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

List of Figures

List of Tables

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.*

- 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 learns the mapping function between depth image observations and the fused depth image given a large amount of training data.