

Faculty of Engineering and Information Technology
University of Technology Sydney

Recovering Dense 3D Motion and Shape Information from RGB-D Data

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

by

Yucheng Wang

April 2017

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 part of the collaborative doctoral degree and/or 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 a research candidature conducted jointly with Beijing Institute of Technology as part of a collaborative Doctoral degree.

Signature of Candidate

*To my parents
for their love and support*

Acknowledgments

I would like to express my sincere gratitude to all those who have given me help in the period of pursuing my PhD.

Foremost, I would like to express my sincere gratitude to my supervisors Associate Prof. Jian Zhang for his patience, motivation, enthusiasm, immense knowledge, and continuous support of my study and research. His trust, understanding, and encouragement in my research have always motivated me. His rigorous academic attitude deeply influences with me in the future work and life.

I also would like to appreciate my co-supervisor Associate Prof. Qiang Wu for providing me with continuous support throughout my PhD study and research. Without his professional guidance and persistent help, this thesis would not have been possible.

I am very grateful to Senior Researchers Zicheng Liu, Zhengyou Zhang and Philip A. Chou at Microsoft Research, and I have benefited greatly from every high-quality group meeting. Their cutting-edge research perspectives and solid theoretical foundation make me open-minded and inspire me to work harder.

Finally, thanks to my parents, who have unconditionally supported my research work. Whenever thinking of them, I can face the upcoming new stage of life without fear and full of hope!

Yucheng Wang
April 2017 @ UTS

Contents

Certificate	i
Acknowledgment	v
List of Figures	xi
List of Tables	xiii
List of Publications	xv
Abstract	xvii
Chapter 1 Introduction	1
1.1 Background	1
1.2 Research Issues	4
1.3 Research Contributions	6
1.4 Thesis Structure	7
Chapter 2 Literature Review and Foundation	10
2.1 Color Data-based Methods	11
2.2 Depth Data-based Methods	14
2.2.1 Image-based Methods	15
2.2.2 Video-based Methods	17
2.3 RGB-D Data-based Methods	18
2.3.1 Image-based Methods	18
2.3.2 Video-based Methods	20
Chapter 3 Fusing Depth Imaging and Stereo Vision with Structural Determinant Information Inference	23
3.1 Introduction	23

3.2	Problem Formulation	24
3.2.1	Initialization	25
3.2.2	Inference of SDI	27
3.2.3	Image Observation	29
3.2.4	Depth Observation	30
3.2.5	Fusion of Image and Depth Observation with SDI . . .	30
3.2.6	Iterative Refinement	31
3.3	Experimental Results	31
3.3.1	Evaluation of Matching Cost Measurement	32
3.3.2	Evaluation of Fusion	33
3.3.3	Evaluation of Iterative Refinement	37
3.4	Conclusion	37

Chapter 4	Handling Occlusion and Large Displacement Motion through Improved RGB-D Scene Flow Estimation	40
4.1	Introduction	40
4.2	Related Work	43
4.3	Problem Formulation	47
4.3.1	Notation	48
4.3.2	Feature Consistency in Non-Occluded Regions along the Temporal Domain	49
4.3.3	Spatial Smoothness Regularization	53
4.4	Optimization	54
4.4.1	Proposal Generation	55
4.4.2	Proposal Fusion	61
4.5	Experimental Results	61
4.5.1	Quantitative Evaluation on Accuracy and Efficiency . .	62
4.5.2	Qualitative Evaluation on Robustness for Challenging Large Displacement Motion and Occlusion	71
4.5.3	Qualitative Evaluation on Robustness for Challenging Deformable Surfaces	73

4.5.4	Qualitative and Quantitative Evaluation on Robustness and Accuracy for Long RGB-D Video Sequences	74
4.6	Conclusions	79
Chapter 5 Depth Super-Resolution on RGB-D Video Sequences with Large Displacement 3D Motion 81		
5.1	Introduction	81
5.2	Related Work	83
5.2.1	Motion Compensation of Neighboring Depth Images	83
5.2.2	Fusion of Compensated Depth Images	86
5.3	Framework	87
5.4	Motion Compensation	88
5.4.1	Formulation	88
5.4.2	Optimization	91
5.5	Fusion	95
5.5.1	Formulation	95
5.5.2	Training	99
5.6	Experiments	100
5.6.1	Evaluation of Depth Super-Resolution	101
5.6.2	Evaluation of Motion Compensation	107
5.6.3	Evaluation of Fusion	114
5.7	Conclusion and Discussion	115
Chapter 6 Conclusions and Future Work 117		
6.1	Conclusions	117
6.2	Future Work	119
Bibliography 121		

List of Figures

1.1	The RGB-D data captured by consumer-level sensor Kinect . . .	4
1.2	The profile of work in this thesis	9
3.1	The structure of the Bayesian belief networks	26
3.2	The relationship of occlusion between pixels.	27
3.3	The errors of different matching cost measurement in occluded areas for the scenes Aleo and Moebius	32
3.4	The results of the scenes Arts, Laundry, and Cloth4 generated	34
3.5	The errors in non-occluded and all areas of results	35
3.6	The results of the scenes Aleo and Moebius	36
3.7	The errors in non-occluded and all areas of disparity map for Art, Laundry, and Cloth4 after the iterative refinement.	38
4.1	The notation of the 2D pixels and 3D points.	49
4.2	The optimization methods for original 2D PatchMatch and our RGB-D PatchMatch. The modules in blue are proposed in the original PatchMatch. The modules in green are proposed by ours.	56
4.3	The NRDC and over-parameterized scene flow representation .	59
4.4	The selection of parameter λ	63
4.5	The 2D projection of 3D scene flow on the image plane	65
4.6	The performance of different PatchMatch algorithms on Middlebury Venus, Cones and Teddy	66
4.7	RGB-D image pairs Nici, Stefan and Largehand	69

LIST OF FIGURES

4.8	The scene flow results on RGB-D image pairs Nici, Stefan and Largehand	70
4.9	The scene flow results of two frames in the Tshirt4 sequence	72
4.10	The tracking results on Tshirt4 sequence and Human Hand Waving sequence	75
4.11	The accumulated error of landmark points on Tshirt4 sequence and Human Hand Waving sequence under large-displacement motion	76
5.1	The framework of our video-based depth super-resolution method	85
5.2	The representation of 3D over-parametrized motion for super-pixels	92
5.3	The graph of super-pixels and its forward propagation order	94
5.4	The CNN architecture of "DepthFusionNet" for fusion	97
5.5	The qualitative comparison of the results of the scaling factor 4 on the NewTsukuba dataset	104
5.6	The qualitative comparison of the results under the super-resolution scale factor 4 on the Sintel dataset	105
5.7	The qualitative results of different methods under the super-resolution scale factor 2 on two sequences <i>Tidying</i> and <i>Reading</i>	108
5.8	The MSE of different motion estimation methods on the entire sequences	109
5.9	The MSE of different motion estimation methods on frame 48 of video NewTsukuba	111
5.10	The MSE of different motion estimation methods on frame 868 of video NewTsukuba	112
5.11	The MSE of different methods under the super-resolution scale factor 4	113

List of Tables

4.1	The classification of RGB-D scene flow methods	44
4.2	The evaluated errors under different evaluation criteria and average running time	64
4.3	The tracking error under regular and large displacement motion in sequences Tshirt4 and Human Hand Waving	77
5.1	Three problems caused by large displacement 3D motion in motion estimation methods	84
5.2	The MSE of the entire sequence under different super-resolution scale factors	103
5.3	The MSE of the entire sequences under the super-resolution scale factor 4 of different fusion methods	115

List of Publications

Papers Published

- **Yucheng Wang**, Jian Zhang, Zicheng Liu, Qiang Wu, Philip Chou, Zhengyou Zhang, Yunde Jia (2015), Handling Occlusion and Large Displacement through Improved RGB-D Scene Flow Estimation. *IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT)*.
- **Yucheng Wang**, Jian Zhang, Zicheng Liu, Qiang Wu, Philip Chou, Zhengyou Zhang, Yunde Jia (2014), Completed Dense Scene Flow in RGB-D Space. In: *Asian Conference on Computer Vision Workshop*.
- **Yucheng Wang**, Huijun Di, Bingjie Wang, Wei Liang, Jian Zhang, Yunde Jia (2014), Depth Super-resolution by Fusing Depth Imaging and Stereo Vision with Structural Determinant Information Inference. in *IEEE International Conference on Pattern Recognition*.
- Yong Duan, Mingtao Pei, **Yucheng Wang**, Min Yang, Xiameng Qin, Yunde Jia (2015), A Unified Probabilistic Framework for Real-Time Depth Map Fusion. *Journal of Information Science and Engineering*.
- Yan Liang, Wanxuan Lu, Wei Liang, **Yucheng Wang** (2014), Action Recognition Using Local Joints Structure and Histograms of 3D Joints. in *International Conference on Computational Intelligence and Security*.

LIST OF PUBLICATIONS

- Bingjie Wang, Wei Liang, **Yucheng Wang**, Yan Liang (2013), Head Pose Estimation with Combined 2D SIFT and 3D HOG Features. in *International Conference on Image and Graphics*.

Papers to be Submitted/Under Review

- **Yucheng Wang**, Jian Zhang, Zicheng Liu, Qiang Wu, Zhengyou Zhang, Yunde Jia (2016), Depth Super-Resolution on RGB-D Video Sequences with Large Displacement 3D Motion, *IEEE Transactions on Image Processing (T-IP)*.

Research Reports of Industry Projects

- **Yucheng Wang**. Advancing 3D deformable surface reconstruction and tracking through RGB-D cameras, Microsoft Research funded project, 2012-2014.

Abstract

3D motion and 3D shape information are essential to many research fields, such as computer vision, computer graphics, and augmented reality. Thus, 3D motion estimation and 3D shape recovery are two important topics in these research communities. RGB-D cameras have become more accessible in recent few years. They are popular for good mobility, low cost, and high frame rate. However, these RGB-D cameras generate low-resolution and low-accuracy depth images due to chip size limitations and ambient illumination perturbation. Thus, obtaining high-resolution and high-accuracy 3D information based on RGB-D data is an important task.

This research investigates 3D motion estimation and 3D shape recovery solutions for RGB-D cameras. Thus, within this thesis, various methods are developed and presented to address the following research challenges: fusing passive stereo vision and active depth acquisition; 3D motion estimation based on RGB-D data; depth super-resolution based on RGB-D video with large displacement 3D motion.

In Chapter 3, a framework is presented to acquire depth images by fusing active depth acquisition and passive stereo vision. Active depth acquisition and passive stereo vision have their limitations in some aspects, but their range-sensing characteristics are complementary. Thus, combining both approaches can produce more accurate results than using either one only. Unlike previous fusion methods, the noisy depth observation from active depth acquisition is initially taken as a prior knowledge of the scene structure, which improves the accuracy of the fused depth images.

Chapter 4 details a method for 3D scene flow estimation based on RGB-D data. The accuracy of scene flow estimation is limited by two issues: occlusions and large displacement motions. To handle occlusions, the occlusion status is modelled, and the scene flow and occluded regions are jointly estimated. To deal with large displacement motions, an over-parameterised scene flow representation is employed to model both the rotation and translation components of the scene flow.

In Chapter 5, a depth super-resolution framework is presented for RGB-D video sequences with large 3D motion. To handle large 3D motion, our framework has two stages: motion compensation and fusion. A superpixel-based motion estimation approach is proposed for efficient motion compensation. The fusion task is modelled as a regression problem, and a specific deep convolutional neural network (CNN) is designed that can learn the mapping function between depth image observations and the fused depth image given a large amount of training data.