

ABSTRACT OF THESIS

RGB-D Cameras' Per-Pixel Calibration and Natural 3D Reconstruction on GPU

Ever since the Kinect brought low-cost depth cameras into consumer market, great interest has been invigorated into Red-Green-Blue-Depth (RGB-D) sensors. Without calibration, a RGB-D camera's horizontal and vertical field of view (FoV) could help generate 3D reconstruction in camera space naturally on graphics processing unit (GPU), which however is badly deformed by the lens distortions and imperfect depth resolution (*depth distortion*). The camera's calibration based on a pinhole-camera model and a high-order distortion removal model requires a lot of calculations in the fragment shader. In order to get rid of both the lens distortion and the *depth distortion* while still be able to do simple calculations in the GPU fragment shader, a novel per-pixel calibration method with look-up table based 3D reconstruction in real-time is proposed, using a rail calibration system. This rail calibration system offers possibilities of collecting infinite calibrating points of dense distributions that can cover all pixels in a sensor, such that not only lens distortions, but *depth distortion* can also be handled by a per-pixel D to Z^W mapping. Instead of utilizing the traditional pinhole camera model, two polynomial mapping models are employed. One is a two-dimensional high-order polynomial mapping from R/C to X^W/Y^W respectively, which handles lens distortions; and the other one is a per-pixel linear mapping from D to Z^W , which can handle *depth distortion*. With only six parameters and three linear equations in the fragment shader, the undistorted 3D world coordinates (X^W, Y^W, Z^W) for every single pixel could be generated in real-time. The per-pixel calibration method could be applied universally on any RGB-D cameras. With the alignment of RGB values using a pinhole camera matrix, it could even work on a combination of a random Depth sensor and a random RGB sensor.

KEYWORDS: 3D Reconstruction; Camera Calibration; RGBD; Image Processing

Author's signature: _____ Sen Li

Date: _____ September 8, 2016

RGB-D Cameras' Per-Pixel Calibration and Natural 3D Reconstruction on GPU

By
Sen Li

Director of Dissertation: Dr. Daniel L. Lau

Director of Graduate Studies: Dr. Caicheng Lu

Date: September 8, 2016

RULES FOR THE USE OF THESES

Unpublished theses submitted for the Master's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the thesis in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this thesis for use by its patrons is expected to secure the signature of each user.

Name

Date

THESIS

Sen Li

The Graduate School
University of Kentucky
2016

RGB-D Cameras' Per-Pixel Calibration and Natural 3D Reconstruction on GPU

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Engineering
at the University of Kentucky

By
Sen Li
Lexington, Kentucky

Director: Dr. Daniel L. Lau, Professor of Electrical Engineering
Lexington, Kentucky

2016

Copyright© Sen Li 2016

TABLE OF CONTENTS

Table of Contents	iii
List of Figures	iv
Chapter 1 Conclusion and Future Work	1
1.1 Conclusion	1
1.2 Future Work	3
Bibliography	5

LIST OF FIGURES

Chapter 1 Conclusion and Future Work

The depth sensor technologies opens a new epoch for 3D markets. Ever since Microsoft brought the low-cost depth camera Kinect into consumer market, RGB-D cameras are famous for their 3D reconstruction applications in research areas of HCI. Graphic processing units (GPUs) can solve large data parallel problems at a higher speed than the traditional CPU, while being more affordable and energy efficient [1]. With floating point calculation and matrix operation, GPUs are specialized many-core processors that are optimized for graphic processing [2]. Without calibration, the horizontal and vertical FoV could help generate 3D reconstruction in camera space naturally on GPU through proportional calculations in fragment shader, which nevertheless is badly deformed by the lens distortions and *depth distortion*. Due to image formation, some methods must be designed to establish the correct correspondence between raw images streams from the camera and the real-time 3D reconstruction feedback. Traditionally, the camera calibration is done based on a pinhole-camera model and a high-order distortion removal model. However, there will be a lot of calculations in the fragment shader to process parameters. There will be at least five intrinsic parameters from intrinsic camera matrix and another five parameters from distortion removal vector, while the distortion correction in this method requires multiplication calculations to achieve the high order polynomial. In order to get rid of both the lens distortion and the *depth distortion* while still be able to do simple calculations on the GPU fragment shader, a new solution is needed.

1.1 Conclusion

In this thesis, a novel per-pixel calibration method with look-up table based 3D reconstruction on GPU in real-time is introduced, using a rail calibration system. This rail calibration system offers possibilities of collecting infinite calibration points with dense distributions

that can cover all pixels in a sensor. Not only lens distortions, but *depth distortion* can also be handled by a per-pixel D to Z^W mapping with data along Z^W -axis collected on the rail. Instead of utilizing the traditional pinhole camera model, two polynomial mapping models are employed in this calibration method. The first model is the two-dimensional 4th order polynomial mapping from R/C to X^W/Y^W during the frames data collection, which takes care of the removal of lens distortions; and the second model is the linear mapping from D to Z^W , which can handle “depth distortion”. The method consists of two big steps: $X^W Y^W Z^W + D$ data collection and mapping parameters determination. D is simply from depth streams. Z^W is from external based on the camera’s position on the rail. And the undistorted (X^W, Y^W) are from the transformation of R/C by a 4th order polynomial mapping model, during which lens-distortions could be removed. This method is claimed as “data-based” calibration method, because both of the two mapping models are determined and calculated by real streams data from the camera.

With the fewest calculations, the undistorted 3D world coordinates (X^W, Y^W, Z^W) for every single pixel could be looked up in real-time based on D from a *column-by-row-by-6* look-up table. Only three linear equations with six parameters need to be calculated in the fragment shader. Two out of six parameters a/b are to determine the per-pixel Z^W , which is generated from per-pixel depth value D ; and the other four parameters $c/d/d/f$ are to determine the per-pixel X^W/Y^W respectively, which are mapped by per-pixel linear beam equation from the per-pixel Z^W . Note that “real-time” here means being able to show an undistorted frame in 3D world space before the start of the second frame processing, and this LUT based 3D reconstruction method can realize real-time very well. The data-based per-pixel calibration method could be applied universally on any RGB-D cameras. With the alignment of RGB pixels using a pinhole camera matrix, it could even work on a combination of a random Depth sensor and a random RGB sensor.

1.2 Future Work

The rail calibration system and per-pixel calibration method with LUT-based real-time 3D reconstruction could be applied universally to all kinds of RGB-D cameras. With a more precise calibration system and corresponding DIP technologies, there could be a huge improvement space for calibration accuracy. Hardware improvement is sometimes more important than a software updating. Concretely the system in the lab now can only handle a working distance Z^W from 1.165m to 2.565m. Considering that the depth resolution deteriorates notably with depth, it might not be a simple linear relationship from D and Z^W , when the depth D value goes much further than the limit of the rail. In that case, the per-pixel D to Z^W mapping could be changed from singular linear to segmented mapping function when D gets larger than a certain value.

Not only the rail, the planar pattern could also be changed based on the resolution of the camera (*e.g.*, size of the dots and side distance of the unit square-shaped of the uniform grid could be larger when the camera has a supper high resolution). A two dimensional object is totally enough for calibrating KinectV2, because the NearIR stream have same pixels' distortions with the depth stream. However, if NearIR stream could not be used to extract points R/C in image space (*e.g.* a structured light based depth sensor) while the size (pixel numbers) of the RGB sensor does not share the same one with the depth sensor, we could make the planar pattern into a “three dimension mode”: change the printed dots into real wholes such that the depth stream could detect a difference between the “dots” and “white background”.

Instead of using a laser distance measurer to manually input Z^W into every frame at the very beginning of the frame data collection, we could add a tracking module onto the rail to active the frame collection by software. In this way, not only Z^W could be traced by the tracking module, but also the frame collection could be automatically done by the activation from the tracking module, recording one frame data after every certain distance of movement has been detected.

Besides the hardware enhancement, softwares like DIP process can also be improved. The order of polynomial model that help map from image space R/C to world space X^W/Y^W could also be heightened for a better accuracy, as long as there will be enough calibrating points to offer the coordinates pairs of both image space and world space. Considering the possible lens-distortions of the RGB sensor, the high order polynomial mapping could also be applied into the color values' alignment from RGB sensor pixels to depth sensor pixels.

Bibliography

- [1] Hui Liang Khor, Siau-Chuin Liew, and Jasni Mohd Zain. A review on parallel medical image processing on GPU. In *International Conference on Software Engineering and Computer Systems (ICSECS)*, Aug. 2015.
- [2] J. E. Gomez-Balderas and D. Houzet. A 3D reconstruction from real-time stereoscopic images using GPU. In *Design and Architectures for Signal and Image Processing (DASIP)*, Oct. 2013.