors are labeled by alphabets in Figure 1.2 which consist of four main challenges and are explained as follows:

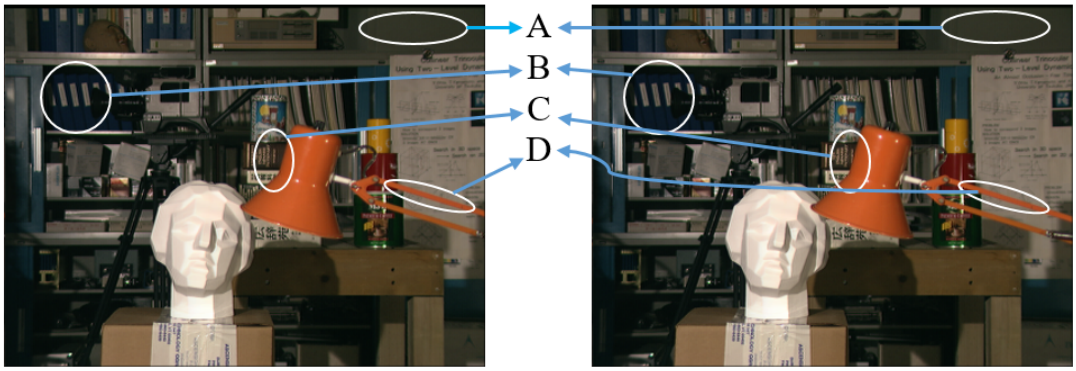


Figure 1.2: A stereo image (i.e., (a) left image (b) right image) of Tsukuba is mapped based on the research challenges (Kordelas et al., 2015).

(i). A (Low texture regions)

The areas labeled A are the most difficult region for the SVDM algorithm to do the matching process. These regions on an image are caused by the plain colour surface and textureless surface regions. Any small regions from the circle in Figure 1.2(a)

image could similarly match to the region within the circle in Figure 1.2(b) image. Additionally, the larger low texture regions on both of the stereo images, it becomes more difficult and challenging due to the pixel intensities look alike to each other.

(ii). B (Repetitive regions)

The second challenge is the areas labeled B. These areas contain the regions with periodic and repetitive surface texture. The algorithm trying to match the pixels on Figure 1.2(a) image with the circle on Figure 1.2(b) image which has a number of possible intensity values may be allocated. The difficulty of matching process occurs when the SVDM algorithm uses wrong matching coordinates. Generally, space objects and man-made objects will normally have many repetitive textures, so this is unavoidable to be a problem that the algorithm must take into consideration.

(iii). C (Occluded regions)

The areas labeled C are the occluded regions. These regions contribute to most general type of difficulty for a stereo matching algorithm. Notice in Figure 1.2(a) image one book is not visible, but matching the similar region in Figure 1.2(b) image the book is almost visible behind the table lamp. Because of the geometric displacement between the cameras, one of the scene is causing another to not be visible to both cameras. Apparently something that cannot be seen by both cameras unable to be matched between the images. On the disparity map, the occluded regions are very hard to estimate or to be filled-in with accurate disparity values. This is because the unknown objects, shapes or structures behind the occluded regions. These regions are getting bigger and hard to be corrected when the baseline of the stereo sensor is expanded.

(iv). D (Discontinuity regions)

A final challenge to SVDM algorithms are depth discontinuities as shown by the

table lamp holder marked by the letter D. The challenge is because of the stereo algorithms use a predetermined sized mask from one image to localize within the other image. If this mask contains the information from the front-most surface and the rear-most surface across a depth discontinuity, several correct disparity values could be assigned. Usually, this leads to increase the error across the depth boundaries. It makes more difficult to get the corresponding points if the discontinuity region sizes are different drastically between the stereo image.

1.4 Problem Statements

The focus of this thesis is to develop a new SVDM algorithm to produce accurate results. This will benefit to expand the relevant of stereo vision in areas that implicate the depth estimation. Even though the SVDM algorithms have been studied for years, the low texture regions, repetitive patterns, and occluded regions are the attributes of difficulties in the SVDM development. Yang (2012) (i.e., SSD), Mei et al. (2013) (i.e., SAD) and Zhu et al. (2015) (i.e., NCC) used window-based techniques at matching cost computation which resulting the disparity map heavily exposed to high noise. The improper or wrong window size selections, it may causes problems at incorrect disparities in the object edges and occlusion boundaries. If the window size is too large and consists of object bound-aries, it will assume similar intensity values which this will make an incorrect assumption. Hence, the fattening effects occurred on the results. While small window size will escape the important information crossing the depth discontinuities. The matching cost compu-tation is the most important step which provides the preliminary performance of SVDM algorithm. Thus, this step must have robust function and minimal noise.

Some existing SVDM algorithms were sensitive to the low texture regions which these algorithms could not determine the correct disparity values on the plain colour regions.