Faculty of Engineering and Information Technology
University of Technology Sydney

High-Quality Depth Maps Acquisition for RGB-D Data

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Yifan Zuo

April 2018

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This thesis is the result of me conducted jointly with Shanghai University as part of a collaborative Doctoral degree.

Signature of Candidate: ________________

Date: ________________

Acknowledgments

For two years in UTS which is a part of dual-doctoral plan, my deepest gratitude goes first and foremost to Dr. Qiang Wu, my supervisor, for the constant encouragement and guidance for my Ph.D work. Through weekly meetings, his immense knowledge and motivation help me address problems in research as well as writing this thesis efficiently.

I would like to express my heartfelt gratitude to my co-supervisor Associate Prof. Jian Zhan and my domestic supervisor Prof. Ping An in Shanghai University for providing me with continuous support throughout my PhD study and research.

I thank my labmates in Global Big Data Technology Center: Xiaoshui Huang, Junjie Zhang, Lina Li, Peng Zhang and Zongjian Zhang and labmates in Shanghai University: Xiwu Shang, Chao Yang, Deyang Liu, Jian Ma and Jianxin Wang for the help and discussions and for all the fun we have had in my PhD study period.

Last but not the least, my thanks would go to my parents for their support and great confidence in me all through these years.

Yifan Zuo
September 2017 @ UTS

Contents

Certificate . . . . **i**

Acknowledgment . . . . **ii**

List of Figures . . . . **vii**

List of Tables . . . . **x**

List of Publications . . . . **xii**

Abstract . . . . **xiv**

Chapter 1 Introduction **1**

1.1 Background 1

1.2 Research Problems 6

1.2.1 Fast Depth Map Estimation 6

1.2.2 Temporal Consistency Enhancement for Depth Video . 7

1.2.3 Explicit Measurement for Edge Inconsistency between Depth Map and Color Image 7

1.2.4 Improved Guided Depth Enhancement via Predefined Model 8

1.2.5 Guided Depth Enhancement via Machine Learning . . 8

1.3 Research Contributions 8

1.4 Thesis Structure 9

Chapter 2 Relevant Theories and Related Work **12**

2.1 Foundation for Passive Depth Acquisition via Stereo Matching 12

2.1.1 Concepts 13

2.1.2 Stereo Matching Constraints 14

2.1.3 Basic Procedure 15
2.1.4 Refinement 17
2.2 Foundation for Active Depth Acquisition via Sensors 19
2.2.1 Degradation Model 19
2.2.2 Related Work 20
2.3 Local Methods and Filters 21
2.3.1 L2 Norm Optimization Filters 22
2.3.2 L1 Norm Optimization Filters 24
2.4 Global Optimization and Graphic Models 25
2.4.1 Relation to Bayesian inference 25
2.4.2 Bayesian Inference for Depth Estimation and Depth Enhancement 26
2.4.3 Discrete Optimization via Graph Cut 27
2.5 Learning-Based Depth SR 30
2.5.1 Sparse Coding 30
2.5.2 Convolutional Neural Network 34
2.6 Summary 35

Chapter 3 Fast Depth Video Construction and Its Enhancement via Temporal Consistency 36
3.1 Related Work and Motivation 36
3.1.1 Related Work 37
3.1.2 Motivation 37
3.2 Fast Depth Estimation 38
3.2.1 Affine Invariant Feature 38
3.2.2 Initial Depth Estimation 40
3.2.3 Depth Map Refinement 46
3.2.4 Experimental Results 47
3.2.5 Conclusion 52
3.3 Temporal Consistency Enhancement for Multi-view Depth Sequences 53
3.3.1 Motivation 53

3.3.2 Proposed Method 54
3.3.3 Experimental Results 57
3.3.4 Conclusion 59
3.4 Summary 60

Chapter 4 Color-Guided Depth Map Enhancement for RGB-D Data via Markov Random Field 61
4.1 Related Work 61
4.2 Challenges in Guided Depth Enhancement 63
4.3 Color-Guided Depth SR via MRF Embedded with Hard-decision Edge Inconsistency Measurement 64
4.3.1 Energy Function Construction for MRF 65
4.3.2 Experimental Results 68
4.3.3 Conclusion 69
4.4 Color-Guided Depth Enhancement via MRF Embedded with Soft-decision Edge Inconsistency Measurement 69
4.4.1 Modifications in MRF Energy Function 70
4.4.2 Soft-decision Edge Inconsistency Measurement 70
4.4.3 Embedding Edge Inconsistency Measurement into MRF 78
4.4.4 Algorithm Complexity Discussion 80
4.4.5 Experimental Results 81
4.4.6 Conclusion 92
4.5 Minimum Spanning Forest Embedded with Soft Edge Inconsistency Measurement for Guided Depth Map Enhancement . 94
4.5.1 Modification on MRF Energy Function 94
4.5.2 The Proposed Method 95
4.5.3 Experimental Results 102
4.5.4 Conclusion 112
4.6 Summary 113

Chapter 5 Guided Depth Map Super-resolution via Deep Learning 115

5.1 Introduction and Motivation 115
5.2 MFR-SR Construction 118
5.3 Experiments 120
5.3.1 Training Data 120
5.3.2 Training Detail 121
5.3.3 Experimental Results 121
5.4 Summary 125

Chapter 6 Conclusions and Future Work 126
6.1 Conclusions 126
6.2 Future Work 128

Bibliography 129

List of Figures

1.1 An illustration of occlusion and textureless regions 2
1.2 Some typical products for each type of sensor 3
1.3 An illustration of depth maps captured by sensors 4
1.4 An illustration of edge inconsistency 5
1.5 The relation among Chapter 3, 4 and 5 10

2.1 An illustration of image rectification 13
2.2 The relation between disparity and depth 14
2.3 An illustration of graph cut with three cut solutions 29

3.1 Illustration of AIF 39
3.2 The flowchart of initial depth estimation 41
3.3 The flowchart of refinement for initial depth 46
3.4 Objective evaluate for initial estimated results 49
3.5 Subjective evaluation for initial estimated results 50
3.6 Objective evaluate for refinement 52
3.7 Framework of temporal consistency enhancement 54
3.8 Static regions of test sequences for evaluation 57
3.9 Subjective evaluation of temporal consistency 60

4.1 An illustration on the advantage of Weighted Bipartite Matching 73
4.2 An illustration of Minimum Weighted Bipartite Matching problem 75

4.3 The bi-direction edge inconsistency measurement for Middlebury dataset "Art . . . 77
4.4 The visual quality comparison for depth map SR on "Dolls" dataset using Pro-Soft method . . . 86
4.5 The visual quality comparison for depth map SR with noise on "Art" and "Moebius" datasets using Pro-Soft method . . . 87
4.6 The visual quality comparison for depth map completion on "Moebius" and "Laundry" datasets using Pro-Soft method . . 89
4.7 The visual quality comparison for depth map SR on "Devil and "Shark datasets using Pro-Soft method . . . 91
4.8 The visual quality comparison for depth map completion on NYU datasets using Pro-Soft method . . . 92
4.9 The visual quality comparison for depth map enhancement with the complex degradation on NYU datasets using Pro-Soft method . . . 93
4.10 Visual comparison of upsampled depth maps using different combination of methods . . . 101
4.11 Quantitative comparison between the results enhanced by different combination of methods . . . 102
4.12 The influence of the amount of the super-pixels in Pro-MSF method . . . 103
4.13 Visual comparison of upsampled depth maps on noise-free Middlebury datasets using Pro-MSF method . . . 107
4.14 Visual comparison of upsampled depth maps on noisy Middlebury datasets using Pro-MSF method . . . 110
4.15 Visual comparison of upsampled depth maps on ToF-Mark datasets "Devil" and "Shark" using Pro-MSF method . . . 111
4.16 Visual comparison of enhanced depth maps on synthetic datasets using Pro-MSF method . . . 113
4.17 Visual comparison of enhanced depth maps on NYU datasets using Pro-MSF method . . . 114

5.1 The architecture of MFR-SR for the case 4× 117
5.2 The visual quality comparison for depth map SR with noise on "Art" and "Reindeer" datasets using MFR-SR 124

List of Tables

2.1 The Edge Weights in Fig. 2.3 30

3.1 Quantitative Evaluation of Initial Estimated Results on Average PSNR of Rendered Color Images 48
3.2 Quantitative Evaluation of Depth Refinement on Average PSNR of Rendered Color Images 51
3.3 Average Running Time Comparison 53
3.4 Consistency Evaluation for Static Regions 58
3.5 Comparison of Average PSNR of Rendered Color Images . . . 58
3.6 Comparison of Depth Coding Performance 59

4.1 λ_s^{pq} Value Table for Pro-Hard Method 67
4.2 Depth SR Result of Pro-Hard Method on Noise-free Middlebury Datasets “Art”, “Book” and “Moebius” 68
4.3 Depth SR Results of Pro-Hard Method on Noise-free Middlebury Datasets “Reindeer”, “Laundry” and “Dolls” 69
4.4 Average Running Time of Pro-Soft Method (16×) 81
4.5 Depth SR Results of Pro-Soft Method on Noise-free Middlebury Datasets “Art”, “Book” and “Moebius” 84
4.6 Depth SR Results of Pro-Soft Method on Noise-free Middlebury Datasets “Reindeer”, “Laundry” and “Dolls” 85
4.7 Depth SR Results of Pro-Soft Method on Noisy Datasets “Art”, “Book” and “Moebius” 87

4.8 Depth SR Results of Pro-Soft Method on Noisy Datasets "Reindeer", "Laundry" and "Dolls" 88
4.9 Depth Completion Result of Pro-Soft Method on Synthetic Datasets 88
4.10 Depth SR Result of Pro-Soft Method on ToF-Mark Datasets . 90
4.11 Depth SR Result of Pro-MSF Method on Noise-free Middlebury Datasets "Art", "Book" and "Moebius" 105
4.12 Depth SR Result of Pro-MSF Method on Noise-free Middlebury Datasets "Reindeer", "Laundry" and "Dolls" 106
4.13 Depth SR Result of Pro-MSF Method on Noisy Middlebury Datasets "Art", "Book" and "Moebius" 108
4.14 Depth SR Result of Pro-MSF Method on Noisy Middlebury Datasets "Reindeer", "Laundry" and "Dolls" 109
4.15 Depth SR Result of Pro-MSF Method on ToF-Mark Datasets 109
4.16 Depth Completion Result of Pro-MSF Method on Synthetic Datasets 112
4.17 Running Time Comparison using Pro-MSF Method 112

5.1 Depth SR Result of MFR-SR on Noisy Middlebury Datasets "Art", "Book" and "Moebius" 122
5.2 Depth SR Result of MFR-SR on Noisy Middlebury Datasets "Reindeer", "Laundry" and "Dolls" 123
5.3 Depth SR Result of MFR-SR on Noisy Middlebury Datasets "Cones", "Teddy", "Tsukuba" and "Venus" 124

List of Publications

Papers Published (The First Author)

- **Yifan Zuo**, Qiang Wu, Jian Zhang, Ping An (2018), Minimum Spanning Forest with Embedded Edge Inconsistency Measurement for Guided Depth Map Enhancement. *IEEE trans. IP*, published online.

- **Yifan Zuo**, Qiang Wu, Jian Zhang, Ping An (2018), Explicit Edge Inconsistency Evaluation Model for Color-guided Depth Map Enhancement. *IEEE trans. CSVT*, vol. 28, no. 2, pp. 439-453.

- **Yifan Zuo**, Qiang Wu, Jian Zhang, Ping An (2016), Explicit Modeling on Depth-Color Inconsistency for Color-Guided Depth Up-sampling. *in* 'Proceedings of the International conference on Multimedia and Expo (ICME16)', IEEE, pp. 1-6.

- **Yifan Zuo**, Qiang Wu, Jian Zhang, Ping An (2016), Explicit Measurement on Depth-Color Inconsistency for Depth Completion. *in* 'Proceedings of the International conference on image processing (ICIP16)', IEEE, pp. 4037-4041.

- **Yifan Zuo**, Ping An, Shuai Zheng, Zhaoyang Zhang (2015), Depth Upsampling Method via Markov Random Fields without Edge-Misaligned Artifacts, *in* 'Proceedings of the International conference on image processing (ICIP15)', IEEE, pp. 2324-2328.

- **Yifan Zuo**, Qiang Wu, Jian Zhang, Ping An (2017), Minimum Spanning Forest with Embedded Edge Inconsistency Measurement for Color-guided Depth Map Upsampling, *in* 'Proceedings of the International conference on Multimedia and Expo (ICME17)', IEEE, pp. 211-216.

- **Yifan Zuo**, Ping An, Liquan Shen, Chunhua Li, Ran Ma (2016), Integration of Color and Affine Invariant Feature for Multi-view Depth Video Estimation. *Imaging Science Journal*, vol. 64, no. 6, pp. 313-320.

- **Yifan Zuo**, Ping An, Ran Ma, Liquan Shen, Zhaoyang Zhang (2014), Temporal Consistency Enhancement on Depth Sequences. *Journal of Optoelectronics Laser*, vol.25, no.1, pp.172-177.

- **Yifan Zuo**, Ping An, Qiu-Wen Zhang, and Zhao-Yang Zhang (2012), Fast Segment-Based Algorithm for Multi-view Depth Map Generation. *in* 'Proceedings of the International Conference on Intelligent Computing (ICIC12)', pp. 553-560.

Papers to be Submitted/Under Review

- **Yifan Zuo**, Qiang Wu, Ping An, Xiwu Shang, Integrate co-sparse analysis model with explicit edge inconsistency measurement for guided depth map upsampling. *Journal of Electronic Imaging*.

Abstract

With the developing of computer vision, the problem of high-quality depth map acquisition is demanding urgent solution. Generally, the methods for dense depth map acquisition consist of two categories: passive and active.

The passive methods based on stereo matching algorithms always compute matching cost volume pixel by pixel, which is time-consuming. This thesis firstly proposes a local depth estimation method using adaptive matching scheme. Furthermore, with the help of affine invariant feature, the performance for matching in textureless regions is improved. Experimental results show that the proposed method can achieve better or comparable performances than the state-of-the-art method in the category of local methods, even with the less running time. In addition, since the depth map is estimated frame by frame, the temporal consistency cannot be guaranteed. This thesis proposes a method to enhance temporal consistency by applying adaptive temporal filtering, which explicitly considers the reliability of depth and the moving attribute of regions. Experiments demonstrate that the proposed algorithm can generate more stable depth sequences and effectively suppress the transient depth errors when rendering virtual images.

Due to the inherent drawbacks of stereo matching, the depth map captured by sensors is more robust, especially for the textureless regions. However, it either suffers from low resolution, or has some holes on the depth map. Active methods are to solve these problems. Since low-quality depth map is always captured with a high-quality color or intensity image and they can be registered with each other on the same coordinate system, low-quality depth map can be refined by using the guidance from such high-quality color/intensity image. This type of active method is called guided depth map enhancement. In consideration of clear expression, this thesis uses color image to stand for color/intensity image in the rest of thesis. The meaning of it is according to the context.

The methods on guided depth map enhancement can be classified into different categories depending on whether external training data is used. Without relying on the external datasets, co-occurrence property between edges on the depth map and the corresponding color image is explicitly exploited. However, because the assumption above is not always true, it leads to texture-copy artifacts and blurring depth edges. Markov-Random-Field-based (MRF-based) methods are popular in guided depth map enhancement. The state-of-the-art solutions are to adjust the affinities of the regularization term in MRF energy function. Actually, these existing methods are lack of explicit evaluation model to quantitatively measure the inconsistency between the depth edges and the corresponding color edges, so they cannot adaptively control the efforts of the guidance from the color image for depth enhancement. In addition, widely used affinity computing scheme for regularization term is based on the depth and color differences between neighbor pixels, which ignores local structure on the depth map. In this thesis, three algorithms are proposed to address the problems above. The first one aims to mitigate artifacts caused by edge misalignment between the depth map and the color image via hard-decision inconsistency checking pixel by pixel. The second one uses a structural quantitative measurement on edges inconsistency which is a soft-decision method. It is more accurate than its hard-decision counterpart above. The third one is to combine such soft-decision edge inconsistency measurement and local structure of the depth map which is modeled on Minimum Spanning Trees (Forest) to acquire more robust depth map. These methods are tested on Middlebury, ToF-Mark and NYU datasets which prove progressive improvements.

In addition to the handcraft models for depth map enhancement, datadriven models are expected to implicitly learning such guidance to obtain superior performances. In this thesis, an end-to-end training method based on convolutional neural network is proposed, which borrows many concepts from existing models, e.g., batch-normalization and residual learning. It upsamples low-resolution depth map progressively and the residual network is constructed to learn high frequency component in multiple scales. This coarse-to-fine scheme can reconstruct high-resolution depth via multi-frequency synthesis. Experimental results show improvement in subjective evaluation and objective evaluation compared with state-of-the-art methods.