TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1   RGBD Cameras

A RGB-D camera, also called depth camera, is a sensing system that capture RGB images along with per-pixel depth information. Usually it is simply a combination of a RGB sensor a depth sensor with certain alignment algorithm between them. Ever since the Kinect brought low-cost depth cameras into consumer market, with PrimeSense 3D sensing technology as the core depth determination principle for its first generation, great interest has been invigorated into RGB-D sensors. As the depth sensor technologies opens a new epoch for three-dimensional (3D) markets, the multimedia technologies development is changing from two-dimensional (2D) to 3D on games, movies, health care aid, military training, *etc.*, all various areas that are in demand of a more direct reflection on the real world [3]. Simple 2D scenes are not able to meet the social demands any more. And with the fast development of multimedia technologies, three-dimensional scene display has become a hotspot in the display field. As an extension of classic 2D video, the 3D dynamic display technology can provide a more comprehensive immersive feeling to users than a 2D video. Gesture recognition, 3D modeling, 3D printing, augmented reality, virtual reality, *etc.*, a lot of ongoing researches and applications on depth cameras are famous now, cooperated with Human Computer Interaction (HCI) technologies.

The PrimeSense's technology had been originally applied to gaming, whose user interfaces are based on gesture recognition instead of using a controller (also called Natural User Interface, NUI). It was best known for licensing the hardware design and chip used in Microsoft's first generation of Kinect motion-sensing system for the Xbox 360 in 2010. The PrimeSense sensor projects an infrared speckle pattern, which will then be captured by an infrared camera in the sensor. A special microchip is employed to compare the captured speckle pattern part-by-part to reference patterns stored in the device, which were captured previously at known depths. The final per-pixel depth will be estimated based on which reference patterns the captured pattern matches best. Other than the first generation of Kinect camera, Asus Xtion PRO sensor, another consumer NUI application product, has also applied the PrimeSense's technology.

As a competitor of PrimeSense Structured Light technology, time-of-flight technology had been applied into PMD[Vision] CamCube cameras and 3DV's ZCam cameras. Based on known speed of light, Time-of-Flight (ToF) camera resolves distance by measuring the "time cost" of a special light signal traveling between the camera and target for every sin-

gle point. The "time cost" variable that ToF camera measures is the phase shift between the illumination and reflection, which will be translated to distance [19]. To detect the phase shifts, light source is pulsed or modulated by a continuous wave, typically a sinusoid or square wave. The ToF camera illumination is typically from a LED or a solid-state laser operating in the near-infrared range invisible to human eyes. Fabrizio *et al*.[2] compared the time-of-flight (PMD[Vision] CamCube) camera and PrimeSense (first generation Kinect) camera in 2011. He showed that the time-of-flight technology is more accurate and claimed that the time-of-flight technology will not only be extended to support colours and higher frame sizes, but also rapidly drop in price. In 2009, 3DV agreed to sell its ZCam assets to Microsoft. And in 2013, Microsoft released the Xbox One, whose NUI sensor KinectV2 features a wide-angle ToF camera.

Unlike the PrimeSense's speckle pattern or KinectV2's ToF, Intel RealSense camera utilizes stereo vision. Its sensor actually has three cameras: two IR cameras (left and right), and one RGB camera. Additionally, RealSense camera also has an IR laser projector to help the stereo vision recognize depth at unstructured surfaces. Compared with KinectV2 camera, RealSense camera is more like a desktop usage to capture faces or even finger gestures, whereas the KinectV2 could do better to capture the full body actions with all joints. The effective distances of KinectV2 and RealSense hardwares are different. The KnectV2 is optimized to 0.5m 4.5m, while RealSense are designed for 0.2m 1.2m depends on different devices.

1.2 Human Computer Interface

Pattern recognition and gesture recognition are of the hottest sustained research activities in the area of HCI. Being a significant part in non verbal communication hand gestures are playing vital role in our daily life. Hand Gesture recognition system provides us an innovative, natural, user friendly way of interaction with the computer which is more familiar to the human beings. Gesture Recognition has a wide area of application including human machine interaction, sign language, immersive game technology *etc*. By keeping in mind the similarities of human hand shape with four fingers and one thumb, [23] presents a real time system for hand gesture recognition on the basis of detection of some meaningful shape based features like orientation, center of mass (centroid), status of fingers, thumb in terms of raised or folded fingers of hand and their respective location in image. Since gestures based on hand and finger movements can be robustly understood by computers by using a special 3D IR camera, users are allowed to play games and interact with computer applications in natural and immersive ways that improve the user experience. In 2010,
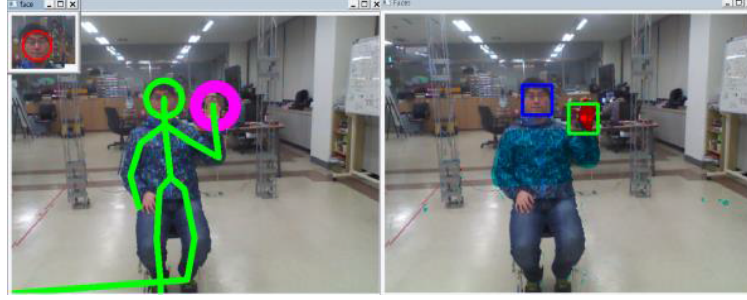
Figure 1.1: Calling Gesture Recognition Using Kinect [33]

Microsoft's first generation Kinect, using PrimeSenses technology, is based on projecting a structured light pattern on the object in order to build and track the skeleton of the users body to control a series of electronic games. [17] also discussed similar techniques for 3D object reconstruction. The appearance of RGBD cameras around revolutionized robotics applications, as the sensor is very easy to use and oers data of enough quality for most applications. By offering an alternative to time-cost procedures such as stereo-vision and laser scanning, RGB-D cameras open up the possibility of real-time robot interaction in human environments. Using this depth-map along with the 2D color images obtained via the Kinect camera, the computational unit of the Xbox 360 console was capable of providing a fairly robust gesture recognition solution. After [16] proposed two methods for gesture recognition, a plethora of applications have been produced around Kinect by various groups of image processing researchers.

A Kinect-based calling gesture recognition scenario is proposed by [33] for taking order service of an elderly care robot. Its proposed scenarios are designed mainly for helping non expert users like elderly to call service robot for their service request. In order to facilitate elderly service, natural calling gestures are designed to interact with the robot. Fig. 1.1 shows the evaluation of gesture recognition when sitting on chair. Individual people is segmented out from 3D point cloud acquired by Microsoft Kinect, skeleton is generated for each segment, face detection is applied to identify whether the segment is human or not, and specific natural calling gestures are designed based on skeleton joints. [12] proposed another smart and real-time depth camera based on a new depth generation principle. A monotonic increasing and decreasing function is used to control the frequency and duty-cycle of the NIR illumination pulses. The adjusted light pulses reflect off of the object of interest and are captured as a series of images. A reconfigurable hardware architecture calculates the depth-map of the visible face of the object in real-time from a number of images. The final depth map is then used for gesture detection, tracking and recognition. Fig. 1.2 shows an example extraction of hand skeleton. In 2013, Jaehong [18] *et al.* develop
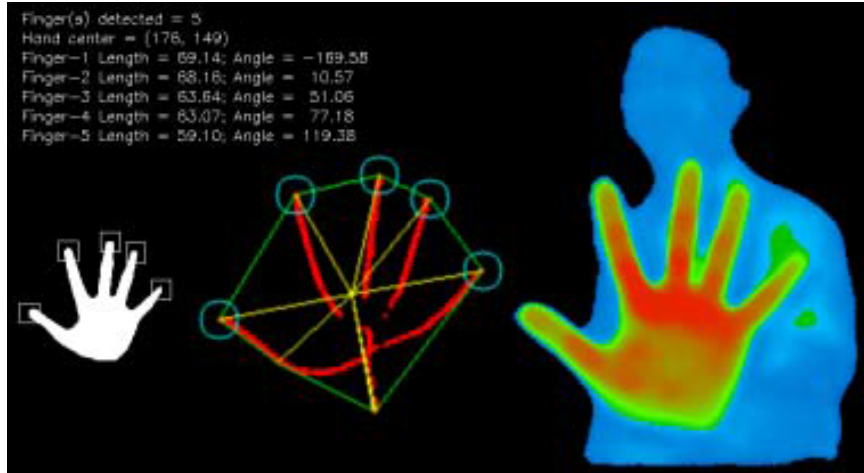
Figure 1.2: Finger Detection using Depth Data [12]

and implement a Kinect-based 3D gesture recognition system for interactive manipulation of 3D objects in educational visualization softwares.

RGB-D cameras own great credits in mobile robotics building dense 3D maps of indoor environments. Such maps have applications in robot navigation, manipulation, semantic mapping, and telepresence. [10] present a detailed RGB-D mapping system that utilizes a joint optimization algorithm combining visual features and shape-based alignment. Building on best practices in Simultaneous Localization And Mapping (SLAM) and computer graphics makes it possible to build and visualize accurate and extremely rich 3D maps with RGB-D cameras. Visual and depth information are also combined for view-based loop closure detection, followed by pose optimization to achieve globally consistent maps. SLAM is the process of generating a model of the environment around a robot or sensor, while simultaneously estimating the location of the robot or sensor relative to the environment. SLAM has been performed in many ways, which can be categorized generally by their focus on localization or environment mapping [29]. SLAM systems focused on localizing the sensor accurately, relative to the immediate environment, make use of sparse sensor data to locate the sensor. Using range sensors such as scanning laser range-finders [15], LiDAR and SONAR [6], many robot applications use SLAM systems only to compute the distance from the sensor to the environment. SLAM systems focused on mapping use dense sensor output to create a high-fidelity 3D map of the environment, while using those data to also compute relative location of the sensor [22] [31]. Many modern SLAM algorithms combine both approaches, usually by extracting sparse features from the sensor and using these for efficiently computing the location of the sensor. This position is then used to construct a map from dense sensor data.
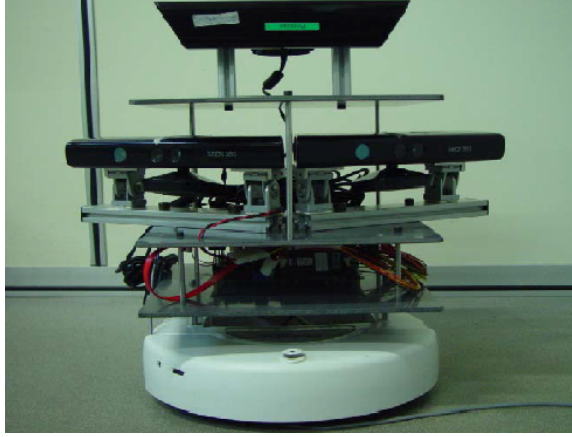
Figure 1.3: SLAM system with Only RGBD Cameras [20]

With a consumer RGB-D camera providing both color images and dense depth maps at full video frame rate, there appears a novel approach to SLAM that combines the scale information of 3D depth sensing with the strengths of visual features to create dense 3D environment representations, which is called RGB-D SLAM. [4] gives an open source approach to visual SLAM from RGB-D sensors, which extracts visual keypoints from the color images and uses the depth images to localize them in 3D. [20] builds an efficient SLAM system using three RGBD sensors. As shown in fig. 1.3, one Kinect looking up toward the ceiling can track the robots trajectory through visual odometry method, which provide more accurate motion estimation compared to wheel motion measurement without being disturbed under wheel slippage. And the other two contiguous horizontal Kinects can provide wide range scans, which ensure more robust scan matching in the RBPF-SLAM framework. Also using RGB-D sensor for SLAM, [21] presents a constraint bundle adjustment which allows to easily combine depth and visual data in cost function entirely expressed in pixel. In order to enhance the instantaneity of SLAM for indoor mobile robot, [30] proposed a RGBD SLAM method based on Kinect camera, which combined Oriented FAST and Rotated BRIEF (ORB) algorithm with Progressive Sample Consensus (PROSAC) algorithm to execute feature extracting and matching. ORB algorithm which has better property than many other feature descriptors was used for extracting feature. At the same time, ICP algorithm was adopted for coarse registration of the point clouds, and PROSAC algorithm which is superior than RANSAC in outlier removal was employed to eliminate incorrect matching. To make the result more accurate, pose-graph optimization was achieved based on General Graph Optimization (g2o) framework. Fig. 1.4 shows the 3D volumetric map of the lab, which can be directly used to navigate robots.

RGB-D camera is also famous in application of doing visual odometry on autonomous

Figure 1.4: 3D Map of RGBD-SLAM with ORB and PROSAC [30]

flight of a micro air vehicle (MAV), helping acquire 3D models of the environment and estimate the camera pose with respect to the environment model. Visual odometry generally has unbounded global drift while estimating local motion. To bound estimation error, it can be integrated with SLAM algorithms, which employ loop closing techniques to detect when a vehicle revisits a previous location. A computationally inexpensive RGBD-SLAM solution is discussed by [25] tailored to the application on autonomous MAVs, which enables our MAV to fly in an unknown environment and create a map of its surroundings completely autonomously, with all computations running on its onboard computer. Fig. 1.5a shows the MAC with an RGB-D sensor (the first generation of Kinect) mounted. And fig. 1.5b shows the reconstruction based on the full point clouds, with the estimated trajectory shown in red dots.

RGB-D sensors can be used on a much smaller scale than SLAM to create more detailed, volumetric reconstructions of objects and smaller environments, which opens a new world to the fast 3D printing. 3D printing is an additive technology in which 3D objects are created using layering techniques of different materials, such as plastic, metal, *etc*. It has been around for decades, but only recently is available and famous among the general public. The first 3D printing technology developed in the 1980s was stereolithography (SLA) [11]. This technique uses an ultraviolet (UV) curable polymer resin and an UV laser to build each layer one by one. Since then other 3D printing technologies have been introduced. Nowadays, some companies like iMaterialise or Shapeways offer 3D printing services where you can simply upload your CAD model on-line, choose a material and in a few weeks your 3D printed object will be delivered to your address. This procedure is quite straight-forward when you got your CAD model. However, 3D shape design tends to

6

<div align="center">(a) RGBD UAV   (b) Reconstruction and Estimated Trajectory</div>

Figure 1.5: RGBD-SLAM for Autonomous MAVs [25]

be a long and tedious process, with the design of a detailed 3D part usually requiring multiple revisions. Fabricating physical prototypes using low cost 3D fabrication technologies at intermediate stages of the design process is now a common practice, which helps the designer discover errors, and to incrementally rene the design. Most often, implementing the required changes directly in the computer model, within the 3D modeling software, is more difficult and time consuming than modifying the physical model directly using hand cutting, caving and sculpting tools, power tools, or machine tools. When one of the two models is modified, the changes need to be transferred to the other model, a process we refer to as synchronization.

KinectFusion, a framework that allows a user to create a detailed 3D reconstruction of an object or a small environment in real-time using Microsoft Kinect sensor, has garnered a lot of attention in the reconstruction and modeling field. It enables a user holding and moving a standard Kinect camera to rapidly create detailed 3D reconstructions of an indoor scene [13]. Not only an entire scene, a specific smaller physical object could also be cleanly segmented from the background model simply by moving the object directly. Fig. 1.6 shows how the interested object (a teapot) is accurately segmented from the background by physically removed. The sub-figure (A) shows surface normals, and sub-figure (B) is the texture mapped model. [8] proposed and introduced a from-Sense-to-Print system that can automatically generate ready-to-print 3D CAD models of objects or humans from 3D reconstructions using the low-cost Kinect sensor. Further, [9] addresses the problem of synchronizing the computer model to changes made in the physical model by 3D scanning the modified physical model, automatically detecting the changes, and updating
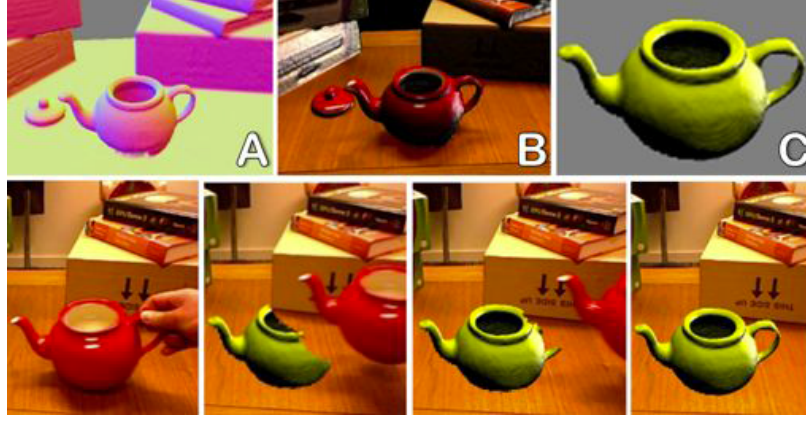
Figure 1.6: Object Segmentation in KinectFusion [13]

the computer model. A new method is proposed that allows the designer to move fluidly from the physical model (for example his 3D printed object, or his carved object) to the computer model. In the proposed process the physical modification applied by the designer to the physical model are detected by 3D scanning the physical model and comparing the scan to the computer model. Then the changes are reflected in the computer model. The designer can apply further changes either to the computer model or to the physical model. Changes made to the computer model can be synchronized to the physical model by 3D printing a new physical model.

1.3 Calibration of RGB-D Cameras

Camera calibration is a necessary step in 3D computer vision in order to extract metric information from 2D images. The calibration of a RGB-D camera aims to be able to generate the world coordinates (X^W, Y^W, Z^W) with corresponding *RGB* values for every single pixel, given the depth steams and RGB streams retrieved from the camera. During the camera calibration, the removal of lens distortions must be involved in order to get the undistorted 3D reconstruction, especially for consumer cameras that utilize a low-cost lens. Imperfect lens shape causes light rays bending more near the edges of a lens than they do at its optical center. The smaller the lens, the greater the distortion. With a wide-angle lens where the field of view of the lens is much wider than the size of the image sensor, barrel distortions will commonly happen. For example, fig. 1.7a shows KinectV2's raw (without distortion removal) NearIR 3D image observing a flat wall, on which a canvas printed with uniform grid dots pattern is hung. A blue rectangle is dropped onto the image, which helps show the "barrel-shape" deformation of the uniform grid pattern.

For decades, much work on camera calibration has been done, starting from the pho-

togrammetry community [1] [5], to computer vision ([28] [7] [32] to cite a few). After the release of the low-cost Microsoft Kinect, a lot of researchers focus their calibrations specially on RGB-D cameras (*e.g.*, Kinect). Jan *et al*. [26] analyze Kinect as a 3D measuring device, experimentally investigate depth measurement resolution and error properties and make a quantitative comparison of Kinect accuracy with stereo reconstruction from SLR cameras and a 3D-TOF camera. Using small data sets of Kinect, Carolina *et al*. [24] present a new method for calibrating a color-depth camera pair, which prevents the calibration from suffering a drift in scale. Aaron *et al*. [27] presented a novel multi-Kinect calibration algorithm that only requires a user to move single spherical object in front of the Kinects observed from multiple viewpoints.

However, those camera calibration methods, using limited number of points to train the non-linear distortion-removal model, are not able to cover all pixels of a sensor. Besides, with one pinhole camera model (3-by-4 transformation matrix) for raw (deformed) 3D reconstruction and another model to remove lens distortion, the final 3D reconstruction cannot be displayed efficiently in real-time. What's worse, those calibration methods are assuming that the depth sensor offers perfectly same precision depth value for all pixels. Whereas in practical depth sensors always have some defects in getting same depth accuracy for all of their pixels, which we will call as "Depth Distortion". Kourosh and Sander [14] provide an analysis of the accuracy and resolution of Kinect's depth data. They show that the random error of depth measurement increases with increasing distance to the sensor, and ranges from a few millimeters up to about 4 cm at the maximum range of the sensor. Fig. 1.7b shows the side view of KinectV2's raw 3D NearIR image observing the
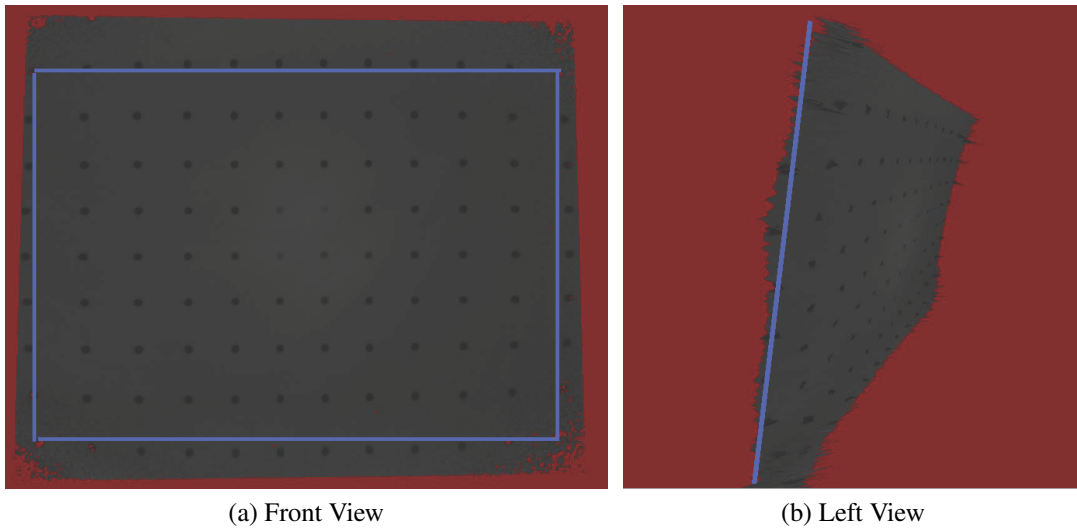

(a) Front View (b) Left View

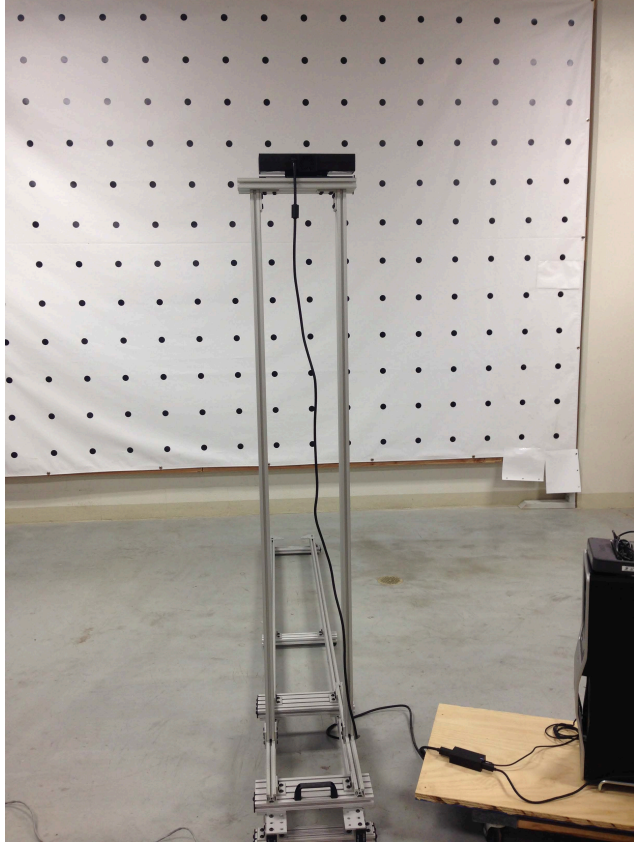Figure 1.7: Raw Distorted KinectV2 NearIR 3D Reconstruction

Figure 1.8: KinectV2 Calibration System

wall, and the blue straight line is added on the left side of the 3D reconstruction. We can tell that most pixels on the left side border are apparently not sitting on a straight line.

In order to get enough calibrating points that can cover all pixels of the sensor, as well as solving the "Depth Distortion", a moving plane with camera on rail calibration system is introduced in this thesis. As shown in figure 1.8, the rail is perpendicular to the uniform round dots pattern, along with the Z^w axis. A RGB-D camera KinectV2 is mounted on the top of the slider, whose moving along the rail generates same effect as a moving pattern plane. This calibration system makes it possible to do per-pixel "Depth Distortion" calibration.

1.4 Summation

The depth sensor technologies opens a new epoch for three-dimensional (3D) markets. Ever since the Microsoft brought the low-cost depth camera Kinect into consumer market, RGB-D cameras are famous in in research areas of HCI, and an accurate RGB-D camera calibration is now important in computer vision. However, the traditional camera calibra-

tion methods are not ideal on accuracy and efficiency. They do not cover all pixels of a sensor, and did not consider the problem of depth "Depth Distortion" either. Besides, those methods need the distortion removal to be a second step after raw 3D reconstruction (from a pinhole camera matrix), which is not an efficient way displaying the 3D reconstruction in real-time.

In this thesis, a novel per-pixel calibration method with look-up table based 3D reconstruction in real-time is introduced, using a rail calibration system shown in fig. 1.8. With the pinhole camera model thoroughly abandoned, this proposed calibration method directly determines a beam equation for every single pixel by collecting undistorted X^w/Y^w with external accurate Z^w values. X^w/Y^w values will be generated from high order polynomial transformation, concretely $4th$ order, directly from image space *row* and *column*. And the accurate Z^w values are imported from external Z^w-aixs tracking module. In short, the proposed calibration method will be totally based on image streams data.

In Chapter 2, a pinhole-camera-model based calibration method is discussed in detail, including the lens distortions removal as the second step. The proposed data-based calibration method is well explained in Chapter 3, where two polynomial models will be introduced. One lens-distortions removal model helps map directly from image plane *row* and *column* to X^w and Y^w separately, during calibration data collection. And another "depth distortion" removal model helps map per-pixel depth (D) to Z^W when generating look-up table for per-pixel beam equation. Z^w values will be totally supported from external BLE optical-flow sensor, will be discussed in Chapter 4. Finally, a *row* by *column* by 6, three dimensional look-up table will be generated for real-time 3D reconstruction. Chapter 5 will talk the future work of RGB-D cameras calibration.

Bibliography

[1] D. C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8), 1971.

[2] Three Depth-Camera Technologies Compared. F. Pece and J. Kautz and T. Weyrich". *First BEAMING Workshop*, June 2011.

[3] Yunxin Duan, Hui Deng, and Feng Wang. Depth Camera in Human-Computer Interaction: An Overview. In *Intelligent Networks and Intelligent Systems (ICINIS), 2012 Fifth International*, 2012.

[4] Felix Endres, Jurgen Hess, Nikolas Engelhard, Jurgen Sturm, Daniel Cremers, and Wolfram Burgard. An evaluation of the RGB-D SLAM system. In *IEEE International Conference on Robotics and Automation (ICRA)*, May 2012.

[5] W. Faig. Calibration of close-range photogrammetry systems: Mathematical formulation. *Photogrammetric Engineering and Remote Sensing*, 41(12), 1975.

[6] Maurice F. Fallon, John Folkesson, Hunter McClelland, and John J. Leonard. Relocating Underwater Features Autonomously Using Sonar-Based SLAM. *IEEE Journal of Oceanic Engineering*, 38(3), Jul 2013.

[7] Olivier Faugeras. Three-Dimensional Computer Vision. *MIT Press*, Nov 1993.

[8] Nadia Figueroa, Haiwei Dong, and Abdulmotaleb El Saddik. From Sense to Print: Towards Automatic 3D Printing from 3D Sensing Devices. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2013.

[9] Ammar Hattab and Gabriel Taubin. 3D Modeling by Scanning Physical Modifications. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug 2015.

[10] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments. In *Experimental Robotics*, 2014.

[11] C. W. Hull. Apparatus for production of three-dimensional objects by stereolithography,. *US Patent US4575330 A*, Mar 1986.

[12] Dan Ionescu, Viorel Suse, Cristian Gadea, and Bogdan Solomon. A single sensor NIR depth camera for gesture control. In *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, May 2014.

[13] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, and Richard Newcombe. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011.

[14] Kourosh Khoshelham and Sander Oude Elberink. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors*, 12(2), Feb 2012.

[15] Henrik Kretzschmar, Cyrill Stachniss, and Giorgio Grisetti. Efficient information-theoretic graph pruning for graph-based SLAM with laser range finders. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep 2011.

[16] Kam Lai, Janusz Konrad, and Prakash Ishwar. A gesture-driven computer interface using Kinect. In *Image Analysis and Interpretation (SSIAI)*, Apr 2012.

[17] P. Lavoie, D. Ionescu, and E. M. Petriu. 3-D object model recovery from 2-D images using structured light. In *Instrumentation and Measurement Technology Conference*, Jun 1996.

[18] Jaehong Lee, Heon Gul, Hyungchan Kim, Jungmin Kim, Hyoungrae Kim, and Hakil Kim. Interactive manipulation of 3D objects using Kinect for visualization tools in education. In *Control, Automation and Systems (ICCAS)*, Oct 2013.

[19] Larry Li. Time-of-Flight Camera – An Introduction. *Texas Instruments Technical White Paper*, SLOA190B, 2012.

[20] Maohai Li, Rui Lin, Han Wang, and Hui Xu. An efficient SLAM system only using RGBD sensors. In *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2013.

[21] Kathia Melbouci, Sylvie Naudet Collette, Vincent Gay-Bellile, Omar Ait-Aider, Mathieu Carrier, and Michel Dhome. Bundle adjustment revisited for SLAM with RGBD sensors. In *IAPR International Conference on Machine Vision Applications (MVA)*, May 2015.

[22] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, and Andrew J. Davison. KinectFusion: Real-time dense surface mapping and

tracking. In *IEEE International Symposium on Mixed and Augmented Reality (IS-MAR)*, Oct 2011.

[23] Meenakshi Panwar. Hand gesture recognition based on shape parameters. In *2012 International Conference on Computing, Communication and Applications*, Feb 2012.

[24] Carolina Raposo, Joao Pedro Barreto, and Urbano Nunes. Fast and Accurate Calibration of a Kinect Sensor. In *International Conference on 3D Vision*, Jun. 2013.

[25] Sebastian A. Scherer and Andreas Zell. Efficient onbard RGBD-SLAM for autonomous MAVs. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013.

[26] Jan Smisek, Michal Jancosek, and Tomas Pajdla. 3D with Kinect. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov. 2011.

[27] Aaron N. Staranowicz, Christopher Ray, and Gian-Luca Mariottini. Easy-to-use, general, and accurate multi-Kinect calibration and its application to gait monitoring for fall prediction. In *Engineering in Medicine and Biology Society (EMBC)*, Aug. 2015.

[28] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4), Aug 1987.

[29] Jodie Wetherall, Matthew Taylor, and Darren Hurley-Smith. Investigation into the Effects of Transmission-channel Fidelity Loss in RGBD Sensor Data for SLAM. In *International Conference on Systems, Signals and Image Processing (IWSSIP)*, Sep 2015.

[30] Guan xi Xin, Xu tang Zhang, Xi Wang, and Jin Song. A RGBD SLAM algorithm combining ORB with PROSAC for indoor mobile robot. In *International Conference on Computer Science and Network Technology (ICCSNT)*, Dec 2015.

[31] Cheng-Kai Yang, Chen-Chien Hsu, and Yin-Tien Wang. Computationally efficient algorithm for simultaneous localization and mapping (SLAM). In *IEEE International Conference on Networking, Sensing and Control (ICNSC)*, Apr 2013.

[32] Zhengyou Zhang. Camera Calibration. (Chapter 2). *Emergin Topics in Computer Vision*, 2004.

[33] Xinshuang Zhao, Ahmed M. Naguib, and Sukhan Lee. Kinect based calling gesture recognition for taking order service of elderly care robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Aug 2014.