A Novel High-Speed Stereo-Vision System for Real-time Position Sensing

by

Niankun Rao

B.E., Harbin Institute of Technology, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES
(Mechanical Engineering)

The University Of British Columbia
(Vancouver)

December 2011

Abstract

Real-time position sensing has a wide range of applications in motion control systems, parts inspection and general metrology. Vision-based position sensing systems have significant advantages over other sensing methods, including large measurement volume, non-contact sensing, and simultaneous measurement in multiple degrees-of-freedom (DOF). Existing vision-based position sensing solutions are limited by low sampling frequency and low position accuracy. This thesis presents the theory, design, implementation and calibration of a new high-speed stereo-vision camera system for real-time position sensing based on CMOS image sensors.

By reading small regions around each target image rather than the full frame data of the sensor, the frame rate and image processing speed are vastly increased. A high speed camera interface is designed based on Camera Link technology, which allows a maximum continuous data throughput of 2.3Gbps. In addition, this stereo-vision system also includes fixed pattern noise (FPN) correction, threshold processing, and sub-pixel target position interpolation.

In order to achieve high position accuracy, this system is calibrated to determine its model parameters. The primary error sources in this system include target image noise, mechanical installation error and lens distortion. The image sensor is characterized, and its FPN data is extracted, by experiment. The mechanical installation error and lens distortion parameters are identified through camera calibration. The proposed camera calibration method uses the 3D position reconstruction error as its cost function in the iterative optimization. The optimization of linear and nonlinear parameters is decoupled. By these means, better estimation of model parameters is achieved. To verify the performance of the proposed calibration method, it is compared with a traditional single camera calibration method in simulation and experiment. The results show that the proposed calibration method gives better parameter estimation than the traditional single camera calibration method.

The experimental results indicate that the prototype system is capable of measuring 8 targets in 3-DOF at a sampling frequency of 8kHz. Comparison with a coordinate measurement machine (CMM) shows that the prototype system achieves a 3D position accuracy of 18μm (RMS) over a range of 400mm by 400mm by 15mm, with a resolution of 2μm.

Table of Contents

Abstract . . . . **ii**

Table of Contents . . . . **iii**

List of Tables . . . . **v**

List of Figures . . . . **vi**

List of Acronyms . . . . **ix**

Acknowledgments . . . . **x**

1 Introduction **1**

1.1 Background 1

1.2 Prior Art 2

1.2.1 LEPD-based Position Sensing Systems 2

1.2.2 CCD-based Position Sensing Systems 3

1.2.3 CMOS-based Position Sensing Systems 5

1.2.4 Summary of Prior Art 6

1.3 Thesis Overview 6

2 Vision-based Position Sensing Theory **10**

2.1 Imaging Model 10

2.1.1 Rigid Body Transformation 10

2.1.2 Perspective Projection 11

2.1.3 Lens Distortion 12

2.1.4 Image Digitization 14

2.2 ROI-based Image Processing 15

2.2.1 Initial Target Identification 16

2.2.2 Sub-pixel Target Position Interpolation 16

2.2.3 ROI Update 18

2.3 3D Position Reconstruction 19

2.4 Summary 22

3 System Hardware Design **23**

3.1 Hardware Design Overview 23
3.2 High-Performance Electronics Design 24
3.2.1 Image Sensor 25
3.2.2 Camera Control Unit 29
3.2.3 Analog-to-Digital Electronics Design 29
3.2.4 High-Speed Camera Interface Design 33
3.2.5 High-Speed Image Processing 37
3.2.6 Power Distribution System Design 41
3.2.7 Print Circuit Board Design 43
3.3 System Optical and Mechanical Design 45
3.3.1 Infrared Target Design 45
3.3.2 Lens Selection 47
3.3.3 Camera Body Design 49
3.4 Summary 51

4 System Calibration 54
4.1 System Limitations 54
4.2 Fixed Pattern Noise of Image Sensor 55
4.3 Calibration of Optical and Mechanical Parameters 57
4.3.1 Camera Calibration Methods Review 58
4.3.2 Proposed Camera Calibration Method 61
4.3.3 Simulation Results 63
4.4 Summary 67

5 Experimental Results 70
5.1 FPN Characterization and Correction 70
5.2 Position Sampling Frequency Characterization 73
5.3 Characterization of Sub-pixel Target Position Interpolation 75
5.4 Camera Calibration and Off-line 3D Reconstruction 77
5.5 Characterization of Real-time 3D Reconstruction 89
5.6 Case Study: Long-Stroke Planar Motion Stage 93
5.7 Summary 95

6 Conclusions and Future Work 99
6.1 Consclusions 99
6.2 Future Work 101

Bibliography 103

List of Tables

Table 1.1 Comparison of prior art vision-based position sensing systems 6

Table 3.1 Comparison of CCD and CMOS image sensor, data from [28] 26
Table 3.2 Specifications of LUPA-4000 image sensor, data from [30] 28
Table 3.3 Electrical specifications of LTC2299, data from [32] 31
Table 3.4 Comparison of camera interfaces, data from [33] 37
Table 3.5 Summary of camera module power requirements 41
Table 3.6 Hardware specifications of the prototype system 53

Table 4.1 Simulation case 1 - calibration results of Camera 0 65
Table 4.2 Simulation case 1 - calibration results of Camera 1 65
Table 4.3 Simulation case 2 - calibration results of Camera 0 66
Table 4.4 Simulation case 2 - calibration results of Camera 1 67
Table 4.5 Simulation case 2 - comparison of 3D position reconstruction error 67
Table 4.6 Simulation case 3 - calibration results of Camera 0 (0.1 pixel sub-pixel interpolation error) 68
Table 4.7 Simulation case 3 - calibration results of Camera 1 (0.1 pixel sub-pixel interpolation error) 68

Table 5.1 Average gain ratio between even and odd pixels before compensation 71
Table 5.2 Average gain ratio between even and odd pixels after compensation 72
Table 5.3 Average gain ratio between even and odd pixels of small regions 73
Table 5.4 Specifications of Crysta-Apex C7106, data from [47] 78
Table 5.5 Effect of types of lens distortion on the calibration performance of the NK method . 87
Table 5.6 Effect of order of curvature distortion on the calibration performance of the NK method 87
Table 5.7 Effect of skew angle on the calibration performance of the NK method 87
Table 5.8 Calibration results of the Heikkila Method 89
Table 5.9 Calibration results of the NK Method 90
Table 5.10 RMS error of calibration results 90

Table 6.1 Specifications of the prototype system 100

List of Figures

Figure 1.1 Schematic of HiBall system adapted from [3] 3

Figure 1.2 Working principle of three linear CCDs configuration adapted from [8] 4

Figure 1.3 Working principle of Muehlmann's Design (tracking 1, 3, and 8 targets) adapted from [10] 5

Figure 1.4 Prototype of the real-time position sensing system based on stereo-vision 7

Figure 1.5 Illustration of the metrology system of long-stroke planar motor 8

Figure 2.1 Illustration of the rigid body transformation 11

Figure 2.2 Perspective projection of a point from camera coordinate to image coordinate . . . 12

Figure 2.3 Illustration of radial distortion adapted from [14] 13

Figure 2.4 Illustration of the skewness error of image sensor 15

Figure 2.5 Illustration of region of interest 15

Figure 2.6 2D Gaussian distribution 18

Figure 2.7 Schematic of ROI update adapted from [11] 19

Figure 2.8 Schematic of stereo-vision based on two image sensors 20

Figure 3.1 Overall system architecture 24

Figure 3.2 Electronic architecture of camera module 25

Figure 3.3 Structure of CCD and CMOS image sensor adapted from [26] 26

Figure 3.4 LUPA-4000 CMOS image sensor from Cypress Semiconductor 28

Figure 3.5 Analog-to-Digital conversion electronics architecture 30

Figure 3.6 Voltage reference as power supply to DAC8411 32

Figure 3.7 Schematic of signal conditioning circuits 33

Figure 3.8 Passive LC filter between analog and digital power supply in ADC and DAC . . . 33

Figure 3.9 Block diagram of base, medium and full configuration of Camera Link adapted from [35] 36

Figure 3.10 Camera module FPGA architecture 38

Figure 3.11 Tsunami real-time computer mother board developed by Kris Smeds adapted from [36] 39

Figure 3.12 Tsunami I/O Board Designed by Niankun Rao, Richard Graetz and Arash Jamalian 40

Figure 3.13 Power regulation network of camera module 41

Figure 3.14 Passive LC filter between 5V DC power and input of linear regulator 42

Figure 3.15 Camera module control PCB 43

Figure 3.16 Image sensor PCB 44
Figure 3.17 Layer stack-up of camera module control PCB 44
Figure 3.18 Layer stack-up of image sensor PCB 44
Figure 3.19 Analog and digital partition on image sensor board 46
Figure 3.20 Spectrum distribution and radiation pattern of HE8812SG Opnext Infrared LED, from [37] 47
Figure 3.21 Infrared LED driving circuits 48
Figure 3.22 Illustration of the relationship between measurement volume and lens parameters . 48
Figure 3.23 Schematic of two-camera system 50
Figure 3.24 Resolution in Z_W VS camera distance d 51
Figure 3.25 Camera body SolidWorks Model 51
Figure 3.26 Hardware assembly of the prototype system 52

Figure 4.1 Block diagram of system 55
Figure 4.2 DSNU and PRNU illustration adapted from [29] 56
Figure 4.3 Noise model of an image sensor adapted from [40] 56
Figure 4.4 Optical and mechanical error sources in a vision system 58
Figure 4.5 Tsai's camera model with perspective projection and radial distortion adapted from [18] 60
Figure 4.6 Illustration of 2D reprojection error in image coordinate 62
Figure 4.7 Illustration of 3D reconstruction error in world coordinate 63
Figure 4.8 Flow chart of the proposed calibration method 64
Figure 4.9 3D reconstruction error VS. sub-pixel interpolation error 69

Figure 5.1 DSNU of the image sensor at 10μs exposure time 71
Figure 5.2 Histogram comparison between a raw black image and the image after DSNU correction (The mean value is removed) 72
Figure 5.3 Distribution of test regions for average gain difference compensation 73
Figure 5.4 Comparison of a raw LED image and the image after FPN correction 74
Figure 5.5 Ideal frame rate vs. ROI size and number of target 75
Figure 5.6 Real frame rate vs. ROI size and number of target 76
Figure 5.7 Threshold effect on sub-pixel interpolation resolution 77
Figure 5.8 Illustration of camera calibration setup 79
Figure 5.9 Block diagram of camera calibration and its performance evaluation 80
Figure 5.10 Test points distribution 1 81
Figure 5.11 Effect of DSNU correction on calibration performance 82
Figure 5.12 Effect of average gain compensation on calibration performance 83
Figure 5.13 Effect of threshold on calibration performance 84
Figure 5.14 Test points distribution 2 85
Figure 5.15 Test points distribution 3 85
Figure 5.16 Effect of test point distribution on calibration performance 86

Figure 5.17 Effect of test point quantity on calibration performance 86
Figure 5.18 Real-time 3D reconstruction resolution at 8kHz sampling frequency 91
Figure 5.19 Real-time 3D reconstruction accuracy at 8kHz sampling frequency 92
Figure 5.20 Histogram of 3D reconstruction error 93
Figure 5.21 3D reconstruction error plotted in test point sequence 94
Figure 5.22 Infrared LED targets mounting on the moving stage 95
Figure 5.23 Installation of the position sensing system for a long-stroke planar motion stage . . 97
Figure 5.24 Resolution of 6-DOF measurement 98

List of Acronyms

ADC	analog-to-digital converter.
CCD	charge-coupled device.
CMM	coordinate measurement machine.
CMOS	complementary metal-oxide-semiconductor.
DAC	digital-to-analog converter.
DCM	digital clock manager.
DLT	direct linear transformation.
DOF	degree-of-freedom.
DSNU	dark signal non-uniformity.
DSP	digital signal processor.
FOV	field of view.
FPGA	field programmable gate array.
FPN	fixed pattern noise.
HDL	hardware description language.
LED	light-emitting diode.
LEPD	lateral effect photodiode.
LSB	least significant bit.
LVDS	low voltage differential signaling.
PCB	printed circuit board.
PDS	power distribution system.
PRNU	photo response non-uniformity.
PSNL	pixel storage node leakage.
RAM	random-access memory.
ROI	region of interest.
SNR	signal-to-noise ratio.
UART	universal asynchronous receiver/transmitter.
USB	universal serial bus.

Acknowledgments

Firstly, I am most grateful to my supervisor Dr. Xiaodong Lu for his encouragement, guidance and help on my project. He provides me with world class experience and introduce me to the research field of precision mechatronics. Through his undergraduate and graduate courses, I built up my knowledge and skills for my research. He is always patient and enthusiastic to discuss and explain the problems with me both on the theoretical part and experimental results. With his drive to succeed, passion for exploring unknown world and rigorous attitude to research, he sets a good role model for me to learn from. His ability to explain complex phenomenon from first principles and physical intuition leaves me a deep impression. My two-year working with him taught me how to be a good engineer and, more important, how to be a good person. It is my great honor to work with him as my mentor, as well as a friend.

Professor Yusuf Altintas taught me machine tool vibration in his graduate course. With his encouragement and help, I overcame my weakness in this field and did the best in his course, which significantly increases my confidence of studying overseas. Through his course, I had the opportunity to investigate this area in more detail. I appreciate his mentorship. I am thankful to Dr. Hsi-Yung Feng and Dr. Ryozo Nagamune for being my examination committee. Dr. Ryozo Nagamune taught me the courses on modern control and robust control theories, and offered me many helps in job and scholarship application. His rigorous attitude on teaching and research impressed me.

I worked with Kris Smeds for two years. It is him who taught me a lot in electronics design and VHDL programming. When I had questions on circuit design, I often turned to him. His passion, hard-working and well-organized style on research influenced me. I always enjoyed the discussion with him on hardware design and VHDL coding issues. I would like to thank Irfan Usman for his help in my project. He always gave me valuable feedback and suggestions on my mechanical design of camera body. Working with him as the teaching assistant in MECH 421 is one of the most enjoyable experience in UBC. I appreciate his help and feedback when I am writing my thesis. I also thank my labmate Yingling Huang. Her master project in camera calibration provides a foundation for my research. It has also been a great pleasure working with my other colleagues Richard Graetz, Arash Jamalian, Fan Chen and Xiaoling Jin. I must mention my good friends around me in UBC. It is my great fortune to have them. Living with them in UBC makes my overseas studying colorful.

I am most indebted to my parents. They are my source of support, love and encouragement for my study at UBC. They are the reason I have gotten to where I am today and I appreciate their tremendous sacrifices. I will always love them.

Chapter 1

Introduction

1.1 Background

Position sensing has a wide range of applications, such as parts inspection, providing motion feedbacks for control system and general metrology. In applications, such as robotic guidance and control, computer-assisted surgery and planar motion stage, where the position and orientation of the objects are captured simultaneously, multiple degree-of-freedom (DOF) motion sensing is required. In high-performance real-time motion control systems that require high bandwidth, large motion range and high accuracy, the metrology system must provide high position sampling frequency and high position accuracy. Conventional solutions for position sensing include linear variable differential transformer (LVDT), capacitance probe, laser tracker, optical encoder, interferometer and accelerometer, among others. Most conventional position sensors are limited to measurement of a single-DOF, therefore multiple conventional sensors are required when measuring multiple DOFs motion, increasing the system complexity.

Developments in optical sensor technology make the vision-based position sensing system an attractive solution for real-time position sensing where the object to be tracked has multi-DOF motion in a large moving volume. The advantages of a vision-based position sensing system include large measurement volume, non-contact measurement and simultaneous measurement in multiple DOFs. Currently, the charge-coupled device (CCD) image sensor is a mature technology that is widely applied in many vision systems. The image quality generated by CCD image sensors is already excellent but at a low frame rate. Commercial products based on CCD image sensor have already been successfully used in applications, such as computer-assisted surgery and virtual reality. However, because it is limited by low frame rate and high power dissipation, CCD is not desirable for high-speed vision applications.

There is a tradeoff between the sampling frequency and accuracy in position sensing systems based on CCD. On one hand, the position accuracy is increased by using high resolution CCD. On the other hand, high resolution of image sensor increases the frame readout time, causing lower position sampling frequency. Recent developments in complementary metal-oxide-semiconductor (CMOS) technology make the CMOS image sensor a promising solution for high-speed vision systems. Taking advantage of random pixel accessibility of the CMOS image sensor, the region of interest (ROI) readout is enabled by reading small regions around target image rather than the full frame image data, which vastly reduces the readout time of the image sensor.

A preliminary prototype of the optical tracking system using a commercial CMOS camera was developed by Yingling Huang as a proof of concept to achieve high position accuracy by using a CMOS camera [1]. In her thesis, camera model and calibration methods are investigated and discussed. This prototype achieves a 3D position accuracy of 40μm RMS over a range of 500mm by 500mm by 10mm with 1m distance between the camera and the measurement volume. However, the high speed features of the CMOS image sensor are not investigated in her thesis. Based on these results, a customized vision-based position sensing system is developed in this thesis. In this thesis, a prototype of stereo-vision system for real-time position sensing is designed and built based on CMOS image sensor, which features large measurement volume, high sampling frequency and high position accuracy.

In the rest of this chapter, some backgrounds of vision-based position sensing system are presented, and the prior art are reviewed. Next, the overview of this thesis is presented and the thesis structure is outlined.

1.2 Prior Art

A significant amount of research effort has been devoted to the design of vision-based position sensing systems. There are many designs reported in the literature based on various mechanisms and configurations.

Vision-based position sensing systems can be classified according to the optical sensor they use. Non-imaging sensors such as lateral effect photodiodes (LEPDs) are pure analog sensors that determine the centroid of all the light in the field of view (FOV). They require no digital image processing, but care must be taken to ensure that the light seen by the sensor at any given time comes from a single bright target. These sensors are often used with active light source targets that are able to work in time multiplexing mode. On the other hand, image sensors, such as CCD and CMOS, require some digital computations to find the position of target in an image. They have advantages that the positions of multiple targets can be retrieved using a single image frame. Accurate target positions can be obtained even if there is background noise, as long as the image processing algorithm is smart enough to distinguish the actual targets from the background.

1.2.1 LEPD-based Position Sensing Systems

Lateral effect photodiode (LEPD) is a component that generates a signal proportional to the position of the centroid of incident light on its 2D axes. With other optical and electrical components, a system based on this type of sensor can measure angle and distance. LEPD provides position information which is linear and independent of the intensity profile of the illumination. Further, its linearity is independent of the symmetrical defocusing that might occur in optical systems. The major disadvantage of the LEPD is high noise. Because linearity considerations require a small impedance between lateral contacts, the output signal of the LEPD is typically noisy [2].

One position sensing system based on LEPDs is the HiBall system [3]. The HiBall was developed by the University of North Carolina at Chapel Hill in the early 1980s. This system uses arrays of ceiling mounted infrared light-emitting diodes (LEDs) as targets (Figure 1.1). Multiple LEPDs are installed to detect the positions of infrared LEDs relative to the LEPD sensors. These LEPD devices generate a current proportional to the position of the light that falls on the photodiodes. By using multiple LEPD sensors looking outwards in different directions, the HiBall achieves a large FOV. This system achieves a position sampling frequency of 2000Hz. However, this system only achieves an accuracy of 500μm in position measurement.

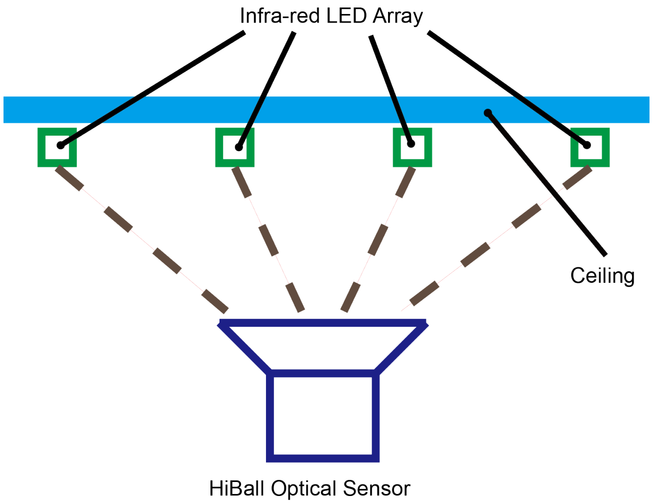


Figure 1.1: Schematic of HiBall system adapted from [3]

1.2.2 CCD-based Position Sensing Systems

Charge-coupled device (CCD) image sensor technology was invented in the late 1960s and has dominated the vision system for the past 25 years. As a result of its large dynamic range, high uniformity and low system noise, CCD can offer high quality images at low frame rate. On the other hand, CCD are limited by their high driving voltage, and their pixel shifting structure is unable to meet the requirement of high-speed vision applications. These disadvantages of the CCD image sensor significantly limit its application in high-speed vision systems [4].

Advanced Realtime Tracking Inc. designed an optical tracking system in 2000, named ARTtrack [5]. ARTtrack uses multiple specially designed CCD cameras which emit flashes of infrared light. These cameras are located surrounding the object being tracked. Light from the cameras bounces off reflective spheres attached to the object and is reflected back to the camera. The orientation and position of the retro-reflectors are known with respect to a reference point. This information is exploited to retrieve the position and orientation of the object. This system can track multiple objects simultaneously in a large measuring depth (up to 6m). The position sampling frequency is limited to 60Hz and the position accuracy is 1mm.

Mathieu Herve, in 2005, introduced a 6-DOF optical tracking system for virtual reality applications, named Cyclope [6]. The Cyclope tracker is based on a single CCD camera and infrared retro-reflectors. The CCD camera captures the image of the retro-reflectors which have a pre-defined pattern. Given

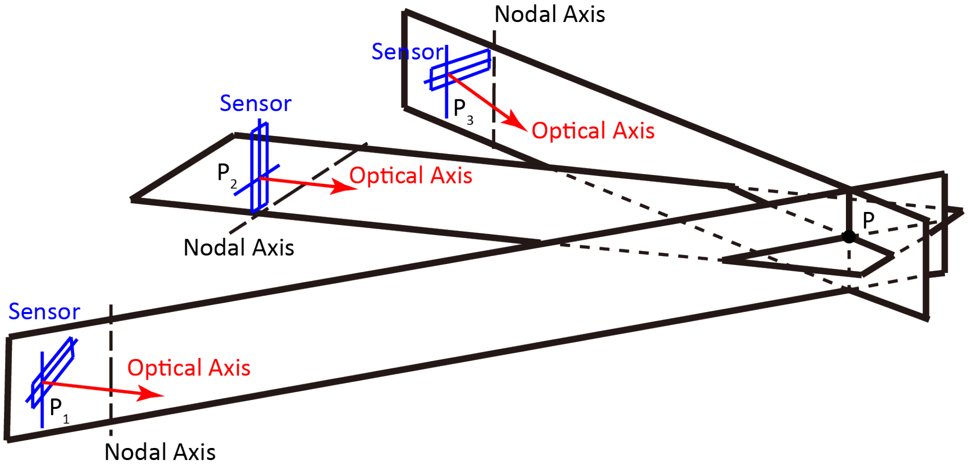


Figure 1.2: Working principle of three linear CCDs configuration adapted from [8]

the 3D positions of these non-coplanar retro-reflectors in the image, the position and the orientation of the reference frame attached to the 3D points with respect to the camera reference frame are retrieved. The Cyclope tracker runs at a position sampling frequency of 60Hz with 30ms intrinsic latency. The position measurement error is 1mm.

The Polaris system [7], launched by Northern Digital Inc. (NDI) in 2005, is a highly versatile, low cost, real-time stereo-vision technology, which supports the needs of medical imaging applications. Polaris uses two area CCDs to capture the motion of active or passive targets which emit or reflect infrared light. Based on the principle of two camera stereo-vision, the position of the target is obtained by triangulation. The Polaris system achieves a position sampling frequency of 60Hz and position accuracy of 250μm, and is mainly limited by its area CCD.

For CCD image sensors, the speed of delivery for pixel data sets the upper limit for frame rate. This limit arises because a CCD sensor must transfer out all of its pixel information in order to empty its transfer register so that it can accept the next image [4]. For a given pixel rate, the higher the image resolution the lower the frame rate. Based on this idea, the linear CCD with much lower image resolution (relative to a high-resolution area CCD) was applied. In 1983, V. Macellari designed an optical tracking system based on three linear CCDs for sensing the motion of human body [8] (shown in Figure 1.2). In 2006, NDI released the OPTOTRAK system using a similar configuration [9]. OPTOTRAK system is a portable coordinate measurement machine (CMM) based on NDI's optical measurement solutions. The system's solid mechanical design ensures a great measurement reliability. By using linear CCDs, the readout time of a single frame is significantly reduced and a frame rate of 4kHz is achieved. However, limited by the linear CCD configuration, OPTOTRAK can only capture one target image in a single frame so that the position sampling frequency is reduced when measuring multiple targets. OPTOTRAK system must be operated with smart targets working in time-multiplexing mode. The position sampling frequency decreases to 575Hz for 6 targets. Meanwhile, OPTOTRAK achieves a position accuracy of 100μm.

1.2.3 CMOS-based Position Sensing Systems

Complementary metal-oxide-semiconductor (CMOS) image sensor technology was developed in the early 1970s, but its performance was limited by the available lithography technology at that time. However, recent developments in lithography technology and process control in CMOS fabrication make CMOS image sensor a strong competitor to CCD [4]. Compared to CCD, CMOS image sensors have low-power consumption. In the applications of high-speed vision systems, they allow short exposure time and fast image readout. CMOSs have a speed advantage as on-board circuitry allows for low propagation delays and conversion to the digital domain physically closer to the actual pixels. Further, CMOS image sensors have the ability of direct pixel access, which allows dynamically changing the ROI location and size. These features make the CMOS image sensor a promising technology in high-speed vision applications. However, CMOS image sensors offer high integration, low power dissipation, small system size and high speed at the expense of image quality [4].

Ulrich Muehlmann, in 2004, designed a new high-speed vision system for real-time tracking applications, combining CMOS image sensors, field programmable gate array (FPGA) technology and universal serial bus (USB) interface [10]. The main advantage of this vision system is the application of the ROI readout of the CMOS image sensor. By this means, when the system is tracking small targets in a wide background, only the small areas around each target are read out, significantly reducing the amount of raw image data which needs to be read in a single frame (Figure 1.3). The position sampling frequency of this system is 800Hz for 8 targets. The correction of image noise is integrated on board. Since the exposure time and illumination conditions affect the image noise data, different calibration maps under different conditions are stored on the board memory. The main bottle neck of this system comes from its USB 2.0 interface between camera modules and host PC. The camera control information, including ROI size, ROI position, etc., and the raw image data share the same USB interface operating in host-slave mode. However, the bandwidth of this USB interface is limited to 480Mbps, and there is high latency due to its inherent micro-frame synchronization mechanism so that the data throughput is greatly limited, causing low position sampling frequency. In addition, this system lacks a good motion predictor to estimate the ROI positions in the next frame. A full frame recovery is required if any target is lost.

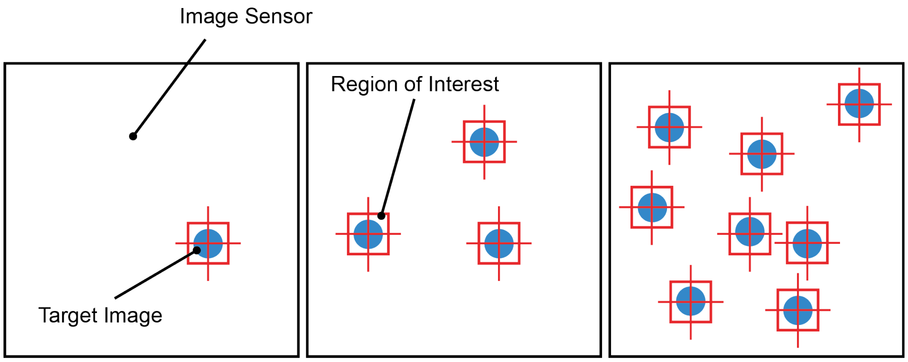


Figure 1.3: Working principle of Muehlmann's Design (tracking 1, 3, and 8 targets) adapted from [10]

Crispin D. Lovell-Smith designed and implemented a prototype inside-out vision-based position sensing system in 2009, named Black Spot [11]. A CMOS image sensor was integrated with digital signal processor (DSP), and multiple camera modules are used to form a camera hub to achieve large measurement volume. The idea of ROI-based image processing is applied: only small regions which contain the target image are processed, which reduces the computation time of the system and, therefore, increases position sampling frequency. However, this idea is not extended to the image readout. This design does not fully take advantage of the ROI readout capability of CMOS image sensor, thus a full frame readout is still required for every processing cycle, severely limiting the image acquisition speed. The universal asynchronous receiver/transmitter (UART) serial interface limits the data throughput between the camera modules and the host PC. The position accuracy is limited to 800 μm caused by poor camera calibration and low resolution of image data (8-bit).

1.2.4 Summary of Prior Art

To summarize the prior art of vision-based position sensing, the systems discussed above are compared in Table 1.1. Shown in Table 1.1, none of the current commercial products or designs is able to achieve high position sampling frequency ($>$10kHz) and high position accuracy ($<$10μm). Based on previous designs, a novel vision-based position sensing system based on CMOS image sensor is developed in this thesis in order to achieve the high position sampling frequency and simultaneously high position accuracy.

Table 1.1: Comparison of prior art vision-based position sensing systems

Name	Sensor Type	Position Sampling Frequency	Accuracy
HiBall	LEPD	2000Hz	500μm
Cyclope	Area CCD	60Hz	1000μm
ARTtrack	Area CCD	60Hz	1000μm
Polaris	Area CCD	60Hz	250μm
OPTOTRAK	Linear CCD	575Hz for 6 targets	100μm
Black Spot	CMOS	60Hz	800μm
Muehlmann’s Design	CMOS	800Hz for 8 targets	Not Available

1.3 Thesis Overview

In this thesis, a new prototype of high-speed vision-based position sensing system based on CMOS image sensor is designed and built (shown in Figure 1.4). The goal of this thesis is to fully realize the high speed potential of the CMOS image sensor through advanced electronics configuration, and develop associated enabling technologies to demonstrate the achievable performance.

To demonstrate the usability of this novel high-speed position sensing system, this system is in-

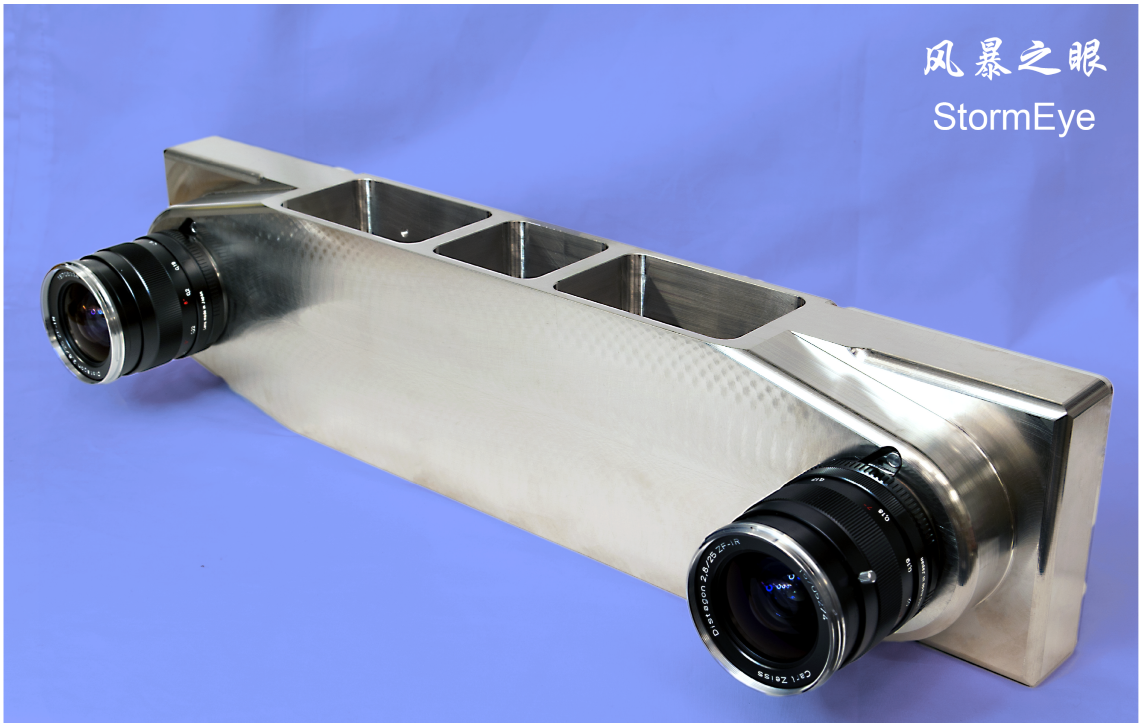


Figure 1.4: Prototype of the real-time position sensing system based on stereo-vision

tegrated as the metrology solution for a long-stroke planar motor (shown in Figure 1.5). The motion stage moves in the range of 1m by 1m in the $X_W - Y_W$ plane with 10mm in Z_W direction. By tracking the position of targets mounted on the armature, the 6-DOF motion of the armature is retrieved.

The thesis is structured as follows:

Chapter 1 introduces the background of vision-based position sensing systems and reviews the prior art. An overview of the high-speed vision-based position sensing system is presented.

Chapter 2 describes the supporting theories behind the system. The imaging model is first built where the transformation of target position from world coordinates to pixel coordinates is given. The ROI-based image processing is presented, including initial target detection, sub-pixel target position interpolation and ROI position update. Further, the 3D position reconstruction algorithm is presented.

Chapter 3 covers the detailed design and implementation of the stereo-vision system hardware. It begins by establishing the design objectives for the stereo-vision system and then presents the electronic hardware, optical and mechanical design. The electronics hardware design covers the imaging electronic architecture and also provides details on the high-performance customized electronics development, integration and realization. The optical and mechanical design is based on the overall system design objectives, providing high image quality and a stable mechanical structure.

Chapter 4 presents the calibration method of the prototype system. The limitations and error sources of vision-based position sensing system are analyzed and discussed. The image sensor is modeled and its noise correction method is presented. Further, the calibration of the stereo-vision system is presented. The prior art of camera calibration methods are reviewed and the proposed calibration method is described. Simulation results are presented to demonstrate the potential performance that the

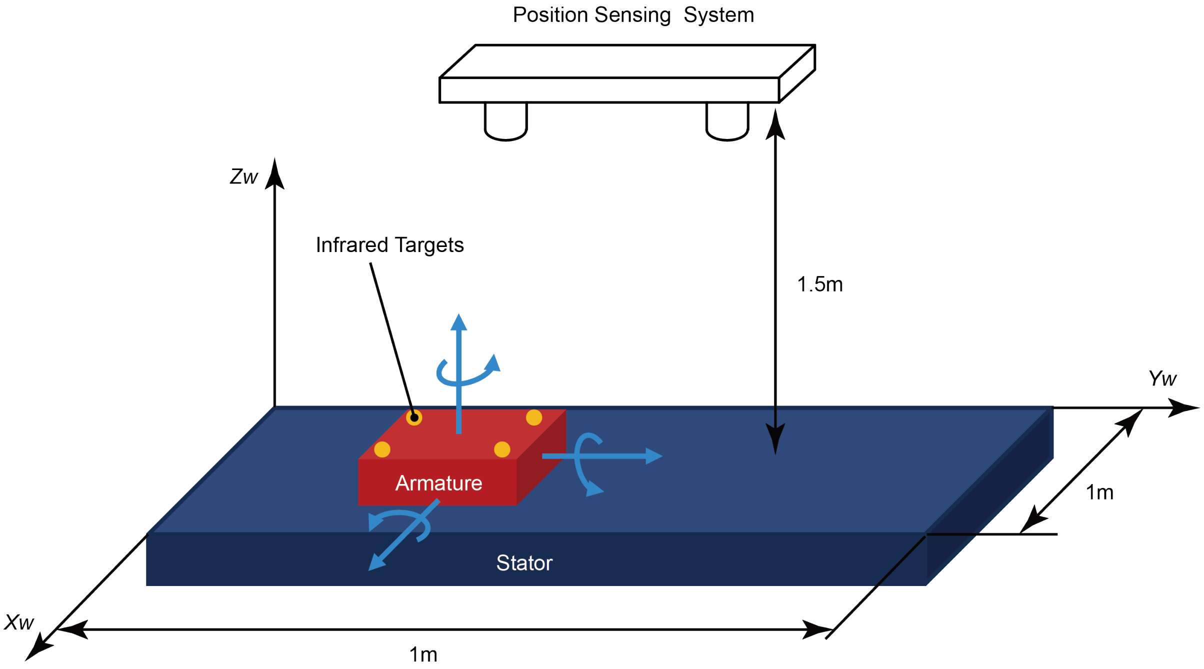


Figure 1.5: Illustration of the metrology system of long-stroke planar motor

proposed calibration method is able to achieve.

Chapter 5 presents the experimental results of the overall system performance. First, the image sensor is experimentally characterized and the effectiveness of image noise correction method is demonstrated by experiment. Second, the position sampling frequency is measured based on different target numbers and ROI sizes. Third, the stereo-vision system is calibrated using CMM. Multiple effects which influence the calibration performance are investigated. Finally, the 3D reconstruction method is realized in real-time, and its precision, repeatability and accuracy are characterized.

Chapter 6 concludes the thesis with an overview of the presented results and a discussion of the system limitations. Future work on the prototype is outlined.

The main contributions of this thesis can be summarized as follows:

- The design and implementation of a new high-speed stereo-vision system. This vision system is designed based on CMOS image sensors and FPGA technology, which features high-performance image acquisition, high-speed camera interface and fast image processing. By advanced electronics design, the camera modules in this system are not only limited to the function of raw image acquisition but also integrated with more image processing functionalities, such as image noise correction and sub-pixel target position interpolation. Therefore, the sampling frequency of this vision-based position sensing system is significantly increased.

- A new calibration method designed for stereo-vision system. The proposed camera calibration system utilizes the cost function which minimizes the 3D reconstruction error in the nonlinear optimization. Therefore, information provided by the 3D test points are fully utilized in the optimization. Especially in the presence of large measurement errors in sub-pixel target position, the proposed calibration method gives better estimation of model parameters than traditional single camera calibration methods. At the same time, this calibration method decouples the optimization of linear and nonlinear parameters in a stereo-vision system, therefore the harmful interaction between them is suppressed. By these means, better estimation of system model parameters is obtained, and therefore higher position measurement accuracy is achieved.

Chapter 2

Vision-based Position Sensing Theory

Vision-based position sensing systems use mathematical models to calculate 3D positions of targets from their raw image data. Various parameters are incorporated in the model that describe the system's physical attributes, such as lens distortions, camera position and sensor pixel size [12]. In this chapter, the position sensing system based on stereo-vision is modeled and the supporting theories are presented. There have been many investigations and studies on the model of imaging systems [13][14][15][16][17][18][19]. Section 2.1 presents the imaging model that describes the target position transformation from the 3D world coordinate to the 2D image sensor coordinate. Section 2.2 presents the idea of ROI-based image processing and investigates the supporting algorithms. The 3D position reconstruction algorithm is presented in Section 2.3.

2.1 Imaging Model

The imaging model describes the transformation of target position from the 3D world coordinate (X_W, Y_W, Z_W) to the 2D pixel coordinate (u, v). This transformation is divided into four steps: rigid body transformation, perspective projection, lens distortion and image digitization.

2.1.1 Rigid Body Transformation

The rigid body transformation represents the coordinate transformation from the world coordinate $P_W(X_W, Y_W, Z_W)$ to the camera coordinate $P_C(X_C, Y_C, Z_C)$ (shown in Figure 2.1). The origin of camera coordinate is built at the optical center of lens; the Z_C axis is the optical axis of the lens; the X_C axis is parallel with the X_I axis of the image sensor; the Y_C axis is parallel to the Y_I axis of the image sensor. The rigid body transformation uses a translation vector T and a rotation matrix R to represent the relationship between world coordinate and camera coordinate. The transformation is expressed as

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + T. \tag{2.1}$$

The matrix R is a 3 by 3 matrix (shown in Equation 2.2) determined by three Euler angles α, β and γ that represent the rotation of the camera coordinate around X_W-axis, Y_W-axis and Z_W-axis, respectively.

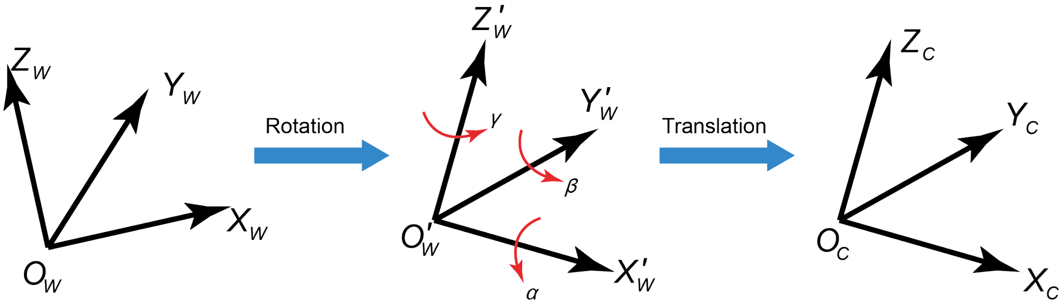


Figure 2.1: Illustration of the rigid body transformation

The vector $T = [T_x, T_y, T_z]^{\mathrm{T}}$ describes the translation between world coordinate and camera coordinate. These six independent parameters which describe the rotation (α, β, γ) and translation (T_x, T_y, T_z) are called extrinsic parameters of the camera.

$$R = \begin{bmatrix} \cos\beta\cos\gamma & \cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\gamma - \cos\alpha\sin\beta\cos\gamma \\ -\cos\beta\sin\gamma & \cos\alpha\cos\gamma - \sin\alpha\sin\beta\sin\gamma & \sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix} \tag{2.2}$$

The rigid body transformation shown in Equation 2.1 can be reorganized into a homogeneous form:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}. \tag{2.3}$$

2.1.2 Perspective Projection

The perspective projection (shown in Figure 2.2) is based on an ideal pinhole model that transforms $P_C(X_C, Y_C, Z_C)$ in camera coordinate to $P_I(X_I, Y_I)$ in image sensor coordinate.

The image position of target P_C in sensor coordinate is denoted as P_I, and its coordinates can be derived as

$$X_I = -d\frac{X_C}{Z_C} \tag{2.4}$$

$$Y_I = -d\frac{Y_C}{Z_C} \tag{2.5}$$

where d is the distance from the lens optical center to the image sensor plane. Equation 2.4 and Equation 2.5 can be further expressed in a homogeneous form:

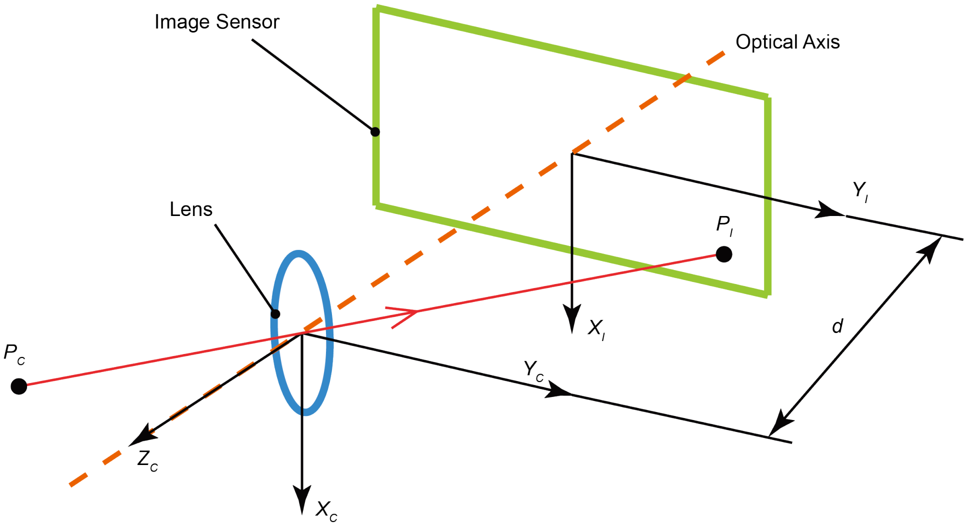


Figure 2.2: Perspective projection of a point from camera coordinate to image coordinate

$$\begin{bmatrix} X_I \\ Y_I \\ 1 \end{bmatrix} = \frac{1}{Z_C} \begin{bmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix}. \tag{2.6}$$

2.1.3 Lens Distortion

The perspective projection presented above is an approximation of real optical projection based on the assumption that the lens is free of optical distortion. However, this assumption is not true in real lenses. The perspective projection is a basis that is extended with some corrections for systematically distorted image coordinate [19]. The distorted position $P_D(X_D, Y_D)$ in image sensor coordinate is expressed as

$$\begin{aligned} X_D &= X_I + \delta_X\left(X_I, Y_I\right) \\ Y_D &= Y_I + \delta_Y\left(X_I, Y_I\right) \end{aligned} \tag{2.7}$$

where $\delta_X\left(X_I, Y_I\right)$ and $\delta_Y\left(X_I, Y_I\right)$ are the optical distortion terms in X_I and Y_I directions, respectively.

Typically, lens distortion is categorized into three types based on its physical cause: curvature distortion, decentering distortion and thin prism distortion [17].

- **Curvature Distortion**: curvature distortion is caused by flawed radial curvature of the lens elements, generating an inward or outward displacement of a given image point from its ideal location [14]. The curvature distortion of a perfectly centered lens has only radial components, and its expression in polar coordinate is governed by:

$$\delta_\rho = K_1\rho^3 + K_2\rho^5 + K_3\rho^7 + \dots \tag{2.8}$$

where ρ is the radial distance from the principal point (u_0, v_0) of the image plane and K_1, K_2,

K_3 ... are the coefficients of radial distortion [13][17]. With positive distortion coefficients $K_i\,(i=1,2,\ldots)$, the distorted image shows a pincushion shape; on the other hand, a barrel shape distortion is introduced by negative K_i (shown in Figure 2.3) [14].

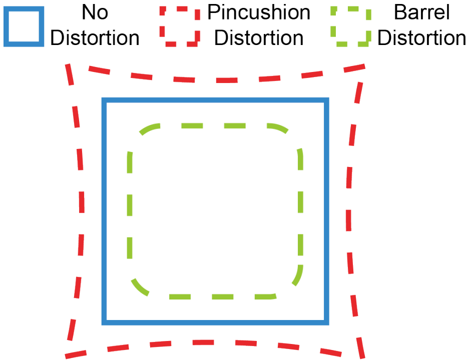


Figure 2.3: Illustration of radial distortion adapted from [14]

At each image point represented by polar coordinate (ρ,ϕ), the image point can also be expressed in terms of Cartesian coordinate (X_I,Y_I):

$$\begin{array}{l} X_I=\rho\cos\phi \\ Y_I=\rho\sin\phi \end{array}. \tag{2.9}$$

Therefore, the amount of curvature distortion in Cartesian image coordinate is derived as

$$\begin{array}{l} \delta_{X,r}(X_I,Y_I)=K_1X_I\left(X_I^2+Y_I^2\right)+K_2X_I\left(X_I^2+Y_I^2\right)^2+K_3X_I\left(X_I^2+Y_I^2\right)^3+\ldots \\ \delta_{Y,r}(X_I,Y_I)=K_1Y_I\left(X_I^2+Y_I^2\right)+K_2Y_I\left(X_I^2+Y_I^2\right)^2+K_3Y_I\left(X_I^2+Y_I^2\right)^3+\ldots \end{array} \tag{2.10}$$

- **Decentering Distortion**: Actual optical systems are subject to various degrees of decentering, that is, the optical center of lens elements are not strictly collinear [16]. This distortion has both radial and tangential components, which is described in polar coordinate by :

$$\begin{array}{l} \delta_{\rho d}=3(j_1\rho^2+j_2\rho^4+\ldots)\sin(\phi-\phi_0) \\ \delta_{td}=(j_1\rho^2+j_2\rho^4+\ldots)\cos(\phi-\phi_0) \end{array} \tag{2.11}$$

where $\delta_{\rho d}$ and δ_{td} are the radial and tangential component of decentering distortion respectively, ϕ_0 is the angle between the positive X_I axis and a line of reference where the maximum tangential distortion happens [16][17].

Similarly, the amount of decentering distortion can be expressed in Cartesian coordinate (X_I,Y_I) in terms of $\delta_{\rho d}$ and δ_{td}:

$$\begin{pmatrix} \delta_{X,d} \\ \delta_{Y,d} \end{pmatrix}=\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} \delta_{\rho d} \\ \delta_{td} \end{pmatrix} \tag{2.12}$$

Notice that $\cos\phi=X_I/\rho$ and $\sin\phi=Y_I/\rho$. By denoting $P_1=-j_1\sin\phi_0$ and $P_2=-j_2\cos\phi_0$, it yields

$$\begin{array}{l}\delta_{X,d}\left(X_I,Y_I\right)=P_1\left(3X_I^2+Y_I^2\right)+2P_2X_IY_I+O[(X_I,Y_I)^4]\\ \delta_{Y,d}\left(X_I,Y_I\right)=P_2\left(X_I^2+3Y_I^2\right)+2P_1X_IY_I+O[(X_I,Y_I)^4].\end{array}\tag{2.13}$$

- **Thin Prism Distortion**: thin prism distortion arises from imperfection in lens design and manufacturing as well as camera assembly. This type of distortion can be modeled by the adjunction of a thin prism to the optical system, causing additional amounts of radial and tangential distortions [16][20]. The thin prism distortion is described in polar coordinate as

  $$\begin{array}{l}\delta_{\rho p}=(i_1\rho^2+i_2\rho^4+\ldots)\sin(\phi-\phi_1)\\ \delta_{tp}=(i_1\rho^2+i_2\rho^4+\ldots)\cos(\phi-\phi_1)\end{array}\tag{2.14}$$

  where $\delta_{\rho p}$ and δ_{tp} are the radial and tangential component of thin prism distortion respectively and ϕ_1 is the angle between the positive X_I axis and the axis of maximum tangential distortion. Denote that $S_1=-i_1$ and $S_2=i_1\cos\phi_1$, the thin prism distortion along X_I and Y_I axis is derived similar to the case of decentering distortion, and it is expressed by

  $$\begin{array}{l}\delta_{X,p}\left(X_I,Y_I\right)=S_1\left(X_I^2+Y_I^2\right)+O[(X_I,Y_I)^4]\\ \delta_{Y,p}\left(X_I,Y_I\right)=S_2\left(X_I^2+Y_I^2\right)+O[(X_I,Y_I)^4]\end{array}.\tag{2.15}$$

In summary, the total distortion equals to the summation of radial distortion, decentering distortion and thin prism distortion:

$$\begin{array}{l}\delta_X\left(X_I,Y_I\right)=\delta_{X,r}\left(X_I,Y_I\right)+\delta_{X,d}\left(X_I,Y_I\right)+\delta_{X,p}\left(X_I,Y_I\right)\\ \delta_Y\left(X_I,Y_I\right)=\delta_{Y,r}\left(X_I,Y_I\right)+\delta_{Y,d}\left(X_I,Y_I\right)+\delta_{Y,p}\left(X_I,Y_I\right)\end{array}.\tag{2.16}$$

2.1.4 Image Digitization

The output of a digital image sensor is a digitized array using pixels to represent location information (Figure 2.4). Due to the manufacturing error of image sensor, the pixel array of the senor is not perfectly square. Denoting the position of a distorted image point as (X_D,Y_D), the location of this point in a digitized pixel coordinate is described as

$$u=u_0+S_xX_D-\frac{S_xY_D}{\tan\theta}\tag{2.17}$$

$$v=v_0+S_y\frac{Y_D}{\sin\theta}\tag{2.18}$$

where θ is the skew angle between u and v axis which can be considered as 90° for most of image sensors [19][18][21], (u_0,v_0) is the principal point where optical axis intersects with image sensor plane, and S_x and S_y are the scaling factor in u and v axis, respectively.

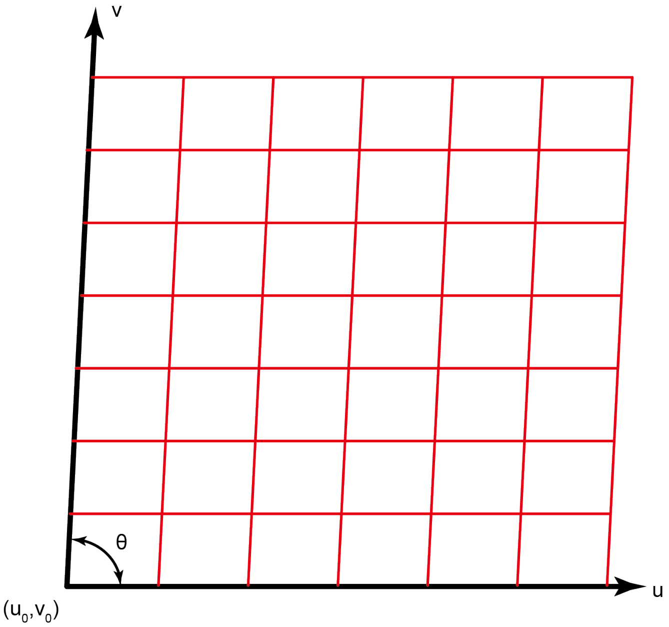


Figure 2.4: Illustration of the skewness error of image sensor

2.2 ROI-based Image Processing

ROI-based image processing is the fundamental idea which enables the system to achieve a much higher position sampling frequency than any other available commercial product and design. A region of interest (ROI) is a rectangular window with sides parallel to the image sensor frame which defines the important area within the overall image (shown in Figure 2.5). By reading small regions around each target image rather than the full frame data of sensor, the frame rate and image processing speed are vastly increased.

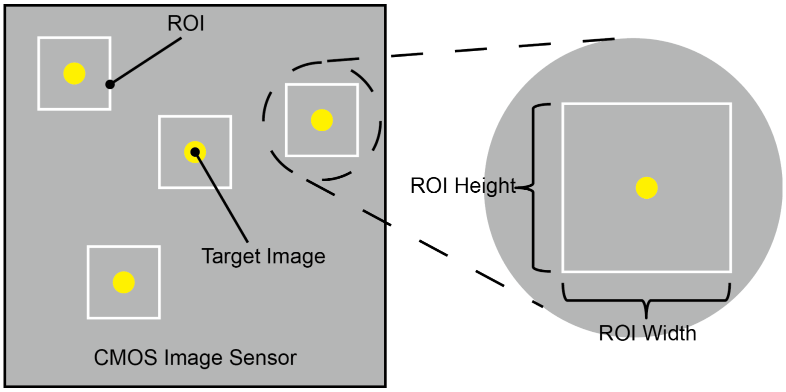


Figure 2.5: Illustration of region of interest

To realize the image processing based on ROIs, several supporting algorithms are required. First, in order to sensing the target position using ROIs, the initial positions of each target in pixel coordinate need to be determined so that the initial ROIs can be created. Second, sub-pixel target position interpolation method is required to achieve high accuracy of target position in pixel coordinate. Finally, target position in pixel coordinate changes due to its motion in world coordinate, therefore the ROI position needs to be updated to ensure that the target image is contained in the ROI. In general,the ROI-based image processing is divided into three parts: initial target identification, sub-pixel target position interpolation and ROI position update.

2.2.1 Initial Target Identification

In order to measure the target positions based on ROI, a full image frame is read out in the first stage, and then the targets are identified from the background. Once the initial target positions are obtained, initial ROIs are built around these targets.

One method to separate the targets from their surroundings is background subtraction [11]. This method subtracts a reference background image frame from the initial full image frame. By this means, an image which contains the difference between targets and background is obtained, therefore the targets are emphasized. Considering that the target images are much brighter than the background, it is easy to detect the bright pixels in the full image and then create the ROIs surround them. This method is computationally cheap and independent of the shape of target image, thus easy to be implemented in real-time embedded system.

2.2.2 Sub-pixel Target Position Interpolation

A key area of photogrammetric measurement is sub-pixel target position interpolation. The subsequent processing in high accuracy 3D position measurement relies on the results of it. There have been a number of theoretical studies as well as practical tests conducted in order to quantify the best performance possible for sub-pixel interpolation algorithms. Shortis reviewed the previous research and investigations on sub-pixel interpolation methods based on digital image sensor [22]. There is a general agreement that precision of 0.01 pixels are theoretically possible, and some practical tests have realized this level of precision. In direct triangulation systems, there appears to be a good correlation between the prediction by simulation and the practical reality [22]. The sub-pixel interpolation precision better than 0.01 pixels is achievable when a high signal-to-noise ratio (SNR) of target image is obtained.

There are two types of sub-pixel target position interpolation methods: threshold-based method and intensity-based method. The threshold-based method, such as binary centroid and ellipse fitting, converts the raw digital image to a binary image by using certain threshold value and then calculates the sub-pixel target position based on the binary image. On the other hand, intensity-based method does not generate a binary image but directly utilizes every pixel value in the interpolation.

Shortis investigated a number of techniques for sub-pixel interpolation [23]. The performance of threshold-based methods and intensity-based methods is compared. Their results show that intensity-based method achieves higher accuracy than threshold-based method in different scenarios of image noise level, saturation level, DC offset and size of target image. Therefore, the intensity-based method is selected. Three candidates of intensity-based method are discussed here: centroid method, squared-centroid method and 2D Gaussian fitting.

Centroid Method

Calculating the centroid of an image is analogous to finding the center of mass of an object. The centroid of an image is the intensity weighted sum of points that constitute the image. The u and v components of the centroid with respect to the ROI origin are expressed as

$$\begin{aligned} u_{center} &= \frac{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I(u,v)u}{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I(u,v)} \\ v_{center} &= \frac{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I(u,v)v}{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I(u,v)} \end{aligned} \tag{2.19}$$

where $I(u,v)$ is the pixel value at position (u,v) within the ROI boundary.

Squared-Centroid Method

The squared-centroid method is very similar to centroid method but uses the square of pixel value as the weight to determine the target position. The squared-centroid method is expressed as

$$\begin{aligned} u_{center} &= \frac{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I^2(u,v)u}{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I^2(u,v)} \\ v_{center} &= \frac{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I^2(u,v)v}{\sum\limits_{v=1}^{m}\sum\limits_{u=1}^{n} I^2(u,v)} \end{aligned} . \tag{2.20}$$

The squared-centroid method emphasizes the main body of target image where the pixel values are much higher than the background [22]. As a consequence, peripheral pixels in the background of ROI are less influential.

2D Gaussian Fitting

The Gaussian fitting method is based on the assumption that the intensity distribution of a point light source is well approximated by a 2D Gaussian distribution (shown in Figure 2.6). It takes the array of pixel values to approximate a 2D Gaussian distribution in the form of

$$I = \frac{K}{2\pi\delta_u\delta_v\sqrt{1-\rho^2}} e^{\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{u-u_{center}}{\delta_u}\right)^2 - 2\rho\left(\frac{u-u_{center}}{\delta_u}\right)\left(\frac{v-v_{center}}{\delta_v}\right) + \left(\frac{v-v_{center}}{\delta_v}\right)^2\right]\right\}} \tag{2.21}$$

where (u_{center}, v_{center}) is the distribution center, δ_u and δ_v are the standard deviations, ρ is the correlation coefficient, and K is the scaling factor. Nonlinear least-square estimator is required to solve the target center (u_{center}, v_{center}).

Comparing these three intensity-based methods, the centroid and squared-centroid method are more favorable than 2D Gaussian fitting in real-time application. Though 2D Gaussian fitting method is able to achieve high accuracy, its computation cost is much higher than the other two methods since the nonlinear least-square estimator requires iteration to solve the fitting problem. On the other hand, the centroid method and squared-centroid method only require multiplications and accumulations that

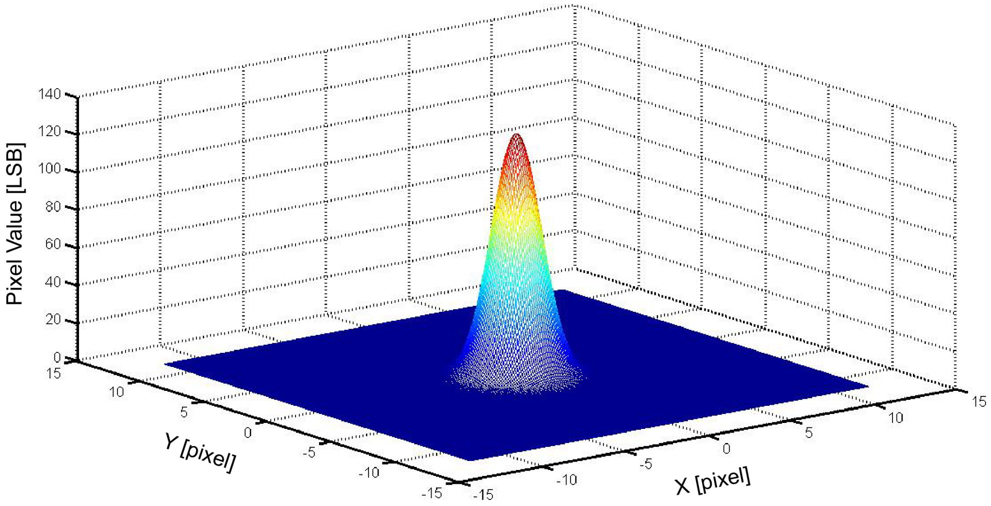


Figure 2.6: 2D Gaussian distribution

can be easily realized by either hardware or software in a real-time embedded system. Chen and Clark compared the use of the squared-centroid method with the centroid method and found small, insignificant differences between them [24]. Shortis also reported that these two types of method produce very similar results [22]. Their results show that both methods are very robust and produce very close results in the experiments. Therefore, both methods are implemented in the system, and their performance are further compared by experiment in Chapter 5.

2.2.3 ROI Update

Tracking the motion of the target from one image frame to another requires that the ROI position must be able to follow the target's movement. Two parameters need to be determined in ROI update: ROI size and ROI position.

When considering the ROI size, there is a tradeoff between ROI size and the allowed target moving speed. Large ROI size allows high moving speed of target but reduces the position sampling frequency since more pixels are read out and processed in each frame. In the contrary, small ROI size increases the position sampling frequency but limits the allowed speed of target motion because the target might move outside the ROI within one frame period.

The update of ROI position can be achieved by placing the next ROI around the center of the target image in the current image frame. As shown in Figure 2.7, the ROI position is updated between frames. The maximum movement S_{max} of target image between frames is determined by the maximum allowed target speed V_{max} and frame period T_{frame}. In order to guarantee that the target image is contained in the ROI boundary, the following constraint can be derived from Figure 2.7:

$$\frac{W}{2} \geq S_{\max} + \frac{d}{2} \tag{2.22}$$

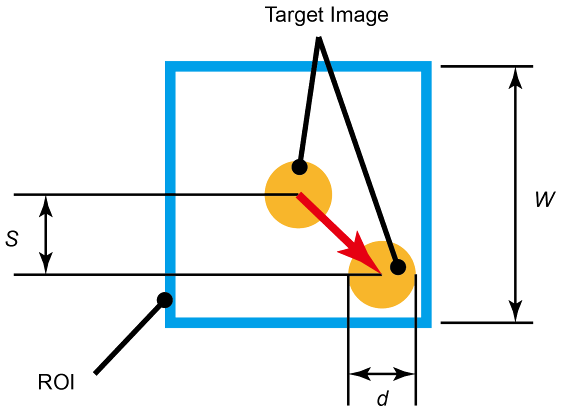


Figure 2.7: Schematic of ROI update adapted from [11]

$$W \geq 2S_{\max} + d = 2V_{\max}T_{frame} + d. \tag{2.23}$$

Assuming that there are N targets to be tracked, the pixel period is T_{pixel} and the exposure time is $T_{exposure}$, ideally it yields

$$T_{frame} = NW^2 T_{pixel} + T_{exposure}. \tag{2.24}$$

Combining the two equations above, the ROI width must meet the constraint that

$$2V_{\max}NT_{pixel}W^2 - W + d + 2V_{\max}T_{exposure} \leq 0. \tag{2.25}$$

Given the desired frame rate and maximum allowed target speed, the required ROI dimension is obtained.

2.3 3D Position Reconstruction

Shown in Figure 2.8, one target located at P_W in world coordinate generates two images at P_{I0} and P_{I1} in two image sensors, respectively. The position information from single camera, such as P_{I0}, is not enough to reconstruct the 3D position of P_W because any target located on the line $P_W P_{I0}$ is able to generate its image at P_{I0} in image sensor 0. That is to say, the depth information is not deterministic based on position information from single camera. Therefore, two cameras are required in order to reconstruct the 3D position in world coordinate. The 3D position reconstruction is the inverse process of imaging model. Assuming that the intrinsic and extrinsic camera parameters are known, the 3D position of the target is able to be retrieved by a linear transformation [15] combining a nonlinear correction of lens distortions.

Notice that the imaging model is linear if the lens distortion terms are removed. From Section 2.1.3, it can be inferred that nonlinear iteration is required to solve the distortion-free position $P_I(X_I, Y_I)$ from the distorted position $P_D(X_D, Y_D)$. This iterative nonlinear solver will generate computation problems when the distortion correction algorithm is implemented in a real-time hardware. To correct the lens distortion using an explicit expression, the inverse mapping is introduced (shown in Equation 2.26)

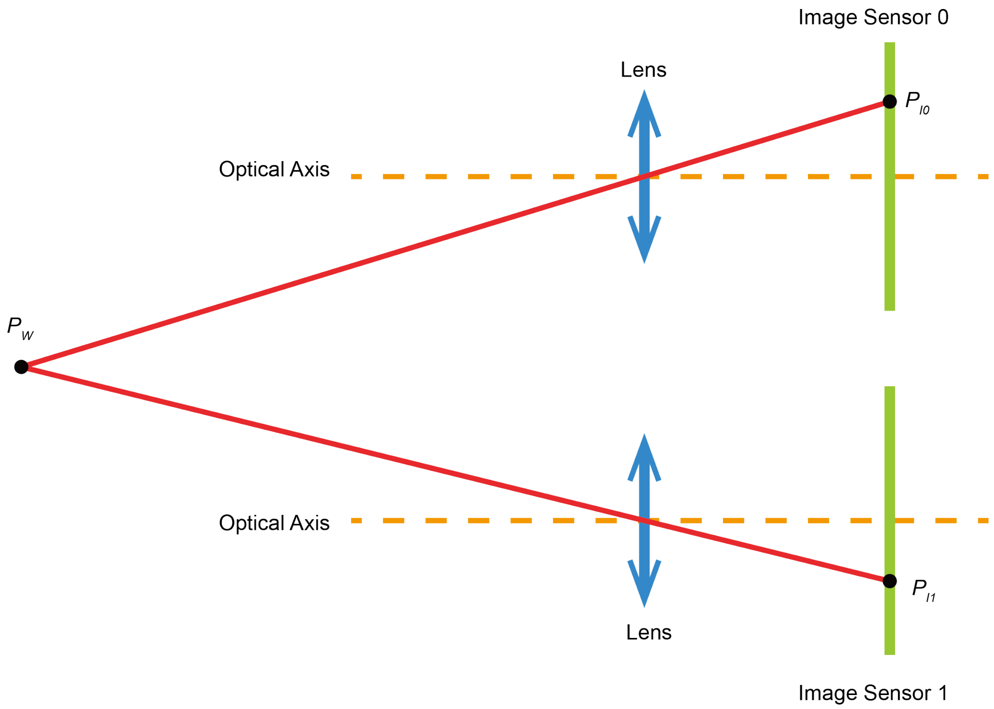


Figure 2.8: Schematic of stereo-vision based on two image sensors

[18][17][19].

$$\begin{array}{l} \delta_X = \delta_{X,r}\left(X_D, Y_D\right) + \delta_{X,d}\left(X_D, Y_D\right) + \delta_{X,p}\left(X_D, Y_D\right) \\ \delta_Y = \delta_{Y,r}\left(X_D, Y_D\right) + \delta_{Y,d}\left(X_D, Y_D\right) + \delta_{Y,p}\left(X_D, Y_D\right) \end{array}. \tag{2.26}$$

Different from Equation 2.16, Equation 2.26 replaces $P_I(X_I, Y_I)$ by using the distorted position $P_D(X_D, Y_D)$ to represent the lens distortion terms δ_X and δ_Y. This replacement is reasonable because the distortion terms δ_X and δ_Y calculated from the ideal position P_I is approximately equal to that calculated from the distorted position P_D [17][19]. By this means, the correction of lens distortion is explicitly expressed and can be computed straightforward (shown in Equation 2.27 and Equation 2.28).

$$\begin{cases} X_D = \frac{(u_d - u_0 + S_x Y_D \cot\theta)}{S_x} \\ Y_D = \frac{(v_d - v_0)\sin\theta}{S_y} \end{cases} \tag{2.27}$$

$$\begin{cases} X_I = X_D - \delta_X\left(X_D, Y_D\right) \\ Y_I = Y_D - \delta_Y\left(X_D, Y_D\right) \end{cases} \tag{2.28}$$

From Section 2.1, it can be derived that the ideal target position in image sensor coordinate (X_I, Y_I) is

$$X_I = -d\frac{r_{11}(X_W - T_x) + r_{12}(Y_W - T_y) + r_{13}(Z_W - T_z)}{r_{31}(X_W - T_x) + r_{32}(Y_W - T_y) + r_{33}(Z_W - T_z)} \tag{2.29}$$

$$Y_I = -d\frac{r_{21}(X_W - T_x) + r_{22}(Y_W - T_y) + r_{23}(Z_W - T_z)}{r_{31}(X_W - T_x) + r_{32}(Y_W - T_y) + r_{33}(Z_W - T_z)} \tag{2.30}$$

where r_{ij} ($i = 1,2,3$ and $j = 1,2,3$) is the component in rotation matrix R.

Equation 2.29 and Equation 2.30 are reorganized by introducing the intermediate parameters $L_j(j = 1,2,...,11)$ [15]:

$$X_I = \frac{L_1X_W + L_2Y_W + L_3Z_W + L_4}{L_9X_W + L_{10}Y_W + L_{11}Z_W + 1} \tag{2.31}$$

$$Y_I = \frac{L_5X_W + L_6Y_W + L_7Z_W + L_8}{L_9X_W + L_{10}Y_W + L_{11}Z_W + 1} \tag{2.32}$$

where $D = -(r_{31}T_x + r_{32}T_y + r_{33}T_z)$,
$L_1 = -\frac{dr_{11}}{D}$,
$L_2 = -\frac{dr_{12}}{D}$,
$L_3 = -\frac{dr_{13}}{D}$,
$L_4 = \frac{d(r_{11}T_x+r_{12}T_y+r_{13}T_z)}{D}$,
$L_5 = -\frac{dr_{21}}{D}$,
$L_6 = -\frac{dr_{22}}{D}$,
$L_7 = -\frac{dr_{23}}{D}$,
$L_8 = \frac{d(r_{21}T_x+r_{22}T_y+r_{23}T_z)}{D}$,
$L_9 = \frac{r_{31}}{D}$,
$L_{10} = \frac{r_{32}}{D}$,
$L_{11} = \frac{r_{33}}{D}$.

It is clear that the parameters $L_j(j = 1,2,...,11)$ are fully determined by the intrinsic and extrinsic camera parameters. Denote that $(X_{I,l},Y_{I,l})$ and $(X_{I,r},Y_{I,r})$ are the ideal target positions in two cameras, respectively. Applying Equation 2.31 and Equation 2.32 to the two cameras yields

$$X_{I,l} = \frac{L_{1,l}X_W + L_{2,l}Y_W + L_{3,l}Z_W + L_{4,l}}{L_{9,l}X_W + L_{10,l}Y_W + L_{11,l}Z_W + 1} \tag{2.33}$$

$$Y_{I,l} = \frac{L_{5,l}X_W + L_{6,l}Y_W + L_{7,l}Z_W + L_{8,l}}{L_{9,l}X_W + L_{10,l}Y_W + L_{11,l}Z_W + 1} \tag{2.34}$$

$$X_{I,r} = \frac{L_{1,r}X_W + L_{2,r}Y_W + L_{3,r}Z_W + L_{4,r}}{L_{9,r}X_W + L_{10,r}Y_W + L_{11,r}Z_W + 1} \tag{2.35}$$

$$Y_{I,r} = \frac{L_{5,r}X_W + L_{6,r}Y_W + L_{7,r}Z_W + L_{8,r}}{L_{9,r}X_W + L_{10,r}Y_W + L_{11,r}Z_W + 1}. \tag{2.36}$$

The equations above can be further rearranged into a matrix format:

$$\begin{bmatrix} X_{I,l}L_{9,l} - L_{1,l} & X_{I,l}L_{10,l} - L_{2,l} & X_{I,l}L_{11,l} - L_{3,l} \\ Y_{I,l}L_{9,l} - L_{5,l} & Y_{I,l}L_{10,l} - L_{6,l} & Y_{I,l}L_{11,l} - L_{7,l} \\ X_{I,r}L_{9,r} - L_{1,r} & X_{I,r}L_{10,r} - L_{2,r} & X_{I,r}L_{11,r} - L_{3,r} \\ Y_{I,r}L_{9,r} - L_{5,r} & Y_{I,r}L_{10,r} - L_{6,r} & Y_{I,r}L_{11,r} - L_{7,r} \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \begin{bmatrix} L_{4,l} - X_{I,l} \\ L_{8,l} - Y_{I,l} \\ L_{4,r} - X_{I,r} \\ L_{8,r} - Y_{I,r} \end{bmatrix}. \tag{2.37}$$

By solving Equation 2.37, the 3D position of the target is retrieved from its 2D positions in image sensor coordinate.

2.4 Summary

In this chapter, the theory behind vision-based position sensing system is presented. An imaging model is built which describes the coordinate transformation from the 3D world coordinate to the 2D pixel coordinate. This transformation is divided into four steps: rigid body transformation, perspective projection, lens distortion and image digitization. Section 2.2 covers the supporting algorithms of ROI-based image processing. Background subtraction is utilized to identify the targets from image background and find their initial positions. Based on the achievable accuracy and computation cost, the centroid and square-centroid method are implemented in the system to calculate the sub-pixel target position in 2D pixel coordinate. The 3D position reconstruction algorithm based two-camera stereo-vision system is given based on a linear transformation combined with a correction of nonlinear lens distortion.

Chapter 3

System Hardware Design

System hardware, including electronics, optical and mechanical design, is the key to achieve the design goals of high position sampling frequency and high position accuracy. This chapter presents the design of a customized hardware prototype of stereo-vision system. Section 3.1 presents the architecture of system and gives the design strategies. In Section 3.2, the detailed electronics design is discussed. Section 3.3 describes the optical and mechanical design, including optical device selection, target design, and camera body design.

3.1 Hardware Design Overview

The goals of designing a customized hardware are to achieve high position sampling frequency and high position accuracy required by this system. The system is designed and implemented based on a two-camera architecture (shown in Figure 3.1). The camera modules and image processing unit are connected by a camera interface, including a data channel and a camera control channel. In image processing unit, the 3D position of the target being tracked is computed and sent to the host PC. To realize the design objectives, the design strategies of system hardware are discussed in this section.

High-Performance Image Acquisition

Image acquisition is the main bottleneck that limits the performance of a vision-based position sensing system. High-performance image acquisition system features high speed, high resolution image data and high image SNR. Fast image readout provides the foundation to achieve high position sampling frequency. Sampling the analog output of the image sensor with high resolution and low noise improves the achievable position accuracy.

High-Speed Camera Interface

The camera interface between camera modules and image processing unit provides functionalities of data transmission as well as camera control. High-speed and high-resolution image data means massive data transmission, demanding high-speed data throughput from camera module to image processing unit with low latency.

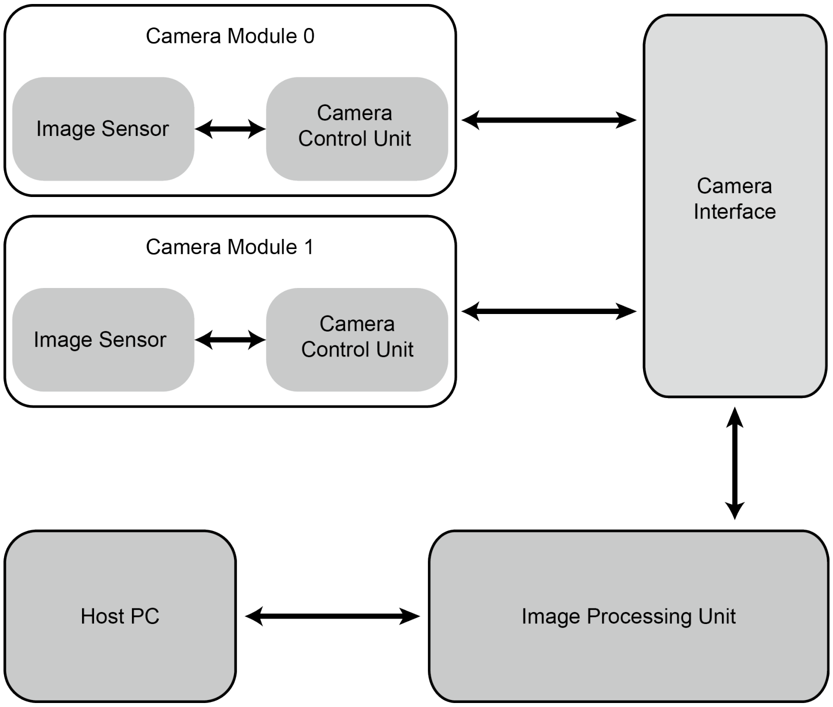


Figure 3.1: Overall system architecture

High-Speed Image Processing

High-speed and high-resolution image data acquisition not only requires a high-speed camera interface but also challenges the capability of image processing. In order to obtain the 3D position of target, the image processing involves image noise correction, sub-pixel target position interpolation and 3D position reconstruction. Therefore, high computation capability of imaging processing hardware is required.

Solid Camera Body and High-Quality Optics

The system is operated in a semi-controlled environment, where the ambient temperature is not fully controlled. Since the 3D position reconstruction relies on model parameters, the temperature fluctuation will introduce thermal deformation of the camera body structure, causing a change in some model parameters. In order to achieve high position accuracy, a thermal-stable camera body is required. At the same time, external vibrations cause the deformation in camera body, which can introduce error in 3D position reconstruction. The designed camera body must be solid to resist the deformation caused by external vibrations as well as temperature fluctuations. Optical components, such as lens, should be carefully selected so as to limit the optical distortion and obtain high spatial resolution.

3.2 High-Performance Electronics Design

Figure 3.2 shows the electronics architecture of camera module. The camera control unit plays the role of a brain that controls the operation of camera module. The image sensor outputs analog signals of pixel value under the timing control of camera control unit. The analog signals are conditioned and

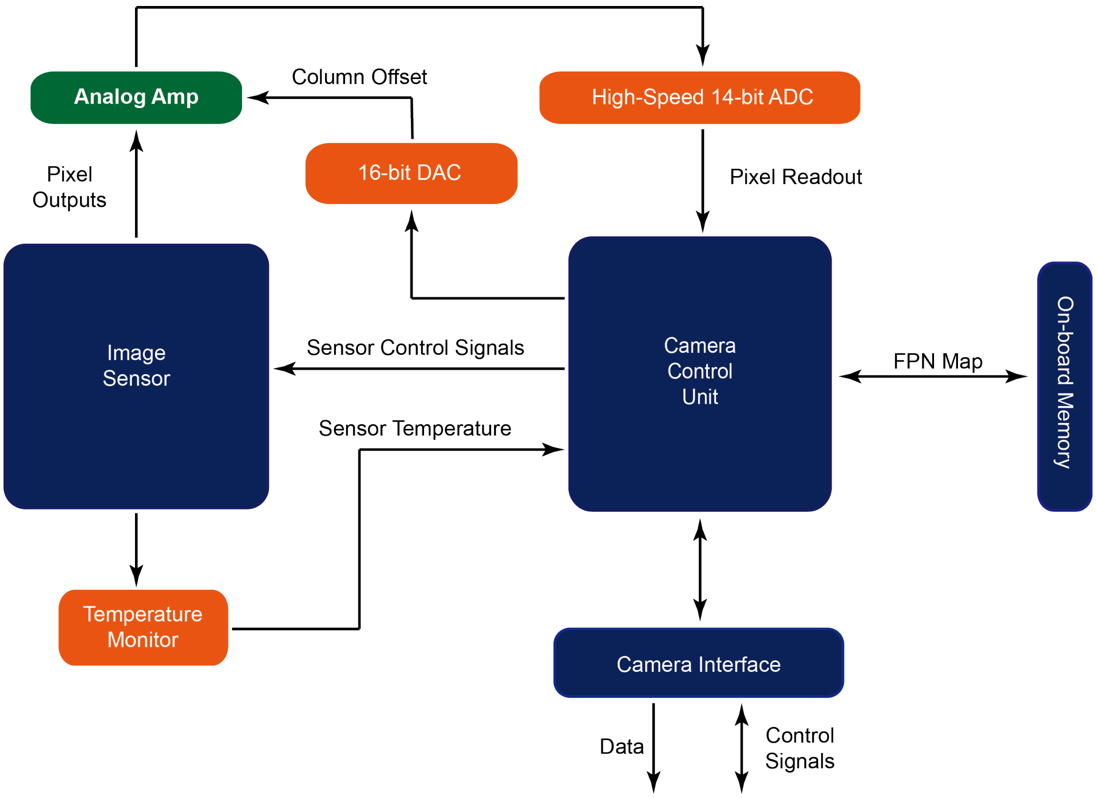


Figure 3.2: Electronic architecture of camera module

amplified, and then sampled by a high-speed analog-to-digital converter (ADC). On-board memory is integrated in the camera module to store the image noise correction data. Data and control information are transmitted through a high-speed camera interface. The detailed designs for each part are described in this section.

3.2.1 Image Sensor

In this section, the selection criteria of image sensor are discussed based on sensor type, pixel rate, resolution, shuttering type and etc..

CCD image sensor and CMOS image sensor are two candidates for the vision applications. CCD image sensor was the first product in vision applications and has dominated in this field for many years. In a CCD image sensor, every pixel's charge is transferred through limited numbers of output nodes to be converted to voltage, buffered, and sent off-chip as an analog signal. All pixel area can be devoted to light capture in a CCD image sensor [4]. On the other hand, in a CMOS image sensor, both photodiode and readout amplifier are integrated within each pixel. The voltage or current output from the pixel is read out over X-Y wires instead of using a shift register. The use of CMOS technology permits ready integration of on-chip timing and control electronics, as well as signal conditioning electronics. ADC can be integrated on chip. This highly integrated imaging system is referred to as a camera-on-a-chip, and represents a second generation solid state image sensor technology [25]. The structure of CCD and CMOS image sensor is shown in Figure 3.3.

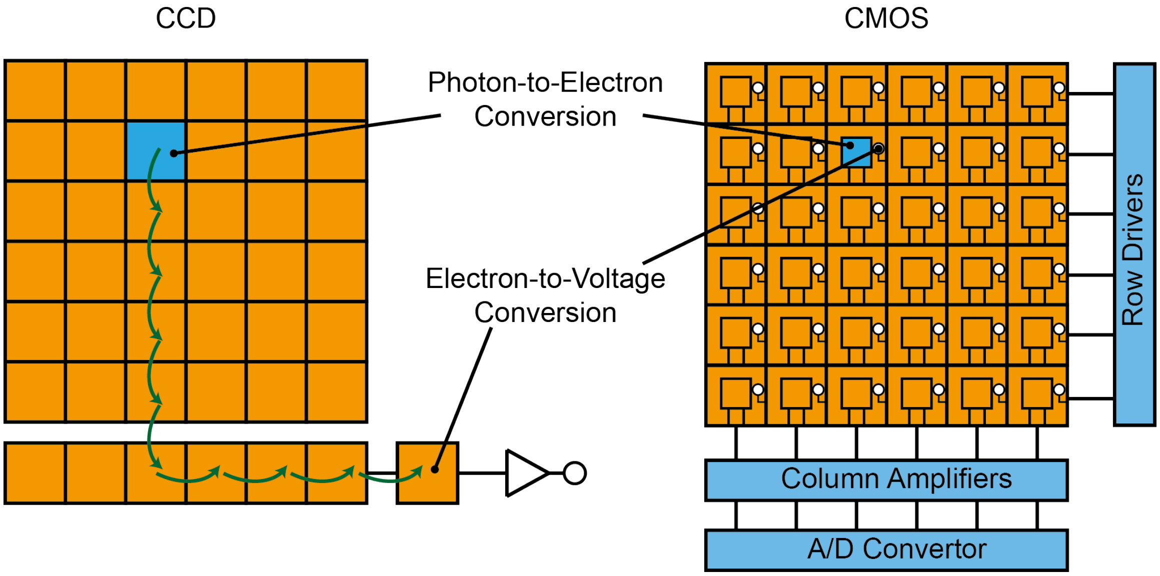


Figure 3.3: Structure of CCD and CMOS image sensor adapted from [26]

When selecting an image sensor, multiple issues must be taken into consideration, relating to both the features of CCD and CMOS devices and the requirements of application. Nixon O. presented the idea that the application determines the choice of image sensor in [27], where criteria of application-oriented image sensor selection are given. Dave Litwiller published a series of paper related to CCD and CMOS imaging technologies [4][26][28], where the development, features and suitable applications of both imaging technologies are reviewed and discussed. The features of CCD and CMOS image sensor are summarized and compared in Table 3.1

Table 3.1: Comparison of CCD and CMOS image sensor, data from [28]

Feature	CCD	CMOS
ROI Readout	No	Yes
Responsivity	Moderate	High
Dynamic Range	Large	Moderate
Uniformity	High	Moderate
Power Consumption	High	Low

One of the most important design goals in this research is to achieve high position sampling frequency (over 10kHz). When selecting the image sensor, the speed features weight more than the others. It is clear that CMOS image sensor has absolute speed advantages over CCD image sensor. ROI readout plays the important role in achieving high speed image acquisition. CMOS image sensor has the ability of direct pixel access that allows to read out a portion of the image sensor. When implementing the CMOS image sensor in a high-speed vision system, only the small regions around each target image are read out and processed, avoiding the time waste on the full-frame readout. ROI readout not only reduces the readout time, but also reduces the computation time during image processing and accelerates the processing speed. Meanwhile, when the targets are moving from one exposure to another, the position and size of ROIs can be dynamically adjusted. The responsivity is the amount of signals that an image sensor delivers per unit of input optical energy. CMOS image sensors have higher signal response to the same light levels because amplifiers are placed at every pixel [28]. High responsivity to illumination of the CMOS image sensor allows short exposure time. Low power consumption simplifies the power distribution system (PDS) design and limits the thermal problems. Conclusively, CMOS image sensor is the only reasonable technology suitable for the high-speed vision applications.

Assuming that the size of ROI is 20 pixels by 20 pixels and the number of ROI is 6 for tracking 6 targets, the readout time for the pixels in 6 ROIs is calculated as

$$t_{\text{readout}} = 6 \times 20 \times 20 \times T_{\text{pixel}} < \frac{1}{10\text{kHz}}. \tag{3.1}$$

where t_{readout} is the readout time and T_{pixel} is the pixel output period. In order to achieve 10kHz position sampling frequency, the pixel rate should be higher than 24MHz estimated from Equation 3.1.

In high-speed vision applications, the vision system needs a way to freeze the motion of targets. In order to freeze the motion of targets and deliver high quality images in high-speed vision applications, image sensors require high-speed shuttering ability. Traditional industry solutions rely on CCD image sensors using interline transfer architecture to deliver this functionality [29]. Recent improvements in CMOS image sensor design have enabled CMOS technology to achieve global shuttering necessary to meet high-speed image acquisition requirement [28]. Global shuttering begins and ends exposure for all pixels simultaneously. Not all CMOS image sensors are capable of global shuttering. Simple pixel design, typically with three transistors structure, offers a rolling shutter. Because this type of sensor can only capture the image row by row and each row represents the target at a different point in time, a blurred image is obtained when capturing fast moving targets [29]. Therefore, more sophisticated CMOS image sensors with global shuttering, typically based on five or six transistors design, are the only candidates.

The resolution of image sensor determines the achievable position precision and accuracy. In this research, a position accuracy of 10μm is desired in a measurement volume of 1m by 1m. Generally speaking, the intensity-based sub-pixel target position interpolation methods can achieve 0.01 pixel accuracy in pixel coordinate [23]. The required resolution of the image sensor is expressed as

$$\left(\frac{1\text{m}}{10\mu\text{m}} \times 0.01\text{pixel}\right) \times \left(\frac{1\text{m}}{10\mu\text{m}} \times 0.01\text{pixel}\right) = 1000\text{pixel} \times 1000\text{pixel} \tag{3.2}$$

In summary, the selected image sensor must satisfy the following basic criteria:

- CMOS type image sensor
- ROI readout support
- over 24MHz pixel rate
- global shuttering
- over 1000 pixels by 1000 pixels resolution

Based on these criteria, the LUPA-4000 CMOS image sensor (Figure 3.4) from Cypress Semiconductor is selected. The specifications of this image sensor are shown in Table 3.2.

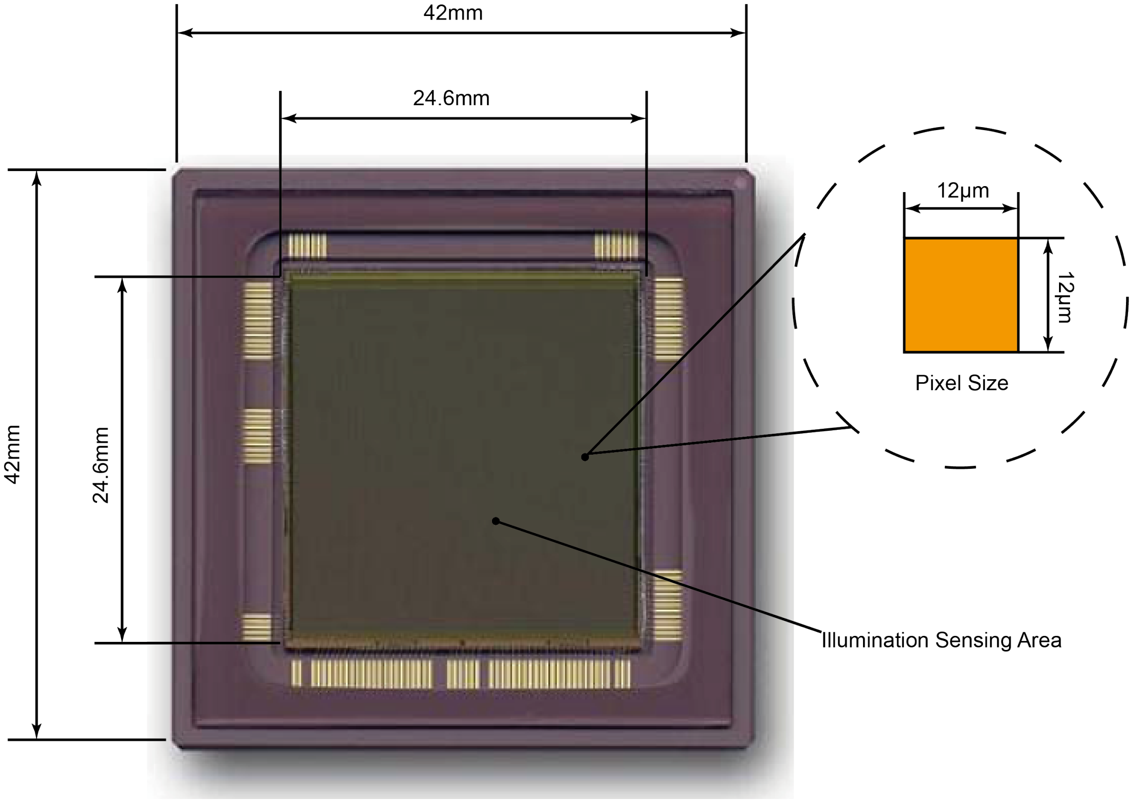


Figure 3.4: LUPA-4000 CMOS image sensor from Cypress Semiconductor

Table 3.2: Specifications of LUPA-4000 image sensor, data from [30]

Parameter	Specification
Sensor Type	CMOS
Sensor Resolution	2048×2048
Pixel Rate	66MHz
ROI Readout Support	Yes
Shutter	Global Shuttering
Dynamic Range	66dB (2000:1)
Power Dissipation	<200mW

According to the datasheet of LUPA-4000 [30], the theoretical frame readout time is calculated as

$$t_{\text{readout}} = \text{FOT} + N_{\text{ROI}} \times N_{\text{col}} \times \left(\text{ROT} + N_{\text{row}} \times T_{\text{pixel}}\right) \tag{3.3}$$

where FOT is the frame overhead time equal to 5μs, ROT is the row overhead time equal to 200ns, T_{pixel} is the pixel period equal to 15.15ns, N_{ROI} is the number of ROI, N_{col} is the column number of the ROI, and N_{row} is the row number of the ROI.

3.2.2 Camera Control Unit

Camera control unit controls and monitors the operation of the camera module. DSP and FPGA are considered as the candidates.

DSP is a specialized processor that is designed specifically for operating complex mathematically orientated calculations very swiftly. Compared to the Von Neumann architecture in the general purpose microprocessor, Harvard architecture or extended Harvard architecture is used in the DSP where instructions and data are stored in different caches and have their dedicated bus [31]. Faster instruction execution is achieved in the DSP. The strong computation capability of DSP meets the requirement of high-speed image processing. However, the processing speed of DSP is limited by its serial instruction stream when the sampling frequency reaches the level of mega-hertz.

FPGA is the best choice for the camera control unit. FPGA contains many programmable logic elements in gate level that can be combined to produce complex high level modules and be used to build very fast algorithmic blocks. In addition, advanced FPGA allows a soft or hard core processor embedded into the logic elements of FPGA. This allows the implementation of some algorithms using FPGA's programmable hardware and developing other parts in software. The FPGA solution is powerful because it combines a fast hardware-based approach with a flexible software-based approach. Further, the I/O resources in FPGA are configurable, providing convenience in camera electronics design.

Spartan-3A XCSD3400A FPGA from Xilinx is selected as the camera control unit. This FPGA is the most advanced in Extended Spartan-3A family, which provides 3400K logic gates resources, maximum 373Kbits distributed random-access memory (RAM), 8 digital clock managers (DCMs) and 469 single ended I/Os. Spartan-3A XCSD3400A has 126 multipliers/DSP48A blocks, which offers high computation capability. Spartan-3A XCSD3400A FPGA is able to work with two digital power supplies: 1.2V for core power and 3.3V for I/O and auxiliary power. Simple power supply allows a compact printed circuit board (PCB) design and potentially less power consumption. According to the analysis based on XPower Estimator from Xilinx, Spartan-3A XCSD3400A consumes 2.3W power based on the designed hardware description language (HDL) in worst case. As a result of low power consumption, the heat sink and other thermal conduction structure are no longer required.

3.2.3 Analog-to-Digital Electronics Design

The selected CMOS image sensor provides pixel analog outputs and digital outputs of pixel values. However, the on-chip ADC has low resolution (10-bit) and low sampling (10MHz). Apparently, the on-chip ADC simplifies the peripheral circuits design of image sensor but is not able to satisfy the design objectives. Therefore, an external analog-to-digital conversion circuit is required in order to achieve high-speed and low-noise image acquisition.

Because the even and odd pixel have separate output amplifiers in the selected image sensor, two identical analog-to-digital conversion circuits shown in Figure 3.5 are required for even and odd pixel, respectively. The analog outputs of image sensor go through analog signal conditioning circuit and then sampled by a high-speed ADC. The sensor analog output offsets for even and odd pixels are adjustable by digital-to-analog converters (DACs).

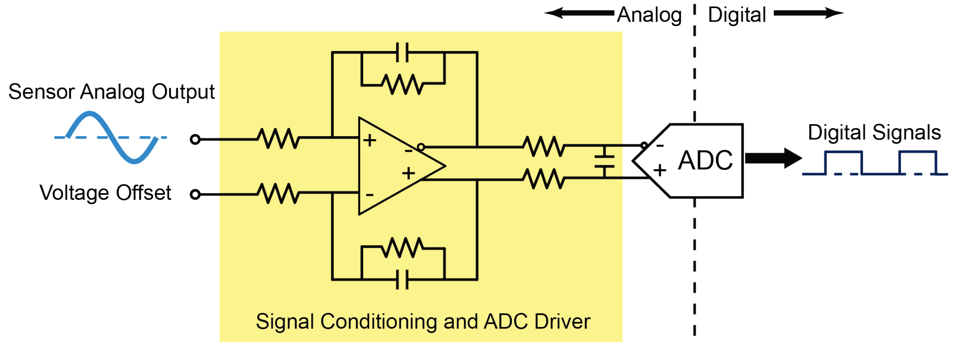


Figure 3.5: Analog-to-Digital conversion electronics architecture

ADC Selection

The sampling rate and resolution are two key specifications in the ADC selection. The sampling rate of the ADC should match the pixel rate of image sensor, and higher sampling frequency allows multiple samples in one pixel period so that multiple samples from one pixel period can be averaged to minimize ADC readout noise. High resolution of ADC reduces the quantization noise, increasing the quality of target image.

There is a tradeoff between ADC sampling rate and resolution. Generally speaking, fast ADCs come with low resolution, and high resolution ADCs come with low sampling rate. Therefore, when selecting an ADC, we should select an ADC with high resolution under the condition that it satisfies the sampling rate requirement.

The dynamic range of of the selected image sensor is 66dB (2000:1), therefore the minimum resolution of the ADC should be 11-bit. The nominal pixel rate of the selected image sensor is 66MHz which is achieved by using parallel analog output amplifiers on the image sensor under a pixel clock frequency of 33MHz. Here, multiple-sample in one pixel rate is desired so that a ADC sampling rate of over 66MHz is required. Considering the ADC with such a high sampling rate are generally based on pipeline architecture, fewer conversion latency cycle is favorable. In order to simplify the power supply design, the favorable ADC should be able to operate with single power supply of 3.3V and have low power consumption.

Based on the criteria mentioned above, LTC2299 from Linear Technology is selected. The specifications of LTC2299 is listed in Table 3.3. Since LTC2299 has two ADC channels in one package, a compact PCB design of ADC circuit is achieved.

Voltage Offset Control of Pixel Output

Analog outputs of even and odd pixels are presented in two parallel output amplifiers in the image sensor. The analog output ranges from 0.3V to 1.3V, and therefore the pixel output must be offset by a DC voltage level in order to meet the input span of the selected ADC. A 0.8V DC voltage is required theoretically to shift the pixel output to the range of -0.5V to +0.5V. Considering the imperfect manufacturing of the image sensor, the pixel offset voltage varies between even and odd pixels so that two dedicated DACs are used to provide the offset voltages to even and odd pixels, respectively.

Table 3.3: Electrical specifications of LTC2299, data from [32]

Specification	LTC2299
Resolution	14
Channel Number	2
Sampling Rate	80Msps
Input Span	$1\mathrm{V_{pp}}$
SNR	73dB @ 70MHz
INL	1.2LSB
Pipeline Delay	5 cycles
Power Supply	Single 3.3V
Power Consumption	444mW

Considering that this offset voltage is not required to change from pixel to pixel, a DAC with high dynamic performance is not required. However, since any noise at the DAC output influences the input of ADC, the selected DAC should provide low noise performance. The noise level of the selected DAC must be limited under one least significant bit (LSB) of the selected ADC. Therefore, the noise level of DAC should satisfy that

$$Noise_{DAC} < \frac{1\mathrm{V}}{2^{14}} = 61\mu\mathrm{V}. \tag{3.4}$$

Based on the requirements above, the DAC8411 from Texas Instruments is selected. DAC8411 is a 16-bit, low-power, low-noise DAC with 6-pin small package. The output noise is $3\mu V_{\mathrm{pp}}$ from 0.1Hz to 10Hz.

Because the selected DAC does not have dedicated input pin of reference voltage, the noise of DAC power supply directly couples to the output noise. As a result of the low supply current required by DAC8411, a precision voltage reference is used to provide the required power (Figure 3.6). The voltage reference LTC6655 from Linear Technology is selected, which features $0.67\mathrm{ppm_{RMS}}$ low voltage noise and $1\mathrm{ppm}/^{\circ}\mathrm{C}$ temperature coefficient.

Signal Conditioning and Analog Front End Design

The input of LTC2299 ADC requires differential input in a range of $1\mathrm{V_{pp}}$ with 1.5V common-mode voltage, but the analog outputs from image sensor are single-ended ranging from 0.3V to 1.3V. Besides shifting the pixel analog output by a DC offset, the analog outputs should be converted from single-ended to differential. Further, the pixel output is unavoidably contaminated with noise, and the anti-aliasing is required in a digital sampling system. Therefore, a signal conditioning and ADC driving circuit is designed to meet these objectives.

Shown in Figure 3.5, the signal conditioning and ADC driving circuit is implemented based on a high performance differential operational amplifier. The candidate differential operational amplifier should satisfy the following criteria:

- High bandwidth for large signal swing ($1\mathrm{V_{pp}}$): over 100MHz.

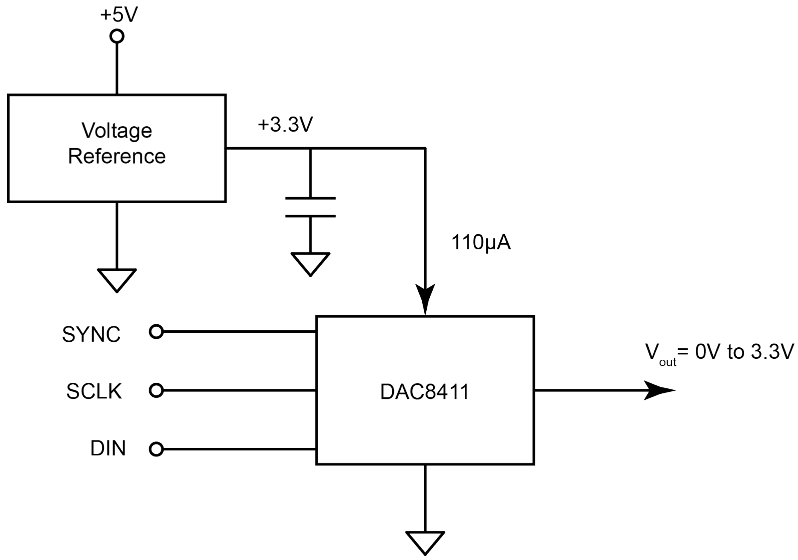


Figure 3.6: Voltage reference as power supply to DAC8411

- High slew rate: The pixel output rate is 33MHz when the even and odd pixel outputs are presented in parallel. The slew rate should satisfy

$$\text{SlewRate} \geq 2\pi f_{\max}V_{peak} = 2\pi \times 33\text{MHz} \times 0.5\text{V} = 104.6\text{V}/\mu\text{s}. \tag{3.5}$$

  The desired slew rate is set to $500\text{V}/\mu\text{s}$.

- Short Settling Time: The pixel output period is 30.3ns, therefore the desiable settling time should be limited within a quarter of this pixel period, that is 8ns approximately.

- Low Voltage Noise: The noise level should be lower than one LSB of ADC.

AD8138 from Analog Devices is selected based on criteria above, which features 265MHz bandwidth for large signal swing, $1150\text{V}/\mu\text{s}$ slew rate, 8ns settling time with 0.01% error and $5\text{nV}/\sqrt{\text{Hz}}$ voltage noise from 100kHz to 40MHz. The schematic of the signal conditioning circuit is shown in Figure 3.7.

Analog/Digital Power Isolation

When digital and analog power planes run on the top of each other, the mutual capacitance between two planes will couple them for high frequency signals and allow noise to transfer from the digital plane to the analog plane. However, ADCs and DACs are components requiring both analog and digital power sources. Clean analog power is required for their sensitive internal components, and on the other hand, separate power source is needed for driving its high-speed digital parts. Directly connecting these two power supplies to the same power plane introduces noise from the ADC or DAC's digital part. However, using the board's digital power plane to power the digital part of ADC and DAC is not an option since it makes the digital power plane overlap the clean analog region. To address this problem, the digital power input pins of the ADC and DAC are connected to the analog power pins through a

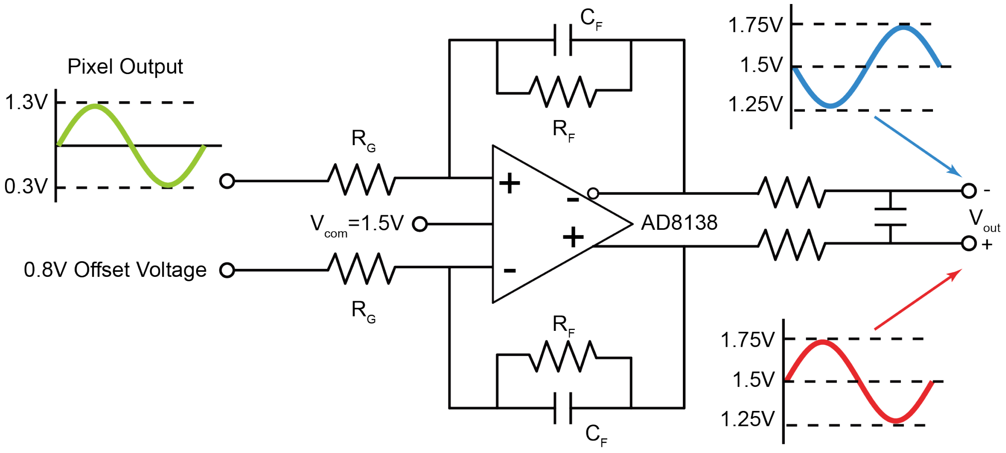


Figure 3.7: Schematic of signal conditioning circuits

passive LC filter (shown in Figure 3.8). The inductor and capacitor create a second order filter with a -3dB bandwidth of $\omega_{-3\text{dB}} = 1/\sqrt{LC}$.

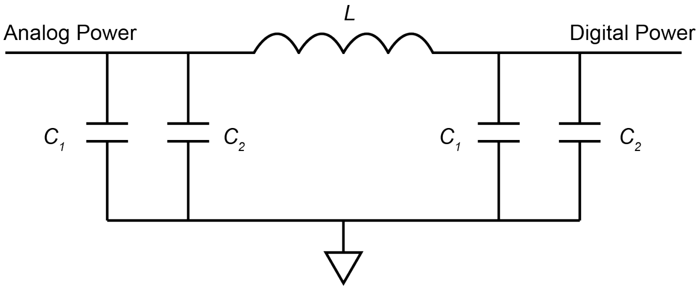


Figure 3.8: Passive LC filter between analog and digital power supply in ADC and DAC

3.2.4 High-Speed Camera Interface Design

High-performance vision system produces high resolution image data in high speed, requiring fast data transmission between camera module and image processing unit. According to the literature review in Section 1.2, the camera interface remained a bottleneck for the prior art vision systems. Communication standards designed for more leisurely data environments have not proved adequate for handling the demands of high-speed vision applications. To address this situation, several camera interface standards have emerged to provide fast and reliable data transmission as well as camera control [33].

There are four major camera buses currently used in the digital vision systems - USB, IEEE 1394, GigE Vision and Camera Link. There is no perfect one-size-fits-all solution, thus when deciding which camera bus is right for the application, decisions and tradeoffs must be made depending on the requirements of the system. Generally speaking, the candidate camera interface is evaluated from the following criteria:

- **Data Throughput**: The data throughput represents the rate at which image data can be transferred over the bus. High data rate, continuous throughput and low latency is desired in our system. According to the hardware design, the pixel frequency is 66MHz and the resolution of each pixel data is 14-bit, therefore, the minimum data transmission rate required is calculated as:

$$14\text{bit} \times 66\text{MHz} = 924\text{Mbps} \tag{3.6}$$

Conservatively speaking, a data throughput over 1Gbps is desired.

- **Interface Flexibility**: Besides transferring the image data from camera module to image processing unit, a desirable camera interface must provide the flexibility of camera control. An independent channel for camera control information is more favorable.
- **Transmission Length**: The transmission length determines the maximum possible distance between the camera modules and image processing unit. Though this specification is not critical, a desired cable length is between 5 meters to 10 meters.
- **Synchronization**: Applying in the application of real-time position sensing system, the vision system is required to work synchronized the overall system. The camera triggering should be addressed and handled easily within the camera interface.

In the rest of this section, the four currently used camera interfaces are described and compared. The final decision is made based on the criteria presented above.

USB

USB interface cameras are generally low-cost. However, currently USB 2.0 has a limited data rate of 480Mbps working in burst mode, and cables for USB are generally less than 5 meters without a repeater. Lacking of a hardware specification for image acquisition devices, the widespread adoption of USB for vision applications is obstructed [33]. USB also does not provide any camera triggering mechanisms. Therefore, it is difficult to synchronize USB cameras with each other and the rest of system.

IEEE 1394

IEEE 1394 was designed for vision equipment. Many IEEE 1394 cameras provide the function of camera triggering, which simplifies system synchronization. The maximum data throughput of IEEE 1394 is 800Mbps in burst mode. IEEE 1394 interface provides point-to-point connection which is limited to less than 5 meters without hubs or repeaters[33]. At the same time, IEEE 1394 has power on the cable, thus most cameras can draw power off the IEEE 1394 bus without need for an external power source.

GigE Vision

The GigE Vision standard is a new camera bus technology for vision systems. GigE Vision features of high data rate, ubiquitous interface hardware, low cost cable, and widespread popularity, which makes it an attractive option for vision systems. Considering the hardware limitations and software overhead, the practical maximum bandwidth is close to 800Mbps [33]. With the cable length reaching 100 meters, GigE Vision is the most favorable camera bus in camera network applications.

One major disadvantage of GigE Vision is its natural drawback of Ethernet. The Ethernet network protocol is very inefficient. When network devices simultaneously send packets of data, these packets may collide midstream. Therefore, a retransmission of the data is required which significantly reduces the effective network bandwidth. Busy networks with multiple network devices vying for throughput result in an increasing number of collisions, dramatically reducing the efficiency of the network and thereby degrading the performance of camera interface [34]. Because the latency of data delivery is unreliable and varies depending on the events that are occurring on the network, GigE-based vision system is not able to guarantee real-time data transmission. Another problem of GigE Vision is the synchronization. It is difficult to use the PC to condition a trigger signal between GigE vision cameras and the rest part of system [33].

Camera Link

Camera Link is a communication interface designed for point-to-point vision applications based on a serial communication protocol. Camera Link interface is based on the National Semiconductor's ChannelLink technology which has been extended to support general-purpose low voltage differential signaling (LVDS) data transmission. The ChannelLink chip performs a 7:1 mux and transmits data as LVDS signals at a maximum speed of 85MHz. ChannelLink consists of a driver and receiver pair. The driver accepts 28 single-ended data signals and one single-ended clock signal. The data is serialized 7:1, and four data streams and a dedicated clock are driven over five LVDS pairs. The receiver accepts the four LVDS data streams and LVDS clock, and then converts to 28 bits single-ended signals and a clock to the board [35].

Camera Link, a high-speed serial digital bus designed specifically for vision systems, features the highest throughput of any camera bus. Camera Link provides a three-tiered bandwidth structure (base, medium, and full) to address a variety of applications (shown in Figure 3.9). The base-configuration acquires at up to 2040Mbps, although the other camera interfaces provide only 800Mbps or less. Medium and full configuration cameras acquire the image data at up to 4080Mbps and 5440Mbps [35].

The Camera Link standard replaces expensive, custom cables with a single, low-cost standard cable with few wires [35]. By the means of transmitting 28 parallel single-ended signals into four high-speed differential pairs, Camera Link Standard reduces cable size and cost, as well as increasing noise immunity and maximum cable length. The maximum cable length defined by Camera Link standard is 10 meters.

Shown in Figure 3.9, besides providing high-speed data transmission based on ChannelLink technology, Camera Link offers a serial communication between camera and frame grabber using LVDS signals. This feature allows users to define their specific serial commands for setting camera operation parameters, such as exposure time, ROI position and ROI size. Camera control information has its dedicated transmission channel separated with the image data transmission. Four Camera Control signals are also provided by the Camera Link cable for camera triggering, offering good synchronization of camera modules with the rest of system. Overall, Camera Link provides the most I/O flexibility and

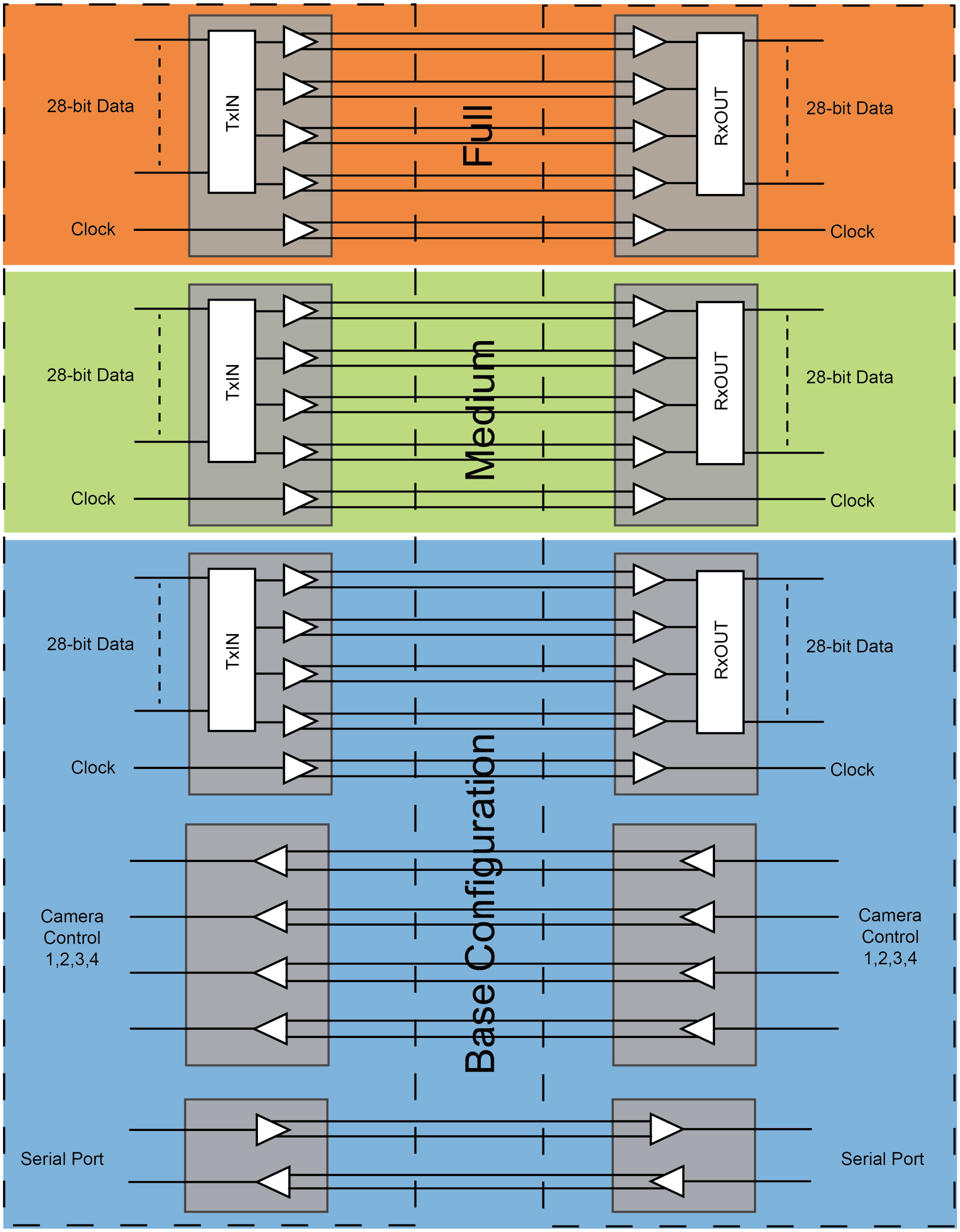


Figure 3.9: Block diagram of base, medium and full configuration of Camera Link adapted from [35]

capability comparing to other camera interfaces.

Table 3.4: Comparison of camera interfaces, data from [33]

	USB	IEEE 1394	GigE Vision	Camera Link
Connection Type	Master-Slave	Point-to-Point	LAN	Point-to-Point
Data Throughput	480Mbps	800Mbps	1000Mbps	>2000Mbps
Image Data Streaming	Burst	Burst	Continuous	Continuous
Transmission Distance	5m	4.5m	100m	10m
Real-Time Signaling	No	No	No	Yes
Synchronization	Difficult	Easy	Difficult	Easy
Camera Control	Yes	Yes	Yes	Yes

The performance of each camera interface is summarized and compared in Table 3.4. It is clear that only GigE Vision and Camera Link are able to provide the data throughput and transmission length required by the system. Considering that the vision system is applied in high performance real-time position sensing applications, it is required that the image data is transferred in real-time with low latency. Since the inherent drawback of Gigabit Ethernet, GigE Vision interface is unable to provide real-time data transmission and its transmission latency is variable depending on the status of Ethernet network. Therefore, Camera Link is the best option for high-speed vision applications.

A customized serial communication protocol similar to RS-232 is designed for camera control communication. The four camera control LVDS pairs in Camera Link standard are used for the synchronization between camera modules and the rest of system. The data transmission, camera control communication and camera triggering are included in one compact Camera Link cable assembly.

3.2.5 High-Speed Image Processing

The goal of image processing in vision-based position sensing system is to retrieve the 3D position of targets being tracked. Fast image acquisition and high resolution image data challenge the speed of image processing. In order to match the performance of image acquisition electronics, the image processing must be designed and implemented in a smart way.

To reconstruct the 3D position from the raw target image, the image processing involves several cascaded sub-processes, including image noise correction, sub-pixel target position interpolation and 3D position reconstruction. For the purpose of accelerating the image processing speed, as well as reducing the data transmission of camera interface, the image processing is divided into two parts: the pre-processing in camera module, including image noise correction and sub-pixel target position interpolation, and the post-processing in a dedicated image processing unit, including 3D position reconstruction. The detailed design of each part is presented below.

Camera Built-in Image Pre-processing

Camera built-in image pre-processing plays a significant role in accelerating the speed of image processing. Considering that the image noise correction and sub-pixel target position interpolation algorithms require simple mathematic operations, such as accumulation and multiplication, they can be

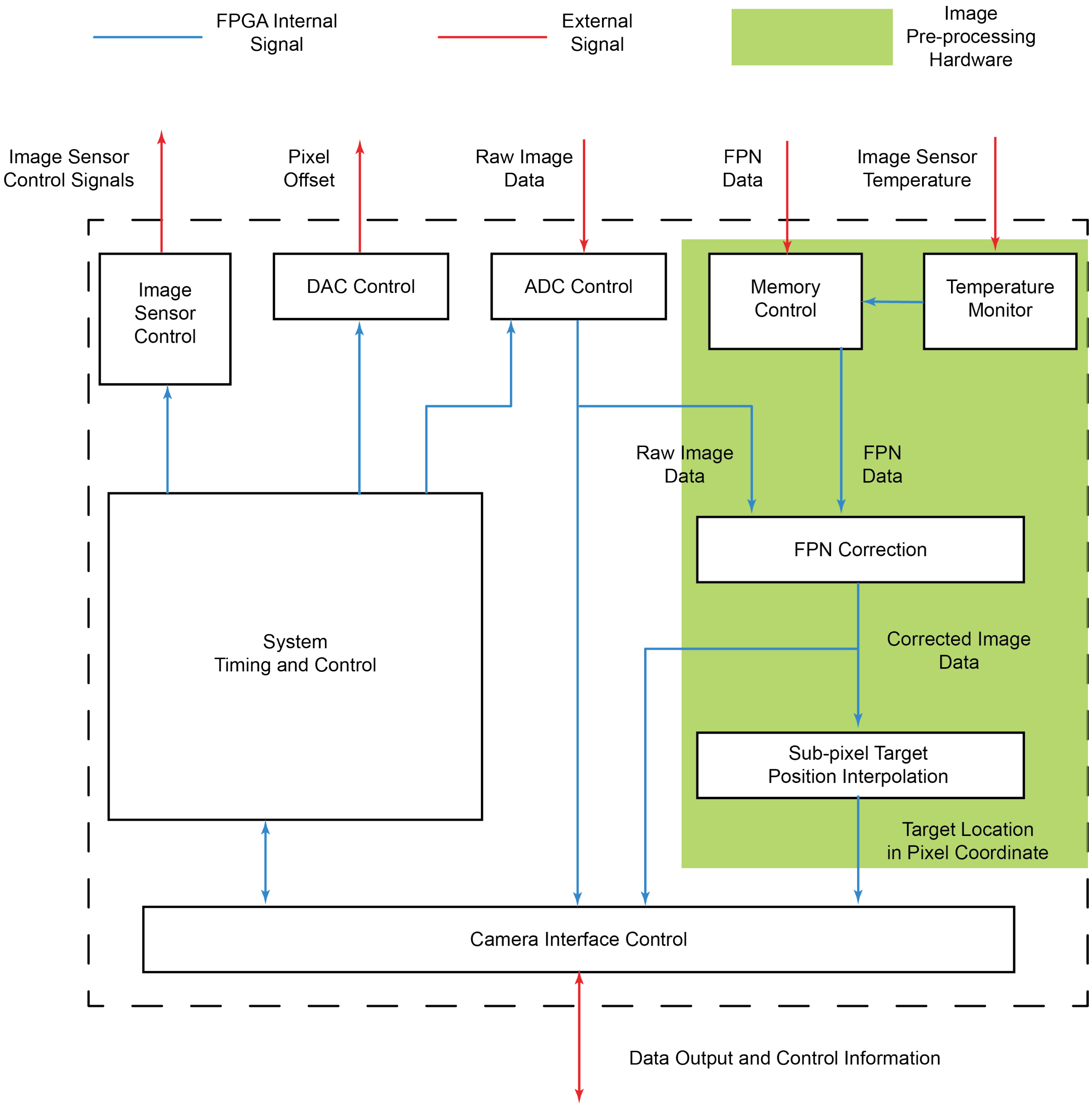


Figure 3.10: Camera module FPGA architecture

realized by hardware inside the FPGA (shown in Figure 3.10). The selected FPGA integrates DSP-48 slices which provides strong computation capability to realize the image pre-processing algorithms.

To realize the fixed pattern noise (FPN) correction, image sensor calibration is required to adapt the CMOS image sensor to the specific lighting conditions. Considering that the FPN is a fixed spatial distribution (time invariant), it can be removed by using a pixel-by-pixel calibration. Once calibrated, the camera modules perform real-time FPN correction of each new image. The calibration stage includes pointing the camera at the typical scene to be imaged and adjusting the offset and gain of the image sensor output according to the specific operation condition. Calibration images are obtained by taking several reference images under certain operation conditions. After calibration is completed, the calibration image is used in the real-time FPN correction. The main idea of the real-time FPN correction

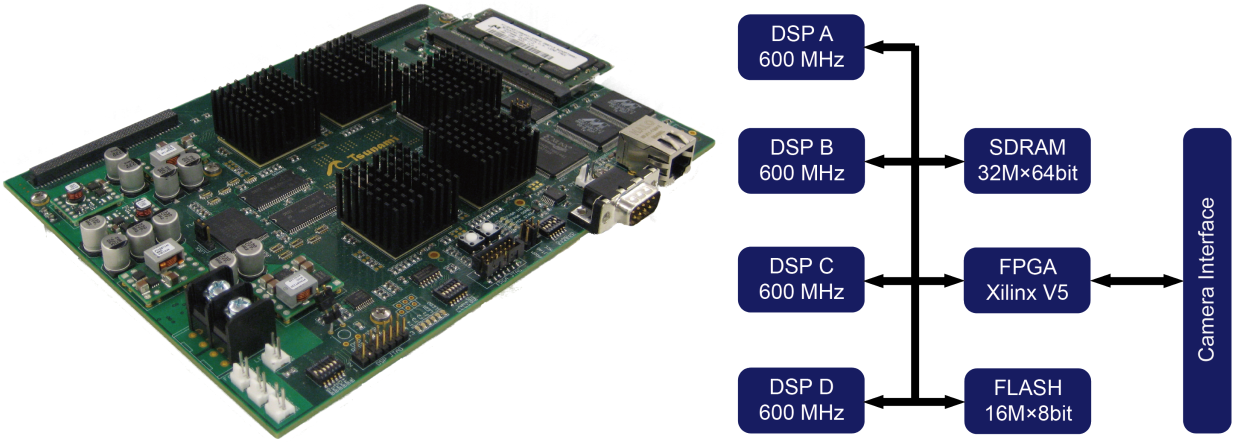


Figure 3.11: Tsunami real-time computer mother board developed by Kris Smeds adapted from [36]

algorithm is based on a first order correction. The advantage of this first order correction lies in the fact that it inherently satisfies the requirement of high-speed application as a result of simple computation required. The detailed analysis of the FPN calibration and correction is presented in Section 4.2.

Because FPN varies with operating conditions (exposure time, pixel sampling rate, lighting conditions and sensor temperature), four on-board non-volatile memory blocks are implemented in each camera module to store the different FPN calibration maps.

Sub-pixel target position interpolation is the process which precisely and accurately determines the target image center in the sub-pixel range based on the target image. The centroid method (Equation 2.19) and squared-centroid method (Equation 2.20) are two intensity-based algorithm widely used. From Equation 2.19 and Equation 2.20, it is clear that these two methods require simple computation, therefore, it is effective to implement them by using hardware in the FPGA. By this means, the sub-pixel target position interpolation is working parallel with the target image acquisition, vastly increasing the processing speed.

Image Post-processing in Tsunami Real-Time Computer

The reconstruction of target 3D position requires complex mathematical computations (shown in Section 2.3). The image post-processing speed is improved by using a dedicated image processing unit. To achieve the high processing speed, a customized real-time computer has been developed by Kris Smeds [36].

This real-time computer is designed based on a multiprocessor architecture [36]. This real-time computer (shown in Figure 3.11), named Tsunami, is built around the Virtex 5 FPGA and four TigerSHARC DSPs. The high speed FPGA frontend provides fast digital signal interface and fast communication to host PC. The TigerSHARC DSPs offer powerful computation capability to deal with complex computation.

Since the Tsunami real-time computer is independent of applications, users are allowed to design their own daughter board and I/O interface based on the requirements of application. A customized

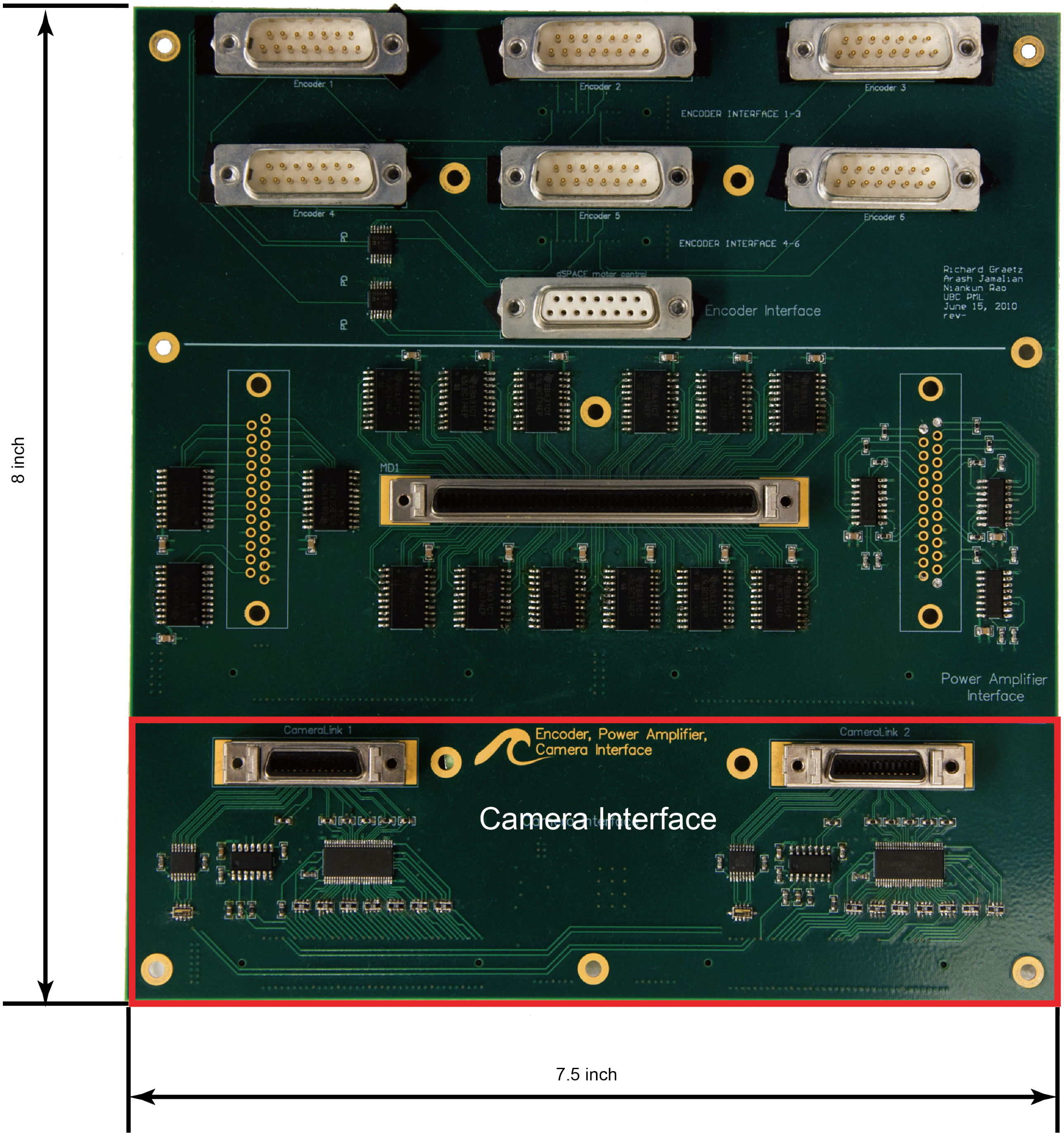


Figure 3.12: Tsunami I/O Board Designed by Niankun Rao, Richard Graetz and Arash Jamalian

I/O board is designed for Tsunami real-time computer to interface the two camera modules (shown in Figure 3.12). For purpose of guaranteeing the digital signal integrity, the impedance and length of high-speed digital signal traces are controlled and matched on the camera interface board.

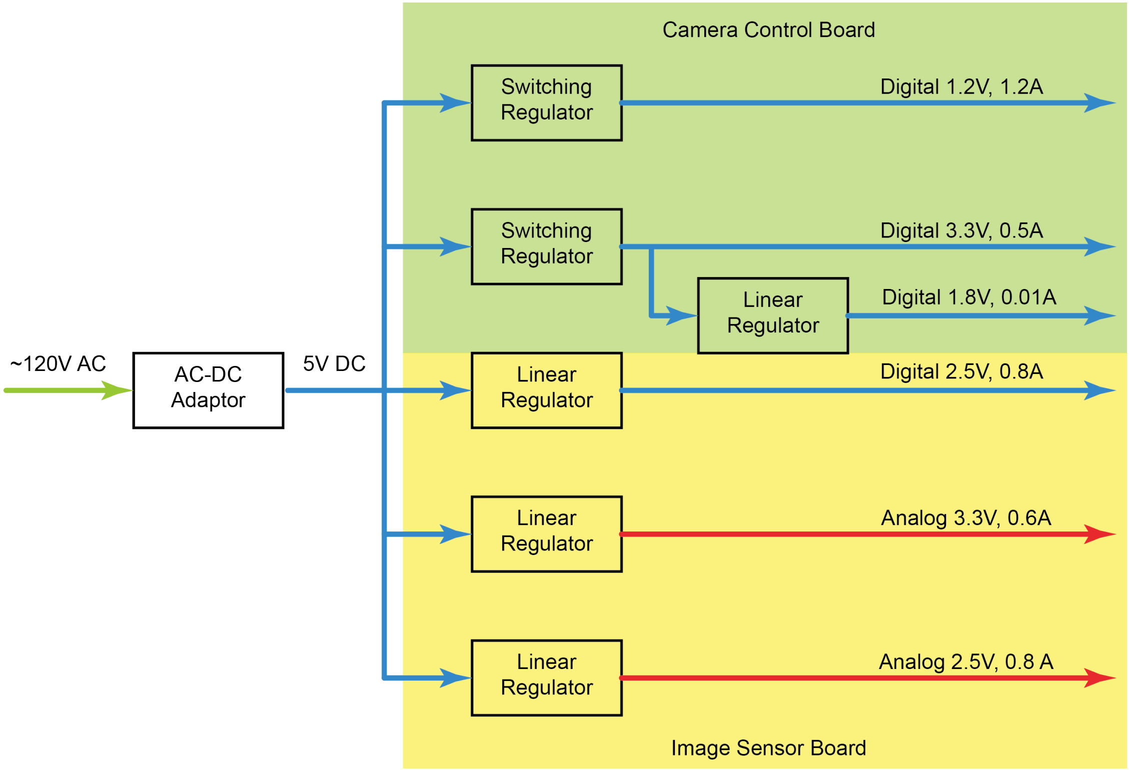


Figure 3.13: Power regulation network of camera module

3.2.6 Power Distribution System Design

The IC chips on the camera module require the DC power supply at different voltage levels. The power quality provided to the electronic devices affects their performance. Therefore, the design goal of power distribution system is to generate the DC voltage levels required by the on-board electronics with high power quality. The summary of the power levels and their required currents is provided in Table 3.5. The power consumption of each camera module is 9W approximately in worst case.

Table 3.5: Summary of camera module power requirements

Voltage Supply		Current [A]		Power Consumption [W]
Type	Level [V]	Required	Designed	
Digital	1.2	1.2	2	1.4
	1.8	0.01	0.1	0.018
	2.5	0.8	1.5	2
	3.3	0.5	1	1.65
Analog	2.5	0.8	1.5	2
	3.3	0.6	1	1.98

Figure 3.13 shows the power regulation network. The input 120V AC wall power is converted to 5V DC power by a AC-to-DC power adaptor. The power regulation network generates both digital and analog power levels, and each power levels has its dedicated power regulator. This separation guarantees that the switching noise on the digital power planes does not transfer to the analog power

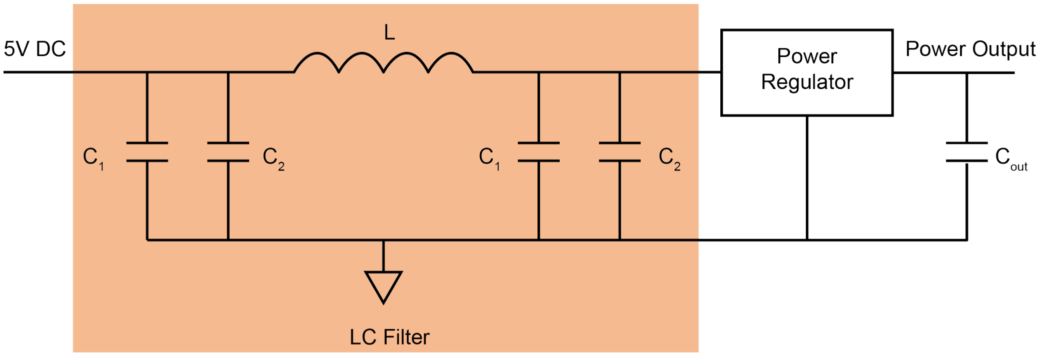


Figure 3.14: Passive LC filter between 5V DC power and input of linear regulator

planes.

Shown in Figure 3.13, linear and switching regulators are used to generate different power levels. The linear regulator is an active device which is operating at its linear region, generating a regulated output voltage at a specific voltage level. The output voltage of linear regulator is continuously regulated with high quality. Analog power levels are all generated by linear regulators. However, the voltage drop between the input and output voltage of linear regulator causes a power loss on the linear power device, reducing the efficiency and generating heat. Compared to linear regulators, switching regulators generate a regulated output by rapidly switching the input supply. Since the supply is either on or off, the switching regulators do not dissipate much power and achieve high efficiency above 90%. Due to their high efficiency, switching regulators do not generate much heat. On the downside, the output voltage of switching regulators has higher ripples due to the on-off operating principle. In the power distribution network, switching regulators are used for generating digital power levels which require high current but not sensitive to voltage ripples. On the other hand, linear regulators are mostly used for generating clean analog power levels. However, the digital 1.8V and digital 2.5V are generated by linear regulators. Due to the low current requirement of the digital 1.8V power level, the power dissipation on the linear regulator is negligible. Meanwhile, the digital 2.5V is required on the image sensor board where lots of sensitive analog devices are located. For the purpose of preventing switching noise contaminating the analog devices, a linear regulator is used.

Switching regulator works in on-off mode, therefore the 5V DC power is unavoidably contaminated with noise from digital power. The voltage ripple at 5V DC power causes high frequency noises at the input of linear regulator for analog power levels. Since linear regulator has low power ripple rejection at the high frequency region, a passive LC filter is implemented between the 5V DC power and input of each regulators in order to guarantee a good isolation between digital and analog power regulators (shown in Figure 3.14). The LC network not only prevents the noise transferring from input side to output side but also prevents the noise transmission in the opposite direction. The -3dB bandwidth of the LC network is $\omega_{-3\text{dB}} = 1\big/\sqrt{LC}$, which is set according to the power ripple frequency.

Based on Table 3.5 and the architecture of power regulation network, the input current at 5V DC in worst case is calculated as:

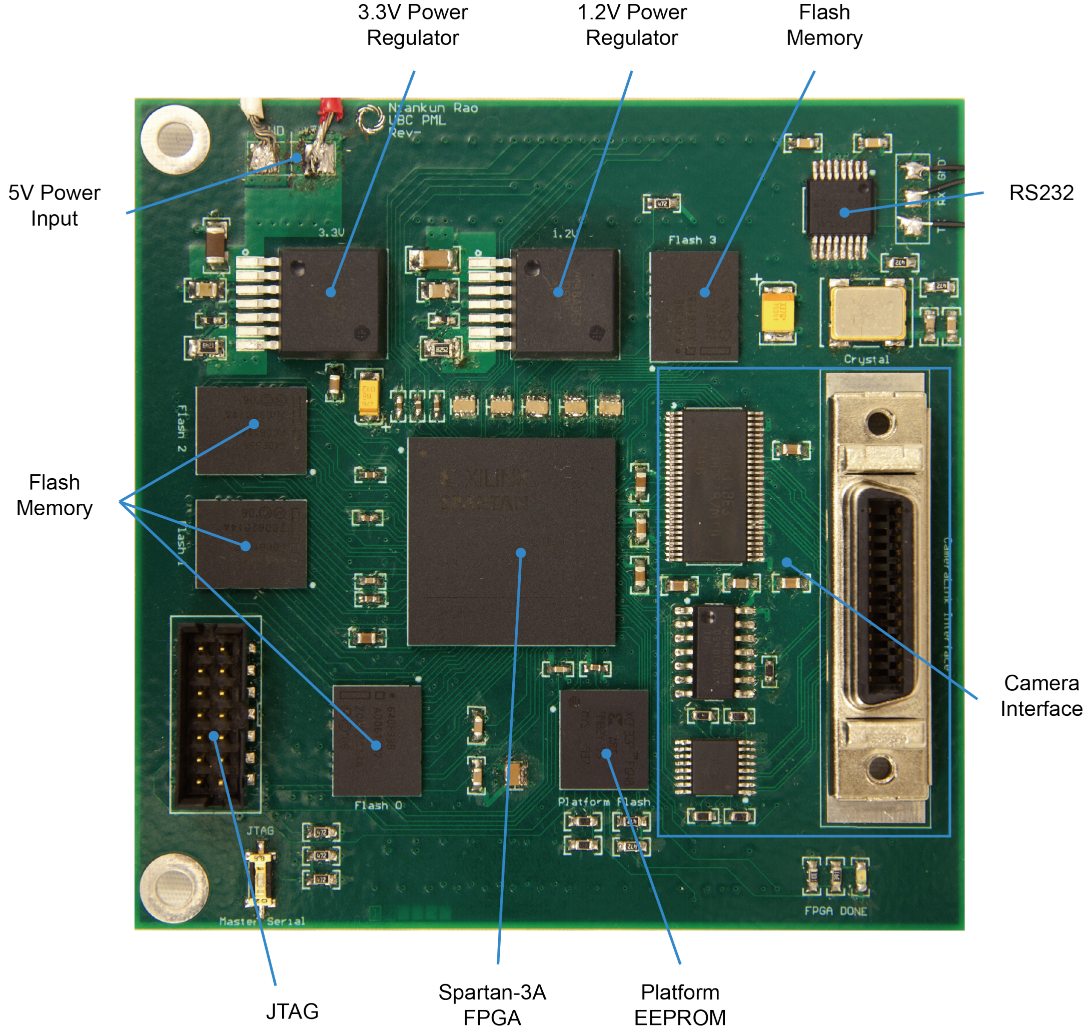


Figure 3.15: Camera module control PCB

$$I_{5\text{V}} = \frac{1.2\text{A} \times 1.2\text{V}}{5\text{V} \times 95\%} + \frac{(0.5+0.01)\text{A} \times 3.3\text{V}}{5\text{V} \times 95\%} + 0.8\text{A} + 0.6\text{A} + 0.8\text{A} = 2.86\text{A} \tag{3.7}$$

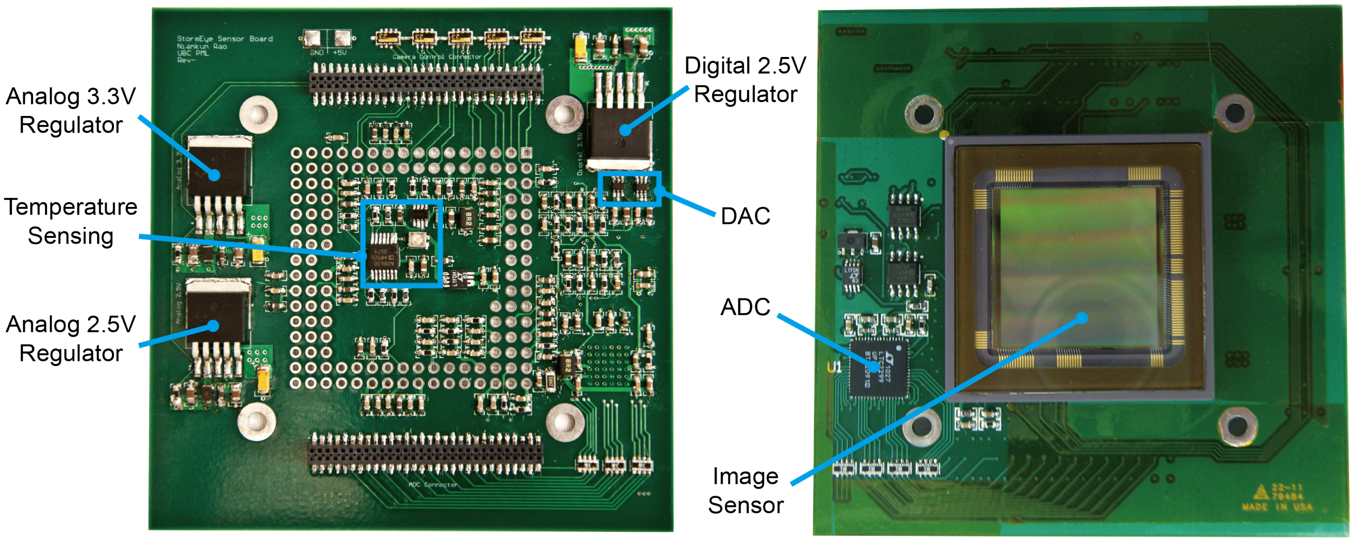
where the efficiency of the switching voltage regulator is 95%. Therefore, the efficiency of the designed power distribution system in worst case is computed as:

$$\eta_{\text{PDS}} = \frac{9\text{W}}{5\text{V} \times 2.86\text{A}} \times 100\% = 63.2\% \tag{3.8}$$

3.2.7 Print Circuit Board Design

Two print circuit boards is designed for each camera module. The camera control board (shown in Figure 3.15) is a pure digital circuit board where FPGA, on-board Flash memories and camera interface are located. The image sensor board is an analog and digital mixed circuit board where the CMOS image sensor, ADC and DACs are located (shown in Figure 3.16).

Both boards are designed with 6 layers: 3 signal layers and 3 power planes. The layer stack-up of



(a) Image Sensor Board Bottom Side

(b) Image Sensor Board Top Side

Figure 3.16: Image sensor PCB

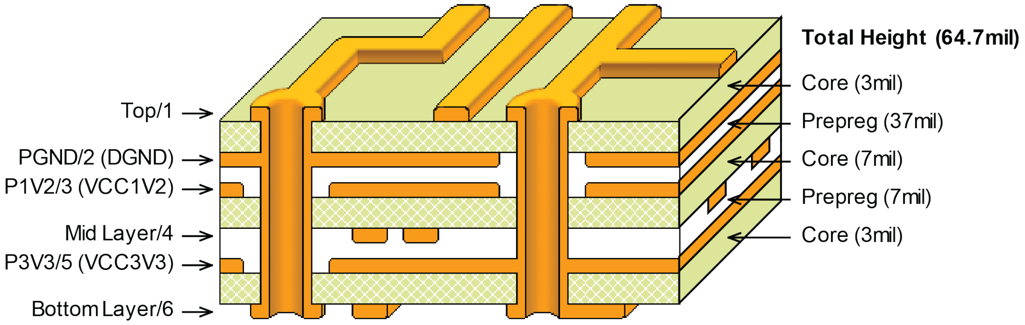


Figure 3.17: Layer stack-up of camera module control PCB

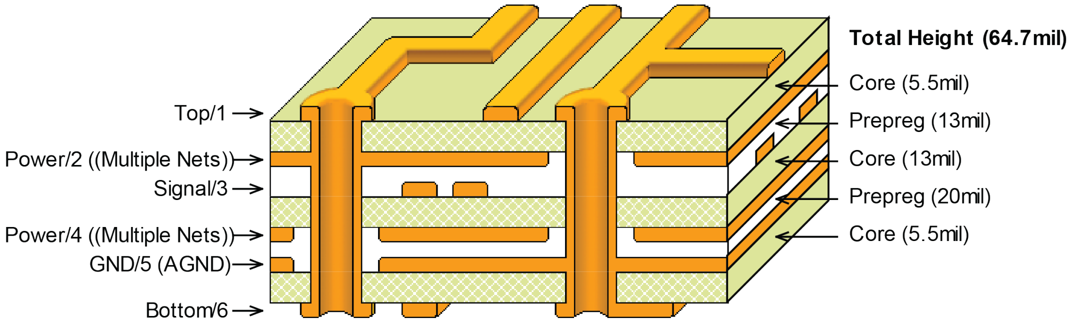


Figure 3.18: Layer stack-up of image sensor PCB

each board is shown in Figure 3.17 and Figure 3.18. Since the two PCBs are designed for high-speed signals, the layer stack-up and signal trace routing are designed to ensure several requirements:

- Independent signals should not be routed in parallel for long distance so that the mutual capacitance and inductance are minimized. Otherwise, the cross-talk among these signals is introduced.
- A reference power plane is required for each signal on its neighboring layer. By this means, the power plane provides a minimized return path for the signal current and minimize the inductance of the signal lines.
- In order to guarantee the signal integrity, signal termination is required for those high-speed signals. In this case, the characteristic impedance of the transmission lines is required to match the termination resistors. The trace width, trace spacing and distance between the adjacent layers are determined by the requirements of controlled impedance. The impedance is controlled at 50Ω for single-ended signals and 100Ω for differential signals. Further, for high speed signals, the trace length in the same signal group, like ADC signals, ChannelLink signals and Flash memory signals, are matched within very small difference in order to guarantee the signals in the same signal group are transmitted with the same propagation delay.
- Both digital and analog signals run on image sensor board. The digital signals operate by fast switching between logic high and low level. The edges of fast switching digital signal generate noise, causing voltage spikes on power planes as well as other adjacent signals. The noise effect on the digital signals can be minimized when the digital signals are properly routed. However, the noise on the analog power planes undermines the precision of the analog components. reducing the certainty of analog operation. Therefore, the analog components need to be isolated from the noise generated by the digital devices. The digital power planes and analog power planes cannot be overlapped with each other, and no digital signal trace can be routed into the analog region on the board. The only common connection between analog and digital part is a single connection between the analog and digital ground. This common connection is necessary to keep the same reference level for both digital and analog sides. A single point connection also enforces a single controlled current return path for interaction between analog and digital side. The controlled current return path does not allow direct and sharp returning current. As a result, the noise associated with the digital signals spreads out over the plane so that its effect is averaged and minimized. Further, the long return path of digital signal current introduces addtional inductance which is able to further filter out the high frequency noise. The analog and digital side partition is shown in Figure 3.19.

3.3 System Optical and Mechanical Design

3.3.1 Infrared Target Design

Vision-based position sensing systems utilize image data captured from image sensors to retrieve the 3D position of an target. Multiple DOF position sensing of the object is realized by using special targets

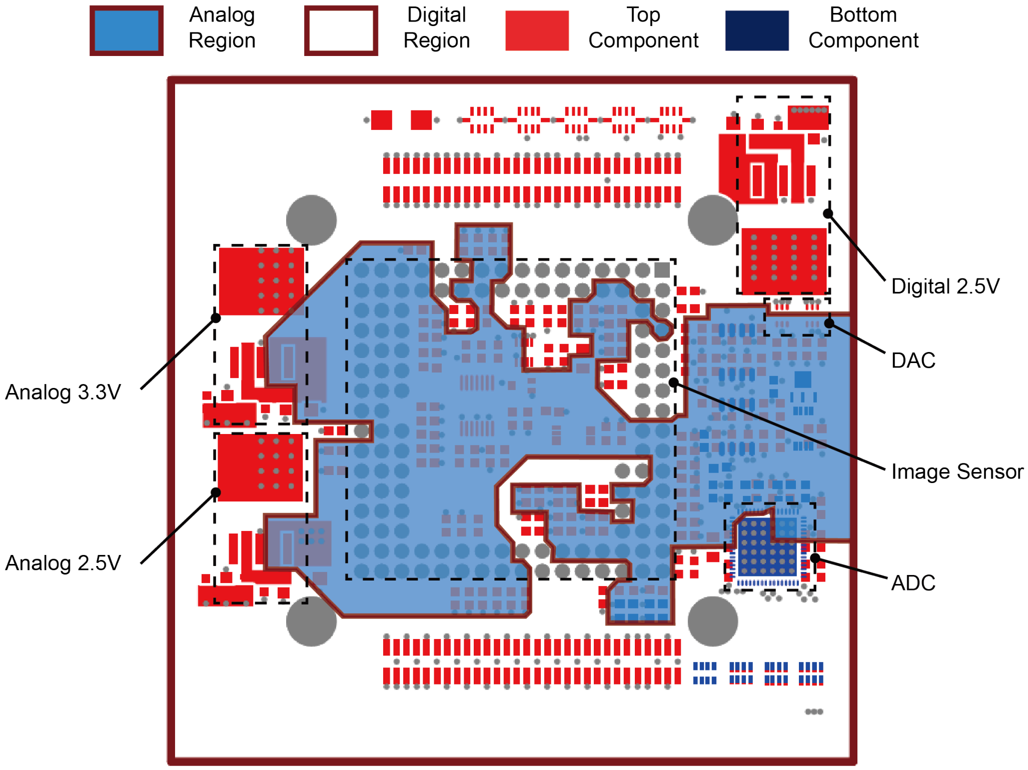


Figure 3.19: Analog and digital partition on image sensor board

attached to the object to be tracked. For the purpose of minimizing the side-effect of ambient light, the target in the system usually works in the infrared light region. Based on the working principles, the infrared target is categorized into two groups: passive target and active target.

- Passive targets are coated with a retroreflective material to reflect the infrared light back that is generated near the camera lens. This kind of targets is mostly spherical so that a large view angle with relative uniform reflection intensity is able to be achieved.
- Active targets use infrared light emitting elements, mostly infrared LEDs. Single or multiple infrared LEDs can be used according to the requirement of the vision system. These infrared LEDs are generally driven and controlled by dedicated electrical circuit.

It is clear that the spherical passive targets offers the visibility from all sides and does not require external driving circuits. However, its sensitive surface, which is difficult to clean, affects the reflected light illumination. Limited by the reflection efficiency, a considerable power loss is introduced, causing a low power reflective light and further requiring longer exposure time in the imaging system. Conclusively speaking, compared to passive target, the active target is more favorable.

The infrared LED utilized in the system is HE8812SG infrared LED from Opnext. This LED emits the infrared light with the peak wavelength of 870nm and keeps a uniform radiation intensity within a range of ±40° (shown in Figure 3.20). The uniformity angle can be further extended by using a diffusor sphere at the expense of lower optical power output.

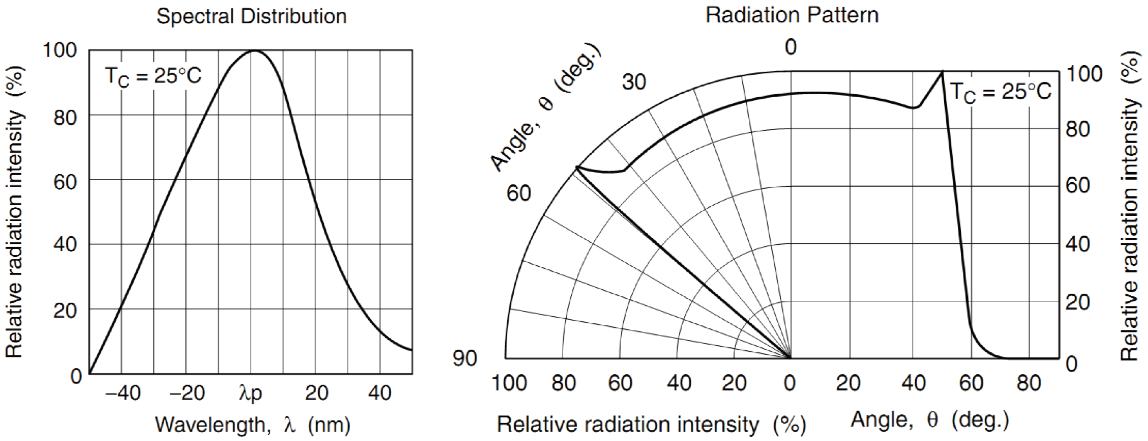


Figure 3.20: Spectrum distribution and radiation pattern of HE8812SG Opnext Infrared LED, from [37]

The active target requires dedicated circuits to drive and control the infrared LEDs. The optical power of the selected infrared LED depends on its forward current I_F. Therefore, a current source circuit is designed to meet the following criteria:

- Adjustable current: 50mA to 150mA
- Low current ripple: $<$ 5mA
- Powered by battery
- Low power dissipation by LED driving circuit

According to the design goals above, the LED driving circuit is designed as shown in Figure 3.21. Rechargeable high capacity battery is used to supply the power. A switching DC/DC convertor is designed to boost the battery voltage to a desired voltage level with high conversion efficiency. The constant current source circuit is realized based on a linear regulator. The output current of the LED driver is determined by the current control resistor R_1 and R_2: $I_{out} = \frac{V_{reference}}{R_1+R_2}$. The experimental result shows that this LED driver achieves a current ripple (peak-to-peak) less than 2mA.

3.3.2 Lens Selection

Designers and integrators of vision systems are always looking for imaging systems with faster speeds and higher resolutions. An imaging system's resolution is determined by its lens and image sensor, therefore, these two key components must be selected in tandem. Higher resolution image sensor does not necessarily guarantee better images. Instead, lens and image sensor must be matched. Choosing a lens is more complicated than choosing a sensor, since multiple lens characteristics must be taken into consideration and traded off against one another [38].

The first lens characteristic asked in the lens selection criteria is the field of view, which is determined by the focal length and angular field. In our application, the object is tracked in a large working volume. Therefore, the lens with large field of view (normally features shorter focal length) is desired.

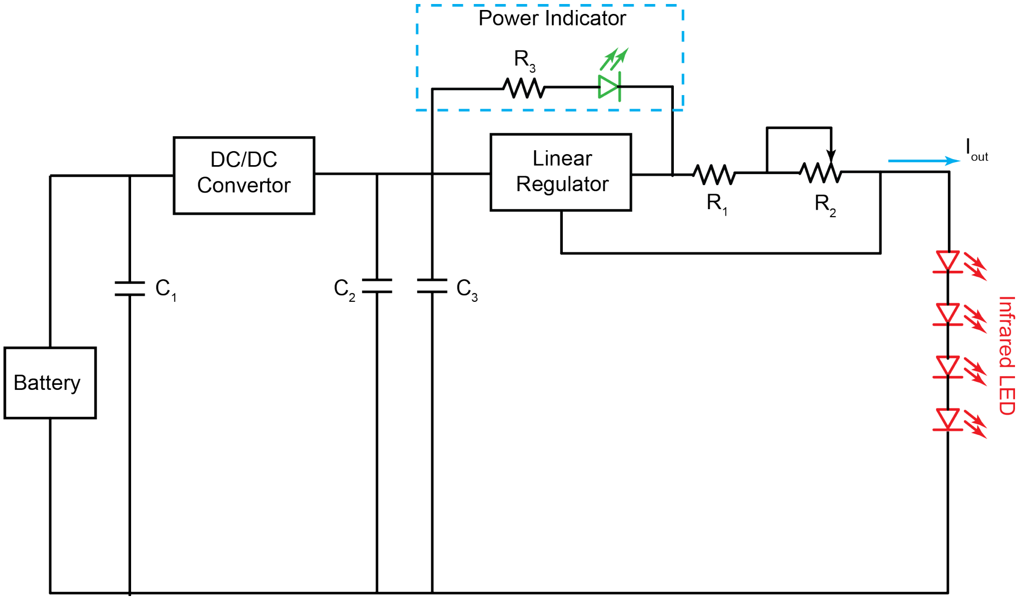


Figure 3.21: Infrared LED driving circuits

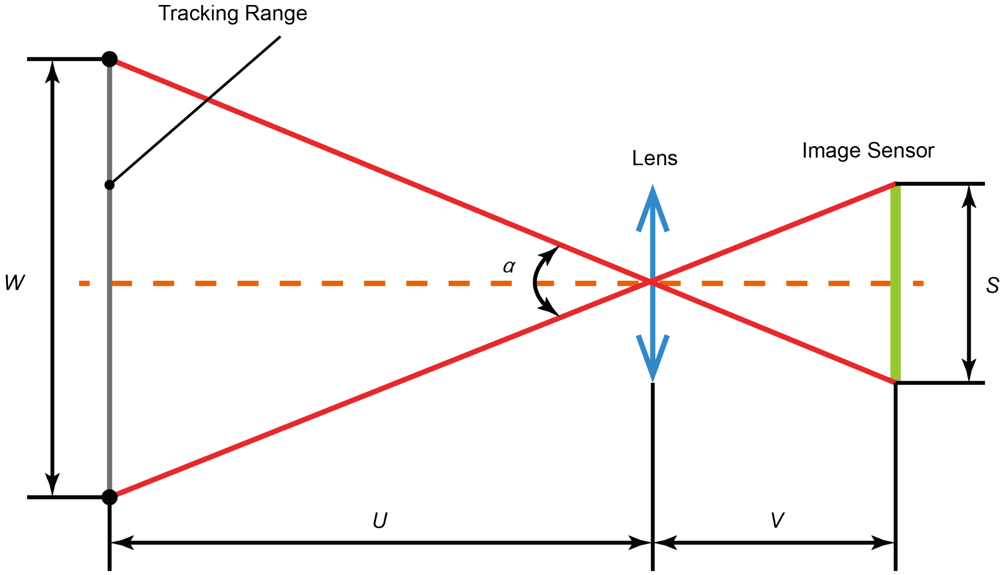


Figure 3.22: Illustration of the relationship between measurement volume and lens parameters

The actual focal length is determined by the desired tracking volume and physical size of image sensor. The Figure 3.22 shows the details.

In Figure 3.22, U is the distance between target and lens; V is the distance between lens and image sensor; W is the moving range of the target in vertical plane; S is the size of image sensor. Based on the geometry shown in Figure 3.22, we can derive that

$$\frac{S}{V} = \frac{W}{U}. \tag{3.9}$$

Further, according to the thin lens formula, it gives that

$$\frac{1}{U}+\frac{1}{V}=\frac{1}{f} \tag{3.10}$$

where f is the focal length of the lens. Combining Equation 3.9 and Equation 3.10 yields

$$f=\frac{SU}{S+W}. \tag{3.11}$$

Here, the size of image sensor S is known to be 24.6mm; the target moving range W is 1m; and the distance U is set to 1.5m. Therefore, it requires that the focal length f is smaller than 36mm. At the same time, the minimum angular field of the lens is derived by

$$\alpha=2\arctan\left(\frac{W}{2U}\right)=36.8^\circ. \tag{3.12}$$

Conservatively, the focal length is set to 25mm and the angular field is set to 40°.

3.3.3 Camera Body Design

The accuracy of 3D reconstruction does not only depends on the image quality and sub-pixel target position interpolation, but is also influenced by the system geometry as well as its stability. The camera body geometry parameters, such as the distance between two camera modules, affect the measurement sensitivity. On the other hand, the geometry parameters are limited by other considerations, including designed measurement volume, overall system dimension and weight. Further, the stability of these geometry parameters determine the repeatability and precision of measurement. Even though the system is designed to operate in a semi-controlled indoor environment, the thermal deformation of camera body due to the fluctuation of ambient temperature degrades the position sensing performance. In addition, the external vibration excites the vibration of camera body, causing a change in system geometry. Therefore, a solid camera body is required.

A single camera module is unable to reconstruct the 3D position of a target because the information it provides is not enough to determine the depth information. With two camera modules, additional information is offered by the second camera, thus the 3D position is reconstructed based on triangulation. Shown in Figure 3.23, the image sensor planes are parallel with the X_W-Y_W plane, therefore, the distance between the two camera modules does not effect the 3D position reconstruction in X_W and Y_W direction, but does affect the performance in retrieving the depth information. Intuitively, when the distance between the two camera modules is zero, it is equivalent that two camera modules merge together and become a single camera module. Under this situation, the depth information is unable to be recovered, therefore the reconstruction error in Z_W direction is infinity. When this distance increases, the accuracy of the measurement in Z_W direction is increasing. However, the accuracy cannot be unlimitedly improved with the increasing distance. When the distance is large enough, the accuracy in Z_W direction does not significantly increase. In an ideal case shown in Figure 3.23 where the lens distortion is not considered, the resolution in X_W-Y_W plane is proportional to the resolution of sub-pixel interpolation in image sensor plane, and this ratio is U/V. At the same time, based on the 3D position reconstruction algorithms described in Section 2.3, the resolution in Z_W direction of a 3D point is pro-

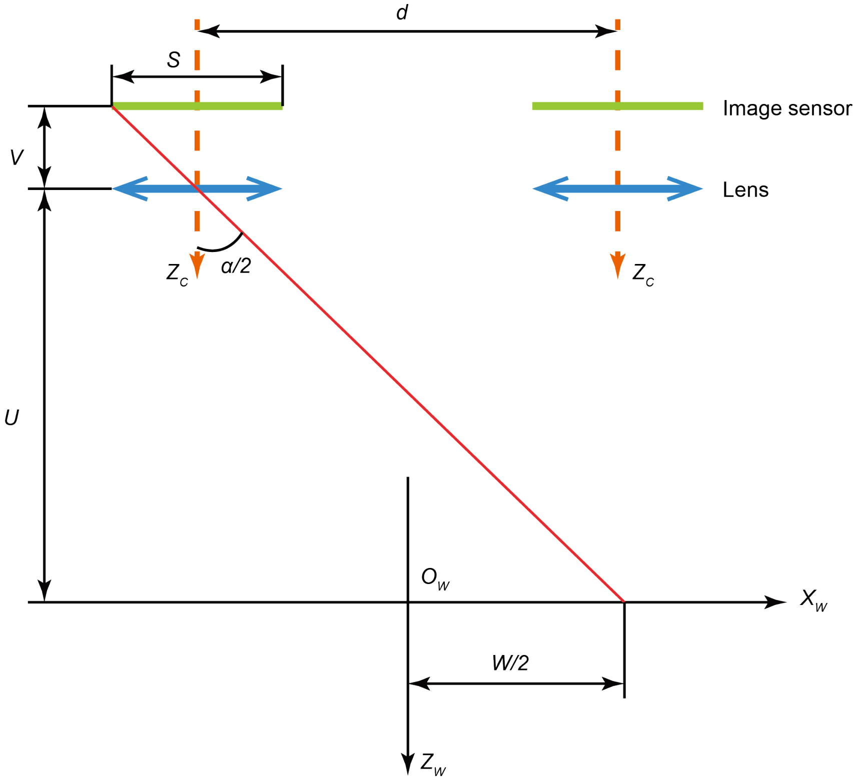


Figure 3.23: Schematic of two-camera system

portional to $1/d$. To further verify the effect of camera distance d on the reconstruction resolution in Z_W direction, a simulation is carried out based on the situation shown in Figure 3.23. In the simulation, the measured 3D point is located at the X_W-Y_W plane 1500mm away from the camera modules, and the resolution of sub-pixel interpolation is assumed to be 0.01pixel. Shown in Figure 3.24, it is clear that the resolution in Z_W direction is improved when the distance between camera modules is increasing.

Meanwhile, the distance between two camera modules is limited by the designed measurement volume. It is required that the targets to be tracked must be kept in the FOV of both camera modules. Therefore, the actual measurement volume is the intersection of the FOV of two cameras. A larger distance between cameras improves the 3D position reconstruction performance at the expense of losing measurement volume. Based on this analysis, it yields from Figure 3.23 that

$$\frac{S/2}{V} = \frac{(W+d)/2}{U}. \tag{3.13}$$

Under the requirement that the measurement range W is larger than 1000mm, it gives that $d < 430.2$mm. In Summary, the distance between two camera modules is set to 400mm.

Once the main geometry parameters of the camera body is determined, a solid camera body (shown in Figure 3.25) is designed and manufactured to maintain the geometry relationship between the two cameras. This camera body is made of invar which is a nickel steel alloy notable for its uniquely low coefficient of thermal expansion (around 1.2ppm/°C from 20°C to 100°C). On the other hand, external vibration and self-weight cause deformation of the camera body. An I-beam structure is designed in

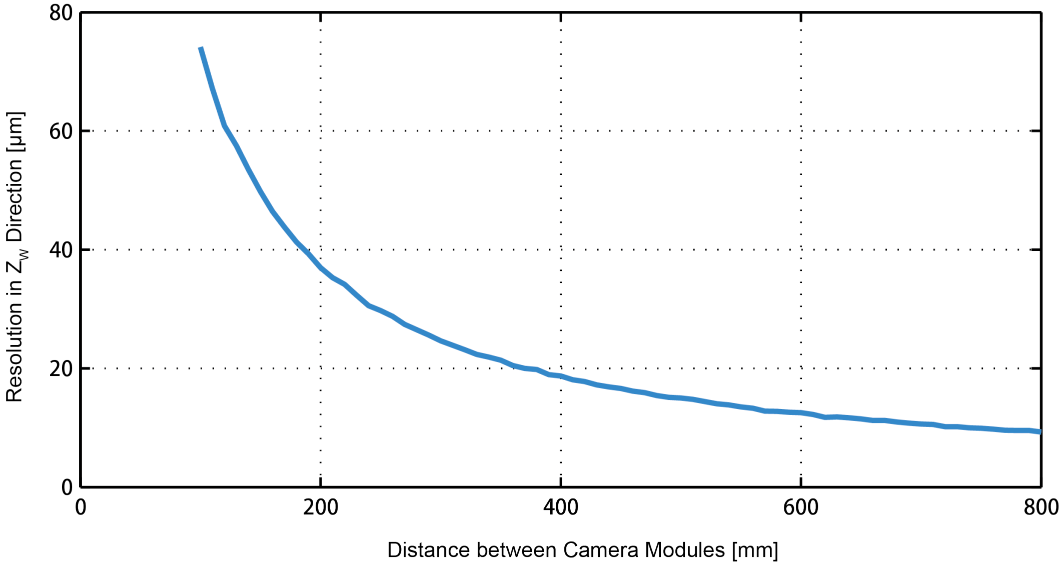


Figure 3.24: Resolution in Z_W VS camera distance d

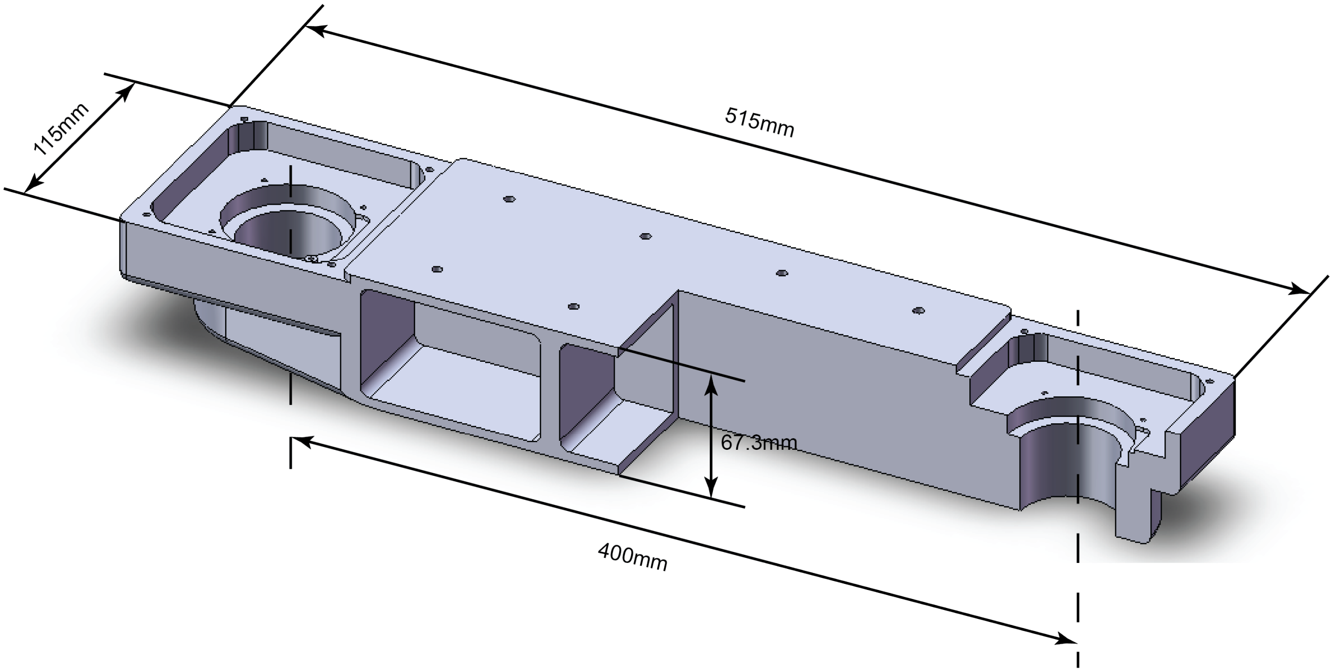


Figure 3.25: Camera body SolidWorks Model

order to make the camera body solid as well as reducing the total weight of the camera body.

3.4 Summary

In this chapter, the need for a set of customized stereo-vision system hardware is explained and the design of hardware is discussed. High-performance electronics, high-quality optics and solid camera body are designed and manufactured as an important contribution of this thesis. The final hardware

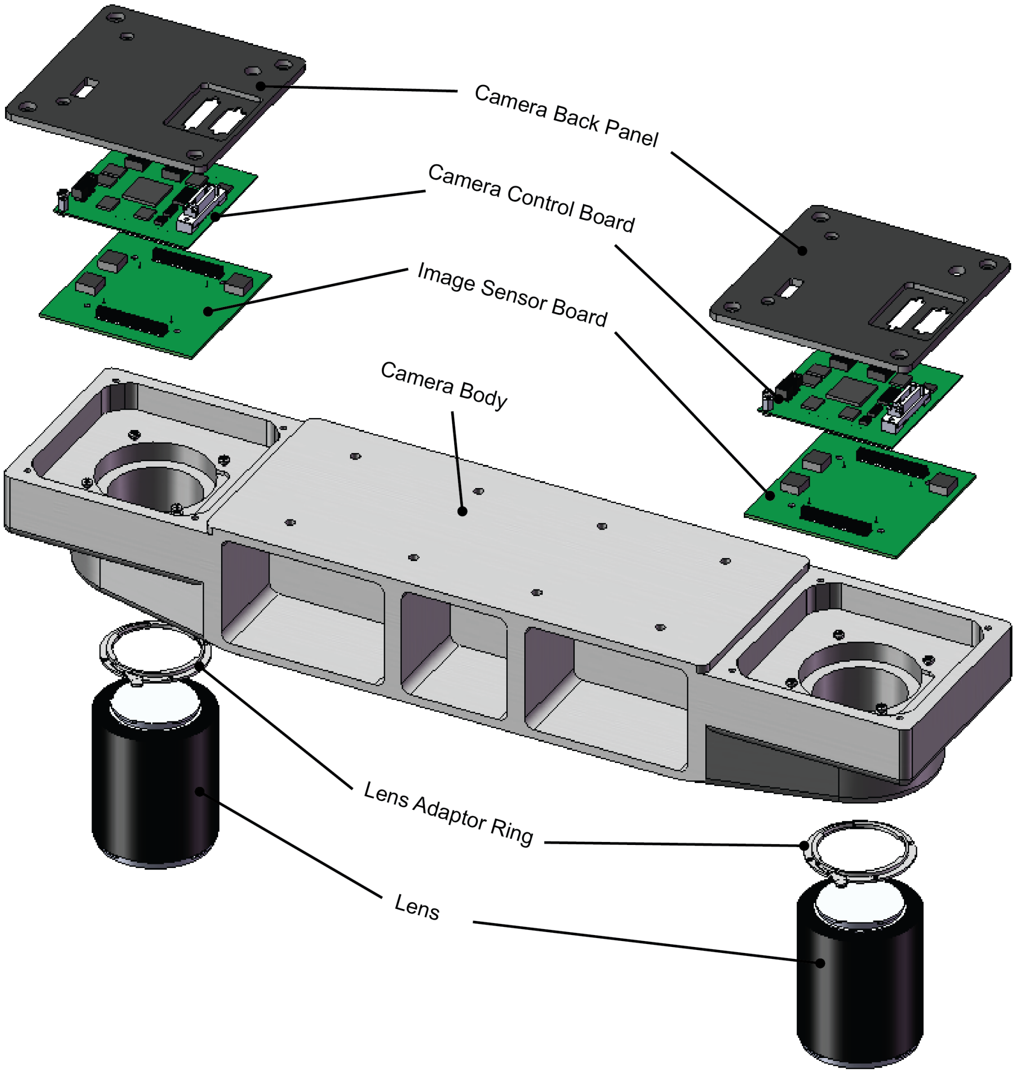


Figure 3.26: Hardware assembly of the prototype system

assembly is shown in Figure 3.26.

The custom electronics based on CMOS image sensor and FPGA architecture features high-performance image acquisition, high-speed camera interface and high-speed image processing. By using the ROI readout mechanism, over 10kHz position sampling can be achieved. The active target is designed, providing a uniform and stable radiation intensity across the measurement volume. Optical lens are carefully selected to satisfy the design goals as well as the selected image sensor. To minimize the effect of thermal expansion and external vibration on system geometry, a solid camera body is designed with optimized geometry parameters, material and mechanical structure. The specifications of system hardware are listed in Table 3.6.

Table 3.6: Hardware specifications of the prototype system

Parameters	Specification
Image Sensor Type	CMOS
Pixel Rate	66MHz
Raw image data resolution	14-bit
Camera Interface	Camera Link
Power Consumption	5W for two camera modules
Position Tracking System Weight	14kg
Marker Type	Active infrared LED

Chapter 4

System Calibration

In order to achieve high position accuracy, vision-based position sensing systems are usually calibrated to determine their model parameters. In this chapter, the calibration of the vision system is presented in detail. Section 4.1 analyzes the primary error sources in the system. Section 4.2 and Section 4.3 present the calibration methods for each error source, including the FPN calibration of image sensor and the calibration of optical and mechanical model parameters.

4.1 System Limitations

The system is designed and implemented based on a two-camera architecture. Ideally, each camera module is modeled as a linear system. The target position in pixel coordinate is obtained through sub-pixel target position interpolation. The 3D position of the target in world coordinate is reconstructed by triangulation based on the target positions in each camera module.

In the ideal situation, there is no 3D reconstruction error if the sub-pixel target position interpolation is perfect and all the model parameters are known accurately, but it is not true for a real world application. Image sensor has noise problems in photon-to-electron conversion and electron-to-voltage conversion, and these noises might be spatial-varying or temporal-varying, as well as temperature-dependent. The real target image is contaminated by the noise from image sensor and its peripheral circuits, therefore the image SNR is reduced which limits the accuracy of sub-pixel target position interpolation. The mechanical installation always has errors, causing discrepant geometry parameters from the nominal design. In addition, as a result of imperfect manufacturing and assembly, a real lens introduces optical distortions. Especially when a large measurement volume is required, the wide-angle lenses are used which usually suffer from severe optical distortions.

In order to achieve high accuracy in 3D reconstruction, these error sources must be modeled and characterized. The block diagram of system is shown in Figure 4.1. The goal of system calibration is to identify and extract the intrinsic and extrinsic parameters in the imaging model and use them in 3D position reconstruction model so that more accurate 3D position is achieved.

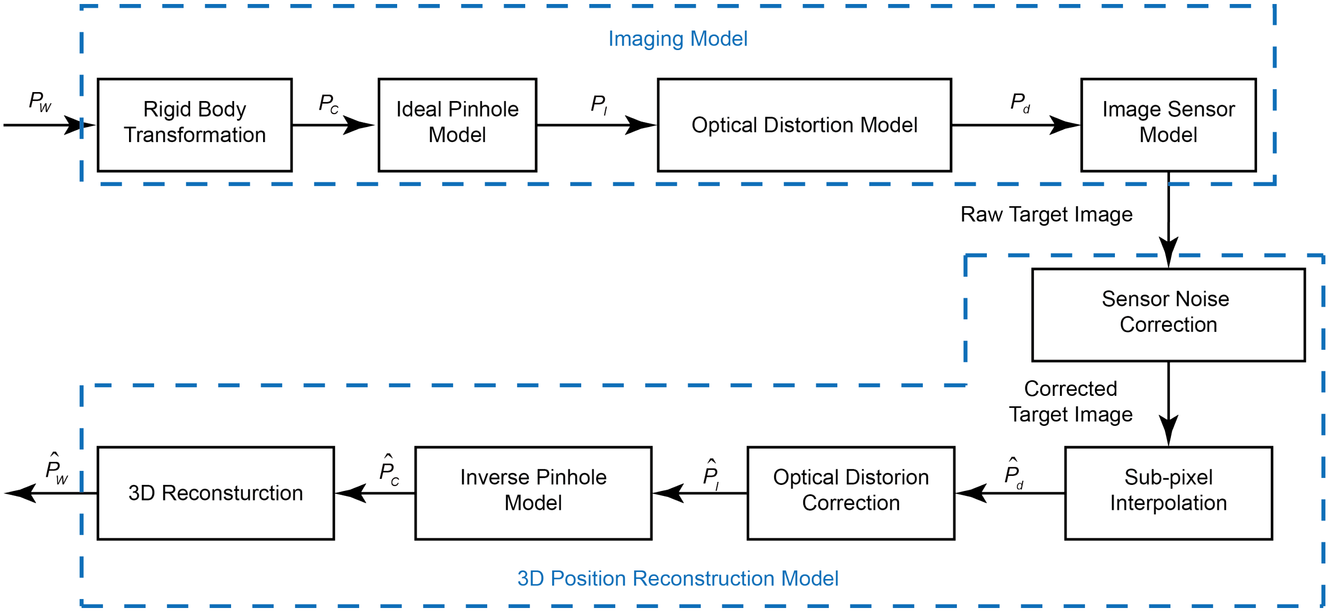


Figure 4.1: Block diagram of system

4.2 Fixed Pattern Noise of Image Sensor

Fixed pattern noise (FPN) is a spatial non-uniformity in an image sensor with multiple pixels. This problem is caused by variations in the pixel size, material or interference with local circuitry. FPN usually refers to two parameters: dark signal non-uniformity (DSNU) and photo response non-uniformity (PRNU) (shown in Figure 4.2). DSNU is the offset from the average across the image sensor with no external illumination. The DSNU is caused by the mismatch between photodiode leakage currents of pixels. The other type of FPN is PRNU, which describes the non-uniform gain between optical power on a pixel versus the electrical signal output. It can be characterized as the local, pixel-dependent photo response nonlinearity and is often simplified as a single value measured at almost saturation level to permit a linear approximation of the non-linear pixel response [39].

Figure 4.3 shows the model of image capture in a CMOS image sensor. From Figure 4.3, the raw output of image sensor is derived as

$$I_{raw} = \left(PRNU + SN_{ph} + K\right) \times L + SN_{dark} + SN_{read} + DSNU \tag{4.1}$$

where I_{raw} is the raw output of the image sensor, K is the conversion gain, $SN_{ph}(I)$ is the photon shot noise, SN_{dark} is the dark-current shot noise, SN_{read} is the readout noise from the signal conditioning and ADC sampling, and L is the light irradiance of illumination. The photon shot noise SN_{ph}, dark-current shot noise SN_{dark} and readout noise SN_{read} are temporally variant. On the other hand, FPN is spatially variant [40].

Although FPN does not change appreciably across a series of captures, it varies with operation conditions. V. Scheinder [41] and D. Joseph [42] investigated the temperature features of FPN in CMOS image sensor and concluded that the FPN in CMOS image sensor is dependent on the operating

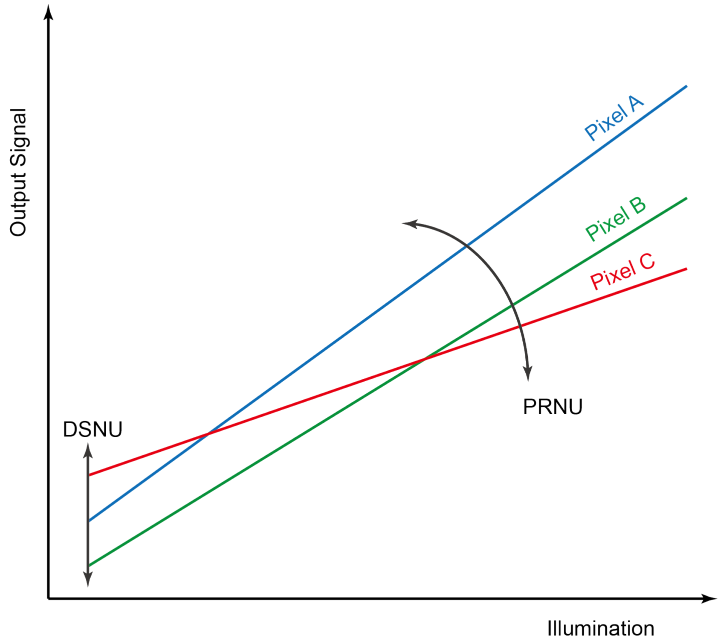


Figure 4.2: DSNU and PRNU illustration adapted from [29]

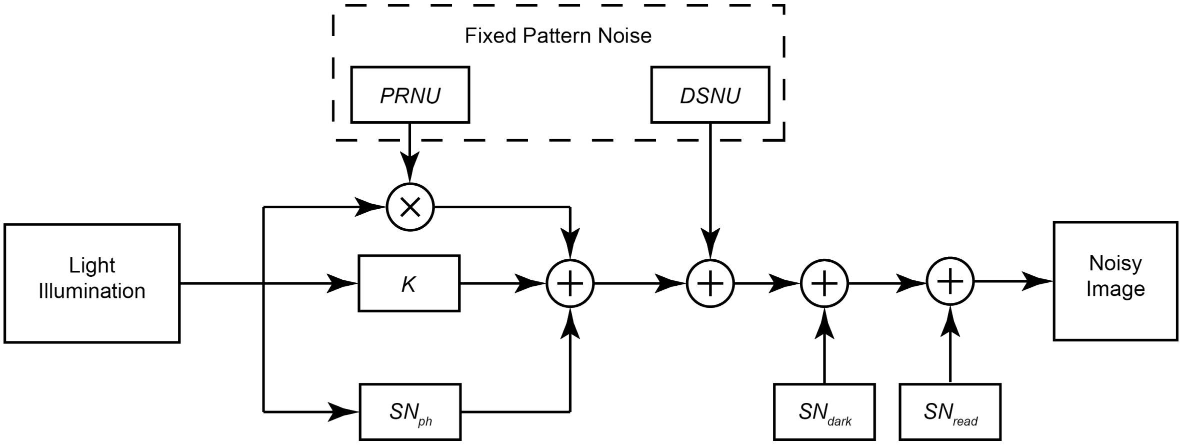


Figure 4.3: Noise model of an image sensor adapted from [40]

temperature conditions. In addition, CMOS image sensor has the common problem of high pixel storage node leakage (PSNL). Due to this problem, pixels that are read later in an image have higher leakage. That is, the readout speed of the image sensor affects the FPN. Therefore, FPN is sensitive to the operating conditions of the CMOS image sensor so that the FPN correction of one image is effective only when this image and the FPN data are collected under the same condition.

FPN is commonly corrected by flat-field correction that uses DSNU and PRNU to linearly interpolate pixel output. The DSNU is measured by analyzing a set of images captured in dark condition with no illumination. By setting $L = 0$, Equation 4.1 becomes

$$I_{dark} = DSNU + SN_{dark} + SN_{read}. \tag{4.2}$$

Based on the assumption that SN_{dark} and SN_{read} are a temporal white noise with zero mean, the temporal averaging of these dark images removes the term SN_{dark} and SN_{read}, yielding

$$\bar{I}_{dark} = DSNU. \tag{4.3}$$

The PRNU is obtained by exposing the pixel array to the same illumination which makes the pixel output signal close to its saturation. Setting the saturation illumination $L = L_{bright}$ in Equation 4.1 yields

$$I_{bright} = \big(PRNU + SN_{ph}(L_{bright}) + K\big) \times L_{bright} + SN_{dark} + SN_{read} + DSNU. \tag{4.4}$$

Similarly, $SN_{ph}(L_{bright})$ is treated as a temporal white noise with zero mean. Temporal averaging of multiple bright images and substituting Equation 4.3 into Equation 4.4 give

$$\bar{I}_{bright} = (PRNU + K) \times L_{bright} + \bar{I}_{dark}. \tag{4.5}$$

Therefore, the PRNU is derived as

$$PRNU = \frac{\bar{I}_{bright} - \bar{I}_{dark}}{L_{bright}} - K. \tag{4.6}$$

Thus, based on Equation 4.3 and Equation 4.6, the image after FPN correction is expressed as

$$I_{corrected} = \frac{I_{raw} - \bar{I}_{dark}}{\bar{I}_{bright} - \bar{I}_{dark}} A = \frac{L}{L_{bright}} A \tag{4.7}$$

where A is range of the image after FPN correction. It is clear from Equation 4.7 that the image after FPN correction is proportional to the input light illumination.

4.3 Calibration of Optical and Mechanical Parameters

In a vision-based position sensing system, optical and mechanical parameters are coupled with each other. The calibration procedure of optical and mechanical parameters in a vision system is called camera calibration. Camera calibration in the context of 3D position sensing is the process of determining internal camera geometric and optical characteristics (intrinsic parameters), and the 3D position and orientation of the camera frame relative to a certain world coordinate system (extrinsic parameters)[18]. Camera calibration is an essential component of photogrammetric measurement, especially in high accuracy close-range measurement. Accurate camera calibration procedures are a necessary prerequisite of the extraction of precise and reliable 3D position information from images [43].

Because of imperfections in the design and assembly of lenses composing the camera optical system, the lenses used in real world application cannot be modeled as an ideal pinhole model. When the light passes through optical lens, the light path is distorted, causing an offset position P_I in the image coordinate. Furthermore, the true geometry parameters in the rigid body transformation are unknown because of the mechanical installation errors. The optical axis of lens may not be orthogonal to the image sensor plane and perfectly intersects with the plane at the nominal center of image sensor. Since the 3D position reconstruction relies on the triangulation using two cameras, 3D position error is introduced by mechanical installation error between two cameras.

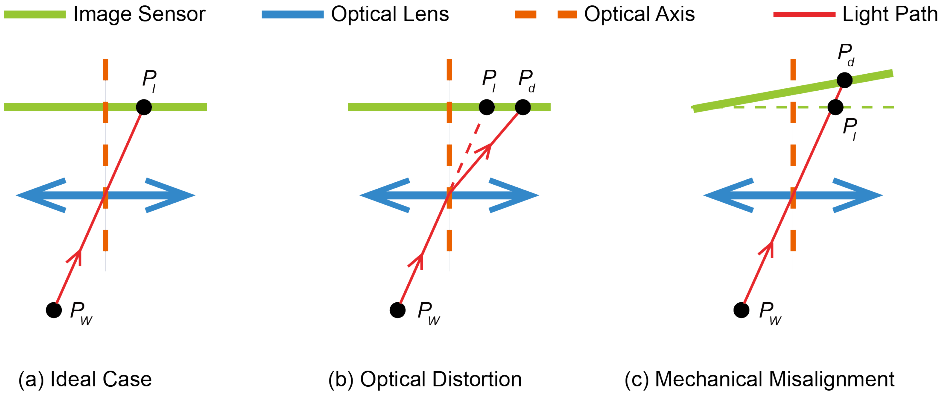


Figure 4.4: Optical and mechanical error sources in a vision system

As a result of the imperfection of lens and mechanical installation, camera calibration is necessary in order to achieve high position accuracy in vision-based position sensing applications. For best serving this purpose, the camera calibration should meet the following criteria [18]:

- **Autonomous**: The calibration procedure should not require operator intervention, such as giving initial values for certain parameters, or choosing certain parameters manually.
- **Accurate**: The camera calibration technique should have the potential of meeting the accuracy requirements. This requires an accurate theoretical modeling of the imaging process as well as calibration algorithms.
- **Versatile**: The calibration technique should operate uniformly and autonomously for a wide range of system setups.

In this section, the prior art of camera calibration methods are investigated and the proposed calibration method is then presented. In Section 4.3.1, the state-of-art camera calibration methods are reviewed and discussed. The proposed camera calibration method is covered in Section 4.3.2. Section 4.3.3 gives the simulation results to preliminarily verify the performance of the proposed calibration method.

4.3.1 Camera Calibration Methods Review

Camera calibration is an important issue in the photogrammetry community. With the increasing need for higher accuracy measurement, it has attracted lots of research effort [15][44][45][18][17][19]. To estimate the model parameters of a stereo-vision system, the traditional methods convert it into multiple single camera calibration problems and solve them individually. The existing techniques for single camera calibration can be classified into the following categories [17]:

- **Closed-form solution**: Model parameters are estimated based on a closed-form solution (e.g., [15] [46]). Intermediate parameters are defined in terms of the original parameters. The intermediate parameters are computed by solving linear equations, and the original parameters are determined from those intermediate parameters [17]. This type of technique is fast because there is no iterative optimization. However, the lens distortion is not incorporated so that the lens distortion effects cannot be corrected. Furthermore, the actual constraints in the intermediate parameters are not considered because of the objective to construct a non-iterative algorithm. Therefore, the solution does not meet the constraints in the presence of measurement noise.

- **Direct nonlinear minimization**: Nonlinear model is established for the camera system. The parameters are searched by using an iterative algorithm with the objective to minimize residual errors of some equations. Many types of lens distortions can be incorporated in this technique. Further, good estimation of model parameters can be achieved if the imaging model is accurate and a good convergence is reached in optimization iteration. However, since the algorithm is iterative, the optimization procedure may end up with a local optimum when the initial guess is bad. In addition, the optimization can be unstable if the procedure of iteration is badly designed. The harmful interactions between nonlinear and linear parameters can lead to divergence or false solution.

- **Multi-step methods**: In this type of method, a direct solution for most of the calibration parameters and some iterative solution for the other parameters are conducted in sequential steps (e.g., [19] [17] [18]). The main advantages of this type of methods is that the major part of model parameters can be derived from a closed-form solution, and the number of parameters to be estimated through iterations is relatively small. However, some existing techniques in this category have their own drawbacks. For instance, Tsai's method [18] can only handle the radial lens distortion and cannot be extended to other lens distortion models. The solution is not optimal since the information provided by the test points has not been fully utilized.

In the rest part of this section, some state-of-art camera calibration methods are reviewed, including direct linear transformation (DLT) method [15], Tsai' method [18] and Heikkila's method [19].

Direct Linear Transformation Method

The DLT method is one of the classical procedures which apply space points calculation from image coordinates. The DLT method is first introduced by Abdel-Aziz and Karara [15] in 1971 and was refined by Hatze [44] in 1988 and Gazzani [45] in 1993. The DLT method gives a closed-form solution which uses eleven intermediate parameters to represent the mathematical relationship between world coordinate (X_W, Y_W, Z_W) and pixel coordinate (u, v):

The intermediate parameters are computed by solving linear equations, and then the model parameters are retrieved from these intermediate parameters. To improve the accuracy of calibration, test points should be included as many as possible and spread uniformly throughout the calibration volume.

Since there is no iterative optimization in the DLT method, the computation speed of this algorithm is fast. However, the DLT method does not guarantee the physical constraints on model parameters, such as the orthogonality of the rotation matrix R. Therefore the calibration accuracy degrades in the presence of sub-pixel interpolation errors. In addition, the DLT method requires the test points to be non-coplanar and calibration volume to be large enough in order to cover the space of measurement volume.

Tsai's Method

R.Y. Tsai proposed a versatile camera calibration method based on a two-stage technique in 1987 [18]. This two-stage technique is aimed at efficient computation of camera external position and orientation relative to object reference coordinate system as well as radial lens distortion, and image scanning parameters.

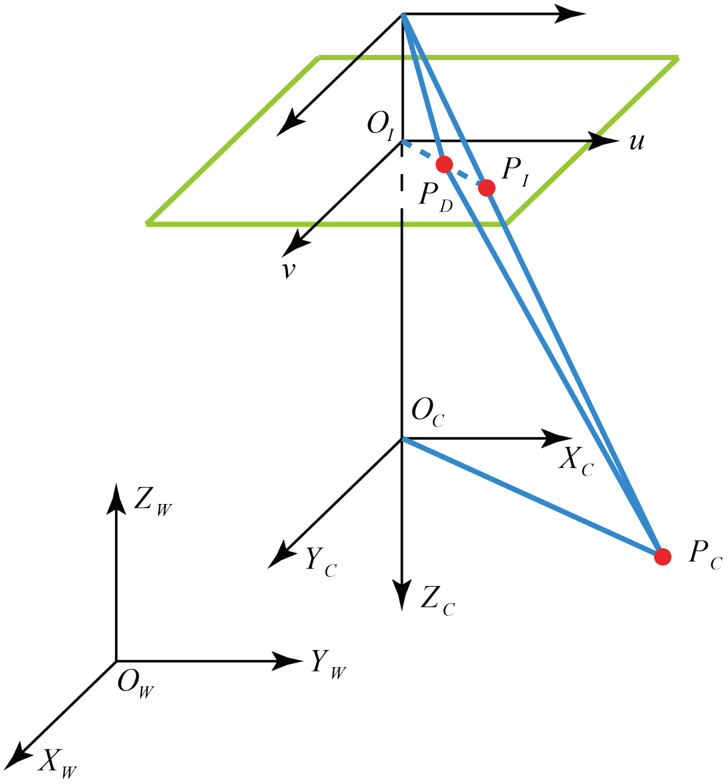


Figure 4.5: Tsai's camera model with perspective projection and radial distortion adapted from [18]

Tsai's method first tries to obtain the estimation of as many parameters as possible using linear least-square fitting methods. In the initial stage, constraints of parameters are not enforced. This does not affect the final results, since these estimated parameters are used as the initial values for the optimization in second stage. In the subsequent stage, the rest of parameters are obtained using a nonlinear optimization method that searches the best fit between the observed image points and those predicted from the identified imaging model. Parameters estimated in the first stage are refined in the process.

The advantage of Tsai's method is that an initial guess of some parameters is given by the first linear stage and fewer optimized parameters in the second nonlinear iteration stage, leading to fast computation speed. Additionally, radial lens distortion is considered here and better 3D reconstruction accuracy is achieved by implementing radial distortion correction. Tsai's method is able to operate uniformly for a wide range of setups and applications.

However, in Tsai's method, only a small part of model parameters are optimized in the iterative optimization. Other parameters are given by the initial linear stage so that the final results are unavoidably harmed by lens distortion. Some parameters, such as the principal point (u_0, v_0) and image sensor scaling factor S_x and S_y, are assumed to be known values provided by the image sensor manufacturer. In addition, this method can only handle the first-order radial distortion and cannot be extended to other types of lens distortion. The harmful interaction between linear parameters and nonlinear parameters in the nonlinear iteration may gives a bad estimation of model parameters.

Heikkila's Method

Heikkila presented a multi-step calibration procedure that is an extension to the two-step method in 1997 [19]. Similar to Tsai's method, a closed-form solution is used to obtain the initial values of linear parameters. All parameters are optimized in a nonlinear iteration in the second step. There are additional steps for correcting the distorted image coordinates. The image correction is performed with an empirical inverse model that accurately compensates for lens distortion.

Compared to Tsai's lens distortion model, Heikkila incorporated radial and decentering distortion in his model to correct the lens distortion.

In the iteration, Levenberg-Marquartdt (LM) optimization method is used to solve the nonlinear least-square problem. However, LM method is a differential-based algorithm that is susceptible to be trapped in local minimum, especially when a bad initial guess is given. Further, the linear and nonlinear parameters are coupled in nonlinear iteration. The Heikkila method designed a cost function in the nonlinear optimization which tries to minimize the 2D reprojection error in pixel coordinate. Considering the errors of sub-pixel target position interpolation, the accuracy of estimation results is decreased.

4.3.2 Proposed Camera Calibration Method

From the literature review in Section 4.3.1, it clearly shows that lots of effort have been devoted into the research of single camera calibration method. Single camera calibration method can be extended to a wide type of camera calibration problems. The traditional way of calibrating a stereo-vision system converts the original problem into multiple single camera calibration problems where the model parameters of each camera are solved individually. As a result, the calibration of a multi-camera system is simplified.

However, single camera calibration methods have their limitations. Because the 3D position in world coordinate cannot be reconstructed with a single camera, the cost function in single camera calibration methods minimizes the 2D reprojection error in pixel coordinate (shown in Figure 4.6). The cost function in single camera calibration methods is express as

$$f(\hat{P}) = \sum_{i=1}^{N} \left\{ \left[u_i - \hat{u}_i(\hat{P})\right]^2 + \left[v_i - \hat{v}_i(\hat{P})\right]^2 \right\} \tag{4.8}$$

where (u_i, v_i) is the measured target position in pixel coordinate, and $\left(\hat{u}_i(\hat{P}), \hat{v}_i(\hat{P})\right)$ is the estimated

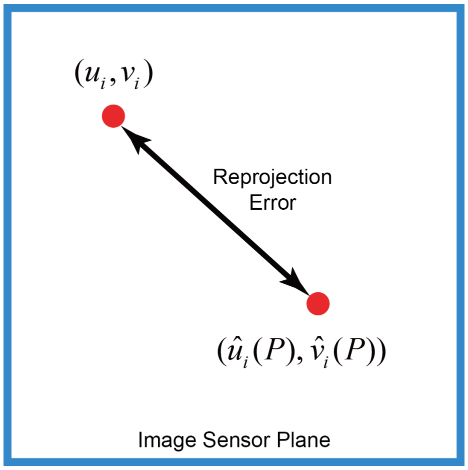


Figure 4.6: Illustration of 2D reprojection error in image coordinate

target position in pixel coordinate based on the estimated camera parameter set $\hat{P}$. The 2D reprojection algorithms are based on the imaging model presented in Section 2.1.

It is known that the depth information of a 3D point is lost in the projection from 3D space to 2D image. therefore the information provided by a 3D test point is not fully utilized. Limited by the SNR of target image, the sub-pixel interpolation is not accurate. The target position in pixel coordinate is subject to the error of sub-pixel interpolation algorithm, and therefore the model parameters estimated from a single camera calibration method are contaminated with error and cannot guarantee a good 3D reconstruction result. Especially when these estimated model parameters are used for 3D position reconstruction, larger reconstruction error is found in depth direction.

For stereo-vision systems, calibrating individual camera separately and combining the results to reconstruct the 3D position does not guarantee an optimized stereo-vision system. With the goal of calibrating a stereo-vision system, it is more reasonable to use the 3D reconstruction error as the cost function (shown in Figure 4.7). The proposed cost function in optimization is expressed as

$$f(\hat{P}) = \sum_{i=1}^{N} \left\{ [X_i - \hat{X}_i(\hat{P})]^2 + [Y_i - \hat{Y}_i(\hat{P})]^2 + [Z_i - \hat{Z}_i(\hat{P})]^2 \right\} \tag{4.9}$$

where (X_i, Y_i, Z_i) is the measured 3D position of test points in world coordinates, and $\left(\hat{X}_i, \hat{Y}_i, \hat{Z}_i\right)$ is the estimated 3D position of test points based on estimated camera parameter set $\hat{P}$. The 3D position reconstruction algorithm is based on a linear transformation combined with a nonlinear correction of lens distortions (discussed in Section 2.3). Usually, the measured 3D position (X_i, Y_i, Z_i) is given by high accuracy machines, such as CMM, therefore the accuracy of the test point position in 3D world coordinate is guaranteed.

In most single camera calibration methods, the linear parameters and nonlinear parameters are coupled in the optimization process. The harmful interaction between linear and nonlinear parameters may introduce instability issue of the nonlinear iteration. Consider that most existing nonlinear methods minimize the cost function using variants of conventional gradient-descent optimization, like Newton method and Levenberg-Marquardt (LM) method. These techniques have well-known problems plagu-

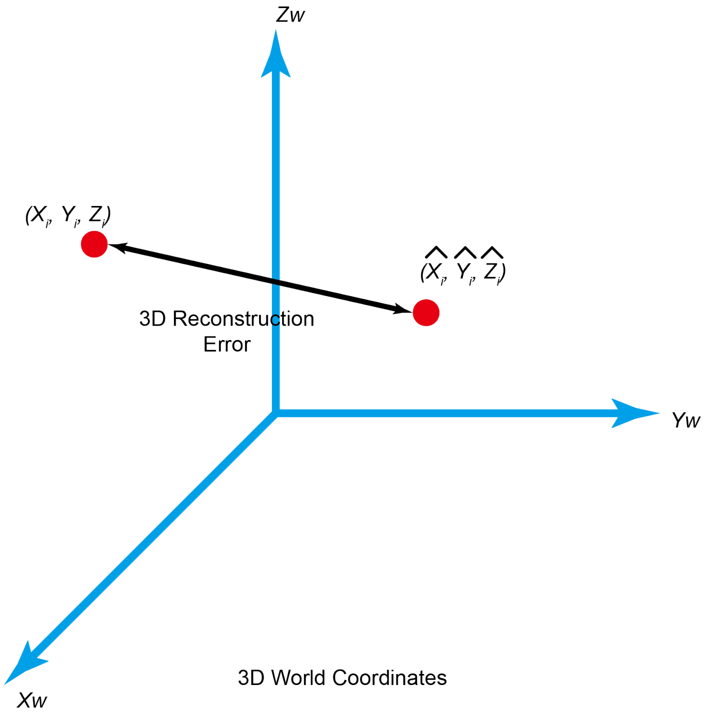


Figure 4.7: Illustration of 3D reconstruction error in world coordinate

ing these differentiation-based methods, such as poor convergence and susceptibility to be trapped in local minimum. This problem is severe when the linear and nonlinear parameters are coupled in the optimization iteration. Therefore, the risk of local rather than global optimization might be severe. On the other hand, this local optimization problem is more severe when a good initial guess cannot be given. Thus, multi-step calibration method must be used in order to provide a proper initial guess of parameters in the first stage.

In camera calibration, the linear parameters are referred to those parameters used in linear transformation, and nonlinear parameters are the other parameters that represent the nonlinear lens distortions. In order to suppress the harmful interaction between linear and nonlinear parameters, the optimization of linear and nonlinear parameters is decoupled in the proposed calibration method. Figure 4.8 shows the flow chart of the proposed calibration method. The initial guess of linear parameters are given by a linear optimization in the first step. In order to obtain a good initial guess of linear parameters, the test points close to image center are used so that the lens distortion effect is minimized. In the second step, the linear parameters are fixed and only nonlinear parameters are optimized. Further in the next step, the nonlinear parameters are fixed and linear parameters are optimized. The optimization iteration is terminated when the 3D reconstruction error satisfies pre-defined constraints.

4.3.3 Simulation Results

In this section, the proposed camera calibration method (named NK method) is compared with the Heikkila method based on the simulation environment. The stereo-vision system to be calibrated is based on a two-camera configuration. Further comparisons by experiment are presented in Section 5.4.

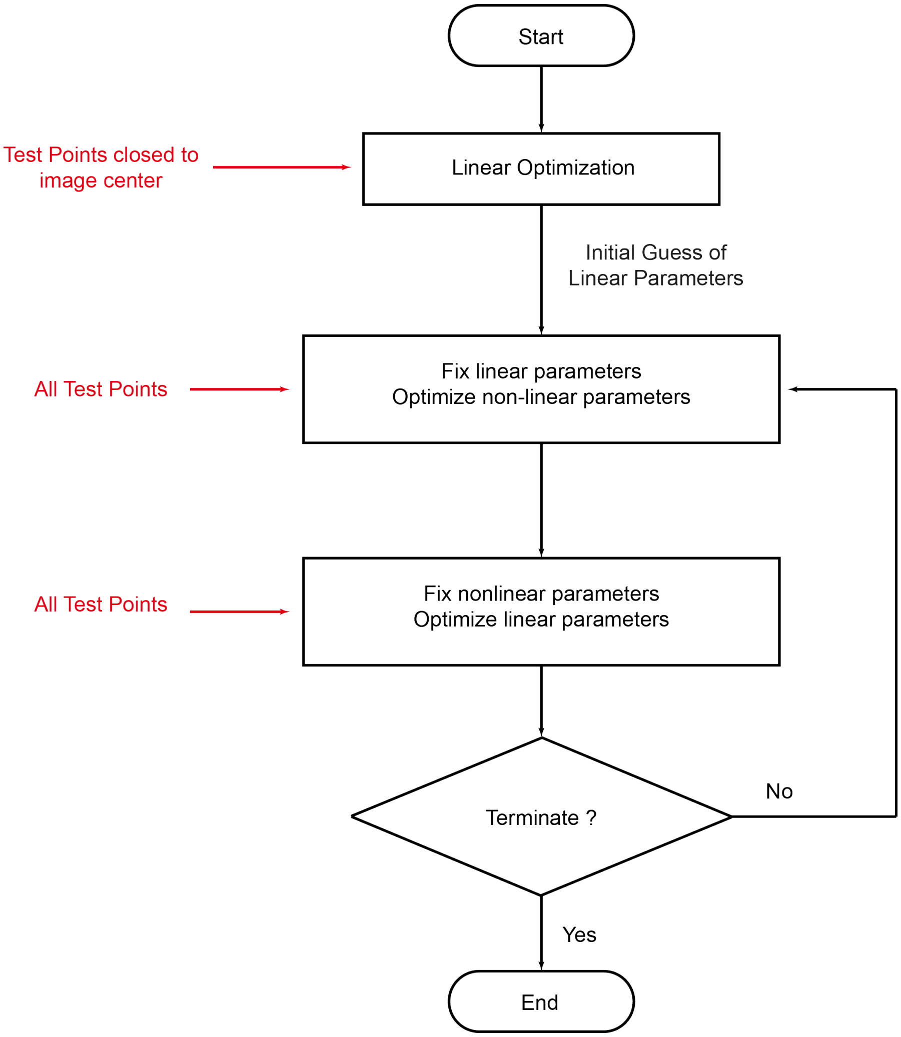

Start
Test Points closed to image center
Linear Optimization
Initial Guess of Linear Parameters
All Test Points
Fix linear parameters
Optimize non-linear parameters
All Test Points
Fix nonlinear parameters
Optimize linear parameters
Terminate ?
No
Yes
End

Figure 4.8: Flow chart of the proposed calibration method

In the simulation environment, 192 test points in 3D world coordinate are used, and their target locations in pixel coordinate are obtained based on the imaging model described in Section 2.1. These test points are evenly distributed in a cubic space of 500mm by 500mm by 10mm. To evaluate the calibration performance, the estimated camera parameters are compared with their ground true values which are pre-defined in the simulation. In addition, the 3D reconstruction errors are also compared.

Three cases are studied in this section. In case 1, the lens distortion is assumed to be zero, and there is assumed to be no sub-pixel interpolation error. Case 2 compares the calibration performance in the presence of lens distortion but no sub-pixel interpolation error. In case 3, both lens distortion and sub-pixel interpolation error are considered.

Case 1

Case 1 is a very ideal case where the lens is assumed to be perfect without optical distortion and the sub-pixel interpolation algorithm is perfectly accurate. In this scenario, the imaging model is simplified to a linear model so that both calibration methods should give a perfect estimation with zero error. Indeed, the simulation results demonstrate that both methods give zero estimation error of camera parameters. The estimation results are shown in Table 4.1 and Table 4.2. Since the estimation of model parameters is perfect, zero 3D reconstruction error is obtained.

Table 4.1: Simulation case 1 - calibration results of Camera 0

Camera Parameters	True Value	Heikkila's Method	NK Method
α [deg]	2	2.000	2.000
β [deg]	5	5.000	5.000
γ [deg]	-179	-179.000	-179.000
T_x [mm]	100	100.000	100.000
T_y [mm]	200	200.000	200.000
T_z [mm]	-900	-900.000	-900.000
$f_x = dS_x$	2141.67	2141.67	2141.67
$f_y = dS_y$	2141.67	2141.67	2141.67
u_0 [pixel]	1024	1024.000	1024.000
v_0 [pixel]	1024	1024.000	1024.000

Table 4.2: Simulation case 1 - calibration results of Camera 1

Camera Parameters	True Value	Heikkila's Method	NK Method
α [deg]	2	2.000	2.000
β [deg]	-20	-20.000	-20.000
γ [deg]	-179	-179.000	-179.000
T_x [mm]	100	100.000	100.000
T_y [mm]	500	500.000	500.000
T_z [mm]	-900	-900.000	-900.000
$f_x = dS_x$	2141.67	2141.67	2141.67
$f_y = dS_y$	2141.67	2141.67	2141.67
u_0 [pixel]	1024	1024.000	1024.000
v_0 [pixel]	1024	1024.000	1024.000

Case 2

Compared to case 1, the lens distortion of lens is considered in case 2 but the sub-pixel interpolation error is assumed to be zero. As a result of lens distortion, the imaging model is nonlinear so that nonlinear iteration is required to solve the model parameters. Considering the numerical error in nonlinear iteration, the calibration method cannot obtain a perfect estimation of the model parameters. The estimation results are shown in Table 4.3 and Table 4.4. Based on these two tables, it clearly shows that the Heikkila method and the NK method achieves very close estimation of the camera parameters. Therefore, the RMS 3D position reconstruction error from each method are compared in Table 4.5. It clear shows that the NK method gives slightly better 3D reconstruction result than the Heikkila method when sub-pixel interpolation error is set to zero and only lens distortion is considered.

Table 4.3: Simulation case 2 - calibration results of Camera 0

Camera Parameters	True Value	Heikkila's Method	NK Method
α [deg]	2	2.000	1.999
β [deg]	5	4.999	5.000
γ [deg]	-179	-179.000	-179.000
T_x [mm]	100	99.990	99.997
T_y [mm]	200	200.001	200.005
T_z [mm]	-900	-899.986	-899.981
$f_x = dS_x$	2141.67	2141.652	2141.641
$f_y = dS_y$	2141.67	2141.651	2141.589
u_0 [pixel]	1024	1024.000	1024.000
v_0 [pixel]	1024	1024.000	1024.000
K_1 $[mm^{-2}]$	-1×10^{-4}	-1.005×10^{-4}	-1.005×10^{-4}
K_2 $[mm^{-4}]$	1.2×10^{-7}	1.026×10^{-7}	1.027×10^{-7}
P_1 $[mm^{-1}]$	2×10^{-5}	2.057×10^{-5}	2.063×10^{-5}
P_2 $[mm^{-1}]$	-1×10^{-5}	-1.034×10^{-5}	-0.999×10^{-5}

Case 3

In case 3, the sub-pixel interpolation error is considered. Here it is assumed that the sub-pixel interpolation error is a white noise with zero mean. Table 4.6 and Table 4.7 show the estimation results of each camera respectively, assuming a random sub-pixel interpolation error with 0.1 pixel standard deviation. It is clear that the estimation error of the Heikkila method increases when the sub-pixel interpolation error is considered. On the other hand, as a result of using 3D reconstruction error as the cost function, the NK method achieves better estimation of model parameters in the presence of sub-pixel interpolation error. This simulation result is consistent with the theoretical expectation in the previous section. To further verify the performance of the NK method, the 3D reconstruction error at different sub-pixel interpolation error level is investigated. In this simulation, varying sub-pixel interpolation error level is considered, and the calibration as well as the 3D reconstruction is conducted at each sub-pixel interpolation error level. From Figure 4.9, it clearly shows that the 3D reconstruction error of each method is very close when the sub-pixel interpolation error is small but the NK method gives slightly better 3D reconstruction result when the sub-pixel interpolation error keeps increasing.

Table 4.4: Simulation case 2 - calibration results of Camera 1

Camera Parameters	True Value	Heikkila's Method	NK Method
α [deg]	2	2.002	2.002
β [deg]	-20	-20.001	-20.001
γ [deg]	-179	-178.999	-178.999
T_x [mm]	500	500.018	499.992
T_y [mm]	200	200.002	199.996
T_z [mm]	-900	-900.071	-899.987
$f_x = dS_x$	2141.67	2141.800	2141.641
$f_y = dS_y$	2141.67	2141.800	2141.712
u_0 [pixel]	1024	1023.990	1024.000
v_0 [pixel]	1024	1023.990	1024.000
K_1 $[mm^{-2}]$	-1×10^{-4}	-1.006×10^{-4}	-1.005×10^{-4}
K_2 $[mm^{-4}]$	1.2×10^{-7}	1.038×10^{-7}	1.028×10^{-7}
P_1 $[mm^{-1}]$	2×10^{-5}	2.052×10^{-5}	2.013×10^{-5}
P_2 $[mm^{-1}]$	-1×10^{-5}	-1.032×10^{-5}	-1.100×10^{-5}

Table 4.5: Simulation case 2 - comparison of 3D position reconstruction error

	Heikkila's Method	NK Method
3D Error RMS [μm]	6.73	6.46

4.4 Summary

The primary limitations on position accuracy of vision-based position sensing systems come from mechanical installation error, lens optical distortion and noise of target image. FPN, including DSNU and PRNU, is a spatial noise of image sensor. A flat-field correction that uses DSNU and PRNU to linearly interpolate the local image non-uniformity is implemented in the system to correct the FPN of image sensor. Further, mechanical installation error and lens distortion introduce error in 3D reconstruction. The camera geometry and optical parameters can be estimated by camera calibration. Traditional camera calibration methods convert a stereo-vision calibration problem to multiple single camera calibration problems. As a result, the reprojection error in 2D pixel coordinate is used as the cost function in optimization process. Considering the error in sub-pixel target position interpolation, the parameter estimation of traditional methods is unavoidably contaminated by large error. The proposed calibration method uses the reconstruction error in 3D world coordinate and decouples the optimization of linear and nonlinear model parameters. The simulation results show that the proposed calibration method achieves more accurate calibration results than the Heikkila method, especially in the presence of large error in sub-pixel target position interpolation.

Table 4.6: Simulation case 3 - calibration results of Camera 0 (0.1 pixel sub-pixel interpolation error)

Camera Parameters	True Value	Heikkila's Method	NK Method
α [deg]	2	1.966	1.998
β [deg]	5	4.933	5.002
γ [deg]	-179	-179.002	-179.000
T_x [mm]	100	99.449	99.998
T_y [mm]	200	198.825	200.005
T_z [mm]	-900	-891.199	-899.971
$f_x = dS_x$	2141.67	2121.200	2141.606
$f_y = dS_y$	2141.67	2120.851	2141.600
u_0 [pixel]	1024	1018.490	1024.000
v_0 [pixel]	1024	1023.300	1024.000
K_1 [mm^{-2}]	-1×10^{-4}	-0.98×10^{-4}	-1.005×10^{-4}
K_2 [mm^{-4}]	1.2×10^{-7}	0.89×10^{-7}	1.029×10^{-7}
P_1 [mm^{-1}]	2×10^{-5}	2.421×10^{-5}	2.068×10^{-5}
P_2 [mm^{-1}]	-1×10^{-5}	-0.47×10^{-5}	-0.980×10^{-5}

Table 4.7: Simulation case 3 - calibration results of Camera 1 (0.1 pixel sub-pixel interpolation error)

Camera Parameters	True Value	Heikkila's Method	NK Method
α [deg]	2	1.961	2.003
β [deg]	-20	-20.105	-20.001
γ [deg]	-179	-179.011	-178.999
T_x [mm]	500	504.058	499.990
T_y [mm]	200	200.405	199.991
T_z [mm]	-900	-906.671	-899.980
$f_x = dS_x$	2141.67	2159.800	2141.641
$f_y = dS_y$	2141.67	2158.609	2141.702
u_0 [pixel]	1024	1023.550	1023.990
v_0 [pixel]	1024	1025.890	1024.000
K_1 [mm^{-2}]	-1×10^{-4}	-0.98×10^{-4}	-1.006×10^{-4}
K_2 [mm^{-4}]	1.2×10^{-7}	1.007×10^{-7}	1.029×10^{-7}
P_1 [mm^{-1}]	2×10^{-5}	3.711×10^{-5}	2.015×10^{-5}
P_2 [mm^{-1}]	-1×10^{-5}	-1.023×10^{-5}	-1.190×10^{-5}

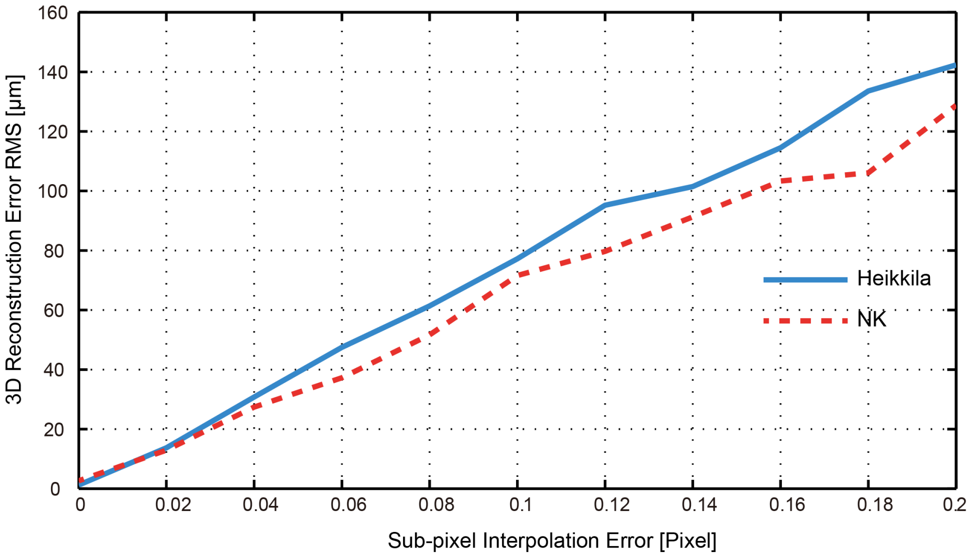


Figure 4.9: 3D reconstruction error VS. sub-pixel interpolation error

Chapter 5

Experimental Results

This chapter presents the experimental characterization of the system. The FPN of image sensor is characterized, and the effectiveness of the FPN correction method is demonstrated by experiment. The position sampling frequency for different ROI sizes and target quantities is characterized. In addition, the stereo-vision system is calibrated to determine its model parameters using a CMM. The 3D position reconstruction accuracy is characterized and examined within the measurement volume. The vision-based position sensing system is integrated as the metrology solution for a prototype of planar motion stage.

5.1 FPN Characterization and Correction

The image sensor is not perfect. In order to achieve a high quality image, the FPN of the image sensor must be characterized and corrected. Based on theoretic analysis in Section 4.2, multiple black images are collected and averaged to obtain the DSNU data of the image sensor. The DSNU data is shown in Figure 5.1.

To verify the effectiveness of DSNU correction, Figure 5.2 compares the histogram of a raw black image and the black image after DSNU correction. It is clear that DSNU correction significantly reduced the spatial standard deviation of black image.

Theoretically, in order to obtain the PRNU of every pixel, a uniform light is required to illuminate the entire the image sensor. However, it is very difficult to generate a uniform light illumination on the entire area of the image sensor. As a result, a pixel-by-pixel PRNU correction is difficult to achieve.

Considering that there are two separate analog channels for even and odd pixels, the average gain non-uniformity between even and odd pixels is obtainable. Therefore, the correction of the gain non-uniformity is downgraded from a pixel-by-pixel correction to an even-odd pixel correction. The image sensor model described in Section 4.2 gives

$$\begin{aligned} I_{Even}(u,v) &= (Gain_{Even}(u,v)) \times L(u,v) + SN_{dark} + SN_{read} + DSNU(u,v) \\ I_{Odd}(u,v) &= (Gain_{Odd}(u,v)) \times L(u,v) + SN_{dark} + SN_{read} + DSNU(u,v) \end{aligned} \tag{5.1}$$

Assuming that the ADC readout noise SN_{read} and dark current shot noise SN_{dark} are white noise with zero mean, the average gain ratio between even and odd pixels is calculated as

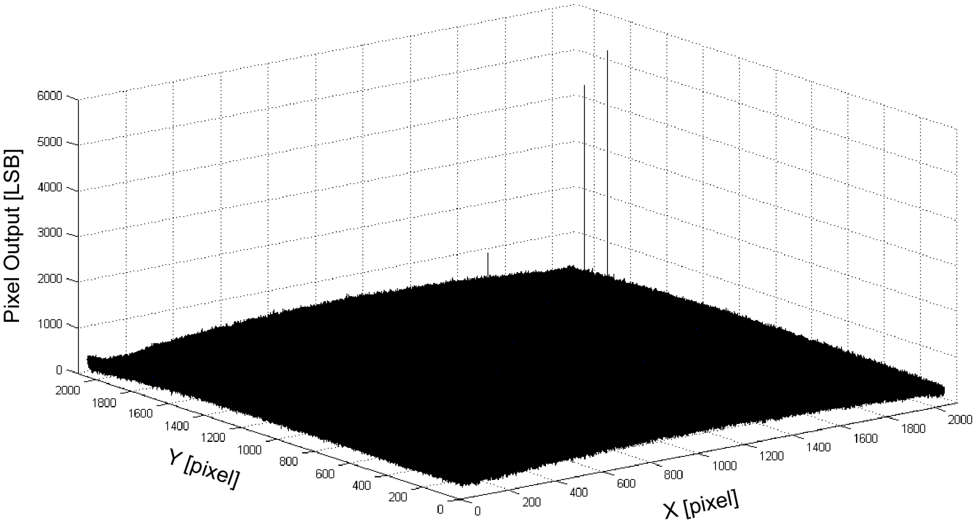


Figure 5.1: DSNU of the image sensor at 10μs exposure time

$$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}} = \frac{\sum (I_{Even}(u,v) - DSNU(u,v))}{\sum (I_{Odd}(u,v) - DSNU(u,v))} \tag{5.2}$$

where $\overline{Gain_{Even}}$ and $\overline{Gain_{Odd}}$ are the average gain of even and odd pixels respectively; I_{Even} and I_{Odd} are the raw pixel values of even and odd pixels, respectively; $DSNU_{Even}$ and $DSNU_{Odd}$ are the odd and even pixels' DSNU, respectively.

To obtain the average gain ratio between even and odd pixels, the image sensor is illuminated by a light source which generates a continuous light intensity distribution across the entire sensor area. Multiple bright images are obtained, and then the average gain ratio between even and odd pixels is calculated based on Equation 5.2. Ten full frame bright images are collected and the average gain ratio between even and odd pixels is calculated for each bright image (shown in Table 5.1).

Table 5.1: Average gain ratio between even and odd pixels before compensation

Image Index	1	2	3	4	5
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$	1.0209	1.0204	1.0206	1.0203	1.0203
Image Index	6	7	8	9	10
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$	1.0196	1.0197	1.0193	1.0193	1.0189

From Table 5.1, it is clear that there is about 2% gain difference between even and odd pixels. After implementing the average gain ratio compensation, another 5 full bright images are collected and the average gain ratio between even and odd pixels are investigated (shown in Table 5.2). After compensation, the average gain difference between even and odd pixels are reduced below 0.1%.

The compensation results shown in Table 5.2 are for the full frame image scale. However, the

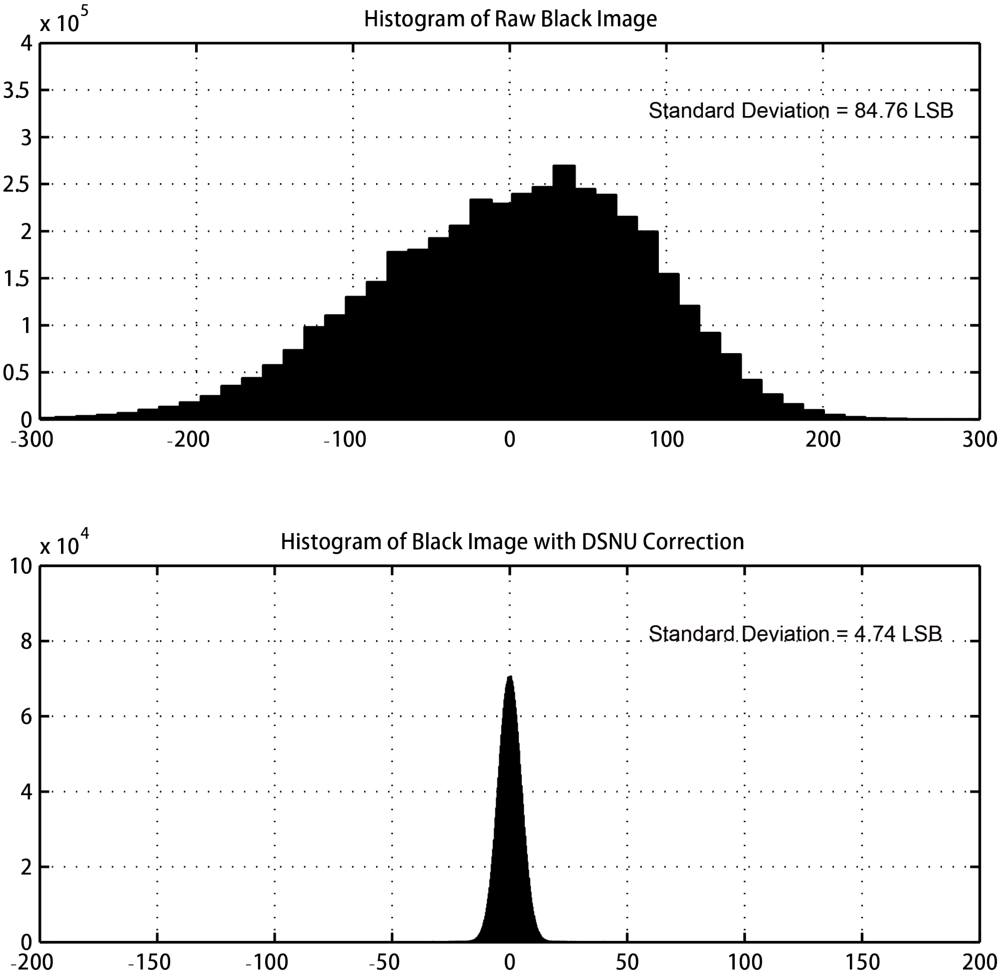


Figure 5.2: Histogram comparison between a raw black image and the image after DSNU correction (The mean value is removed)

Table 5.2: Average gain ratio between even and odd pixels after compensation

Image Index	1	2	3	4	5
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$	0.9992	0.9995	1.0005	1.0001	0.9997

image processing in the system is ROI-based. In order to verify the effectiveness of this average gain ratio compensation, the average gain ratio between even and odd pixels in the ROI scale is investigated. Nine ROI groups located in different areas of the image sensor are read out (shown in Figure 5.3). The average gain ratio before compensation and after compensation is shown in Table 5.3. It is clear that this average gain ratio compensation improves the performance for most of the test regions. Considering the average gain compensation value is calculated based on the global image sensor, it is reasonable that the performance is reduced in some local test regions.

To verify its effectiveness, the FPN correction method is applied in the processing of a real infrared LED image. Figure 5.4 compares the raw LED image and its image after correction. It is clear that the correction method suppresses the spatial non-uniformity of the image sensor.

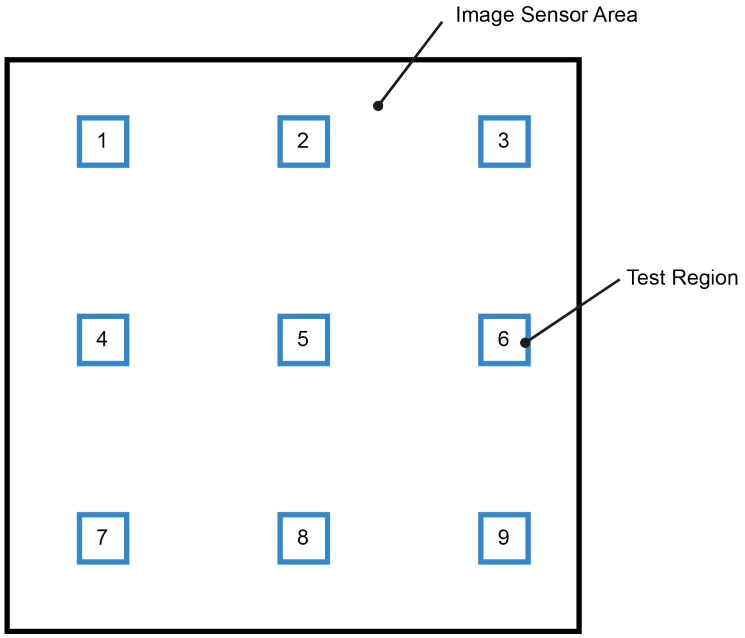


Figure 5.3: Distribution of test regions for average gain difference compensation

Table 5.3: Average gain ratio between even and odd pixels of small regions

Test Region Index	1	2	3	4	5
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$ before compensation	1.0134	1.0208	1.01	1.0151	1.022
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$ after compensation	0.9903	0.9953	0.986	1.0028	1.0059
Test Region Index	6	7	8	9	
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$ before compensation	1.0262	1.006	1.0146	0.9989	
$\frac{\overline{Gain_{Even}}}{\overline{Gain_{Odd}}}$ after compensation	0.9989	0.9876	1.0015	0.9849	

5.2 Position Sampling Frequency Characterization

The position sampling frequency is one of the most important performance characteristics of a position sensing system. As stated in Section 2.2, high position sampling frequency is achieved by ROI-based image readout and processing. Based on the datasheet of the selected image sensor, the theoretical frame period is calculated as

$$T_{\text{frame}} = T_{\text{exposure}} + \text{FOT} + N_{\text{Target}} \times N_{\text{col}} \times \left(\text{ROT} + N_{\text{row}} \times T_{\text{pixel}}\right) \tag{5.3}$$

where T_{exposure} is the exposure time, FOT is the frame overhead time, ROT is the row overhead time, T_{pixel} is the pixel period, N_{Target} is the number of ROI, N_{col} is the column number of ROI, and N_{row} is the row number of ROI. Figure 5.5 shows the ideal relationship of frame rate and ROI size and number of ROI, by ignoring exposure time, frame overhead time and row overhead time.

However, the ideal frame period cannot be achieved due to hardware limitations. First, larger frame

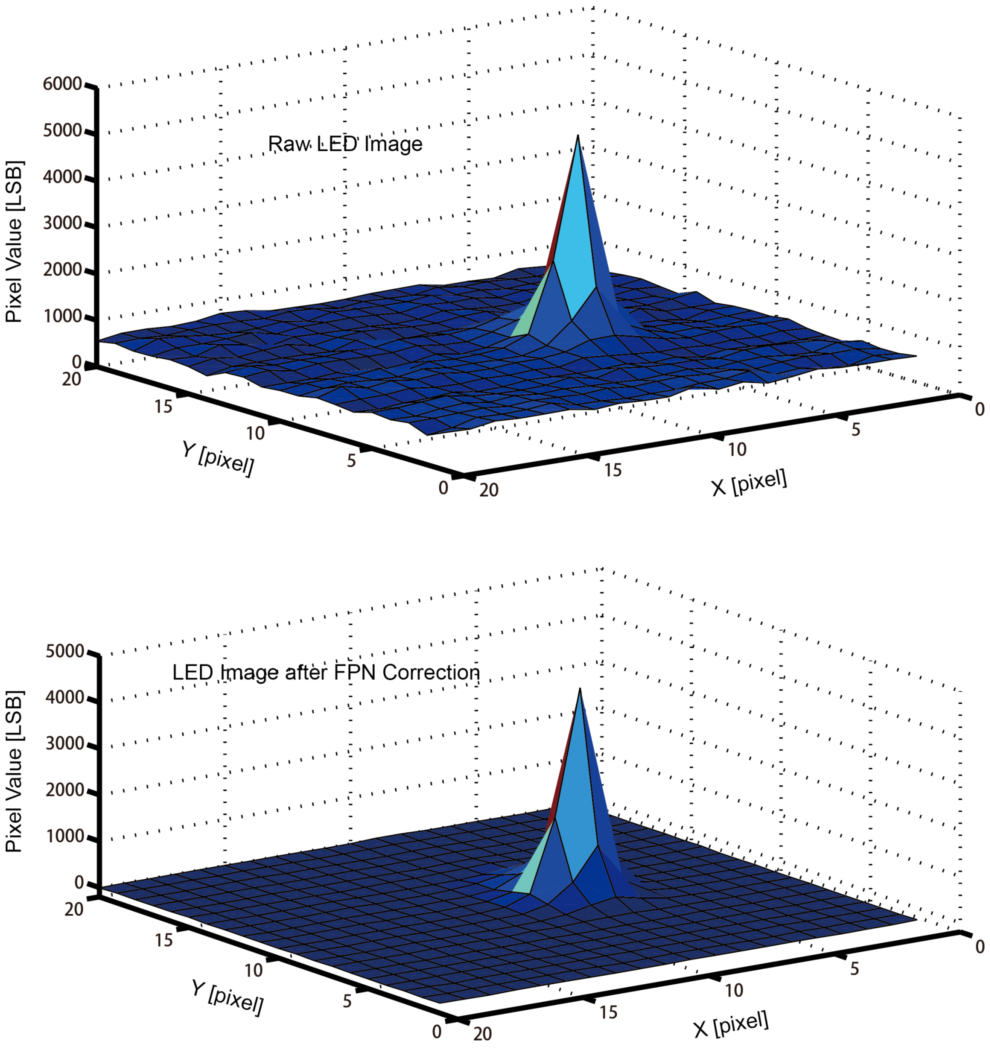


Figure 5.4: Comparison of a raw LED image and the image after FPN correction

overhead time (FOT) and row overhead time (ROT) are required to guarantee that the pixel analog signal becomes stable before readout. Second, the ROI position must be given and loaded to the sensor before the ROI is read, causing an additional ROI update overhead time between the readout of each ROI. Considering all these factors, the actual frame rate is measured and shown in Figure 5.6.

The frame rate is the main bottle neck which limits the achievable position sampling frequency. Based on the ROI readout mechanism of the CMOS image sensor, the frame rate is vastly increased. However, the frame rate is not the only limitation of the position sampling frequency. The sub-pixel target location inside the camera module and the 3D reconstruction computation in the Tsunami real-time computer add additional computation delay. Based on the experimental characterization of the system's computation capability, this system is able to achieve a position sampling frequency of 8kHz for measuring 8 targets in 3-DOF, where the ROI size is set to 14 by 14 pixels.

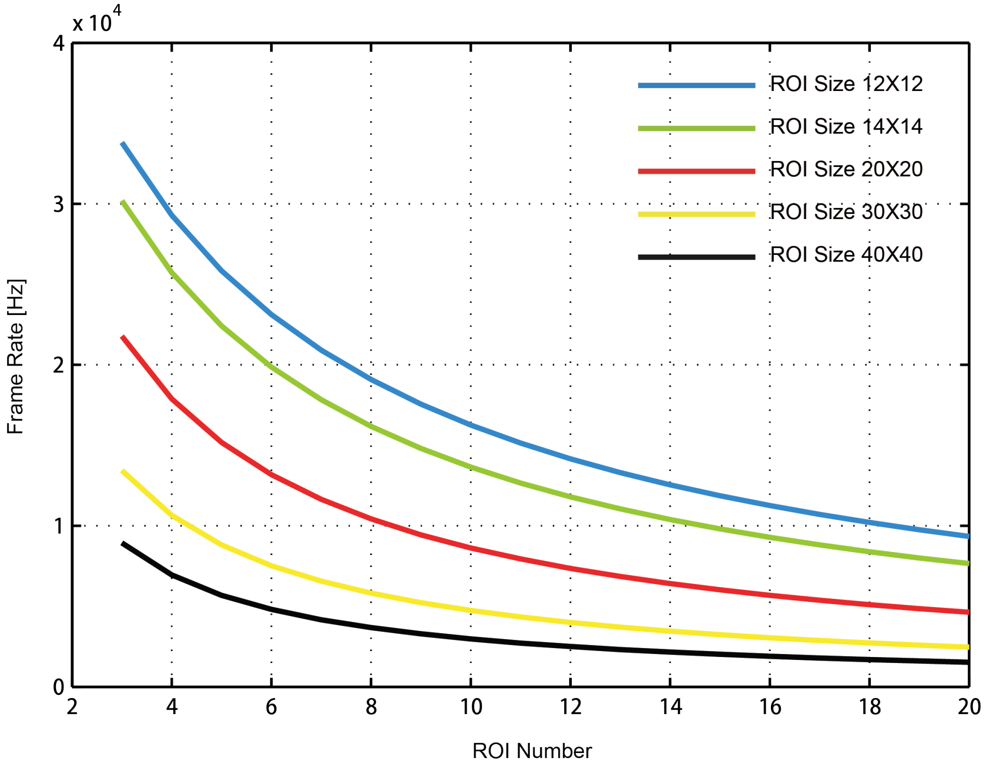


Figure 5.5: Ideal frame rate vs. ROI size and number of target

5.3 Characterization of Sub-pixel Target Position Interpolation

In Section 2.2.2, the sub-pixel target position interpolation algorithms are discussed. Based on the computation cost and achievable sub-pixel interpolation precision, the centroid and squared-centroid method are implemented in each camera module. The sub-pixel target position interpolation is realized. To characterize the performance of the real-time sub-pixel target position interpolation methods, some experiments are conducted and the experimental results are discussed. Limited by the test conditions, the true target position in the image coordinate frame is not accessible, and therefore it is impossible to evaluate the accuracy of sub-pixel target position interpolation methods. Instead, resolution, rather than accuracy, is characterized in this section.

In this experiment, the 3D position of a target is given by the CMM and fixed in the measurement volume. The target position in the pixel coordinate frame is continuously output from the camera module at the position sampling frequency of 8kHz. For each sub-pixel interpolation method, 1000 target positions in image sensor coordinate are collected. The resolution of sub-pixel interpolation is calculated as the standard deviation of these target positions in image sensor coordinate.

The effect of threshold processing is investigated. As presented in Section 5.1, the FPN of the image sensor is relatively constant in the temporal scale, but the experimental result still shows a 6LSB temporal deviation of the sensor FPN with 13-bit ADC resolution. As a result, the online FPN correction is not able to perfectly suppress the ROI background to zero. In order to fully remove the ROI background, threshold processing is considered. The threshold processing is defined as

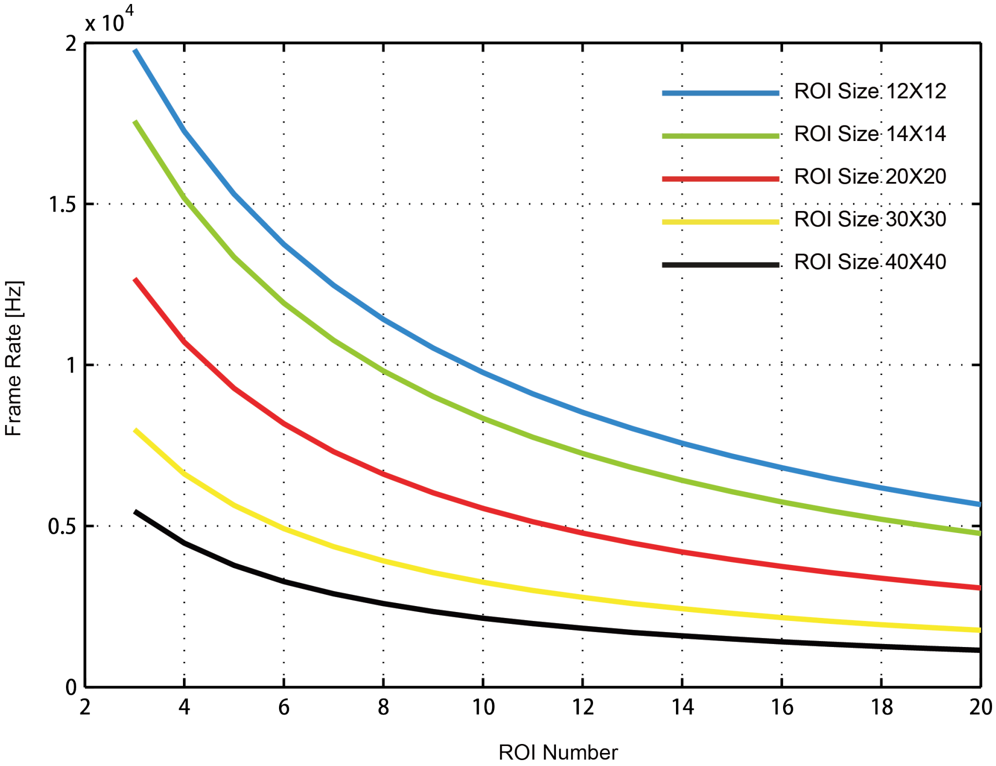


Figure 5.6: Real frame rate vs. ROI size and number of target

$$I_{OUT} = \begin{cases} I_{Raw} - \text{Threshold}, I_{Raw} \geq \text{Threshold} \\ 0, I_{Raw} < \text{Threshold} \end{cases} \tag{5.4}$$

where I_{Raw} is the raw pixel value; I_{OUT} is the pixel value after threshold processing.

The effect of threshold level on the sub-pixel interpolation resolution is shown in Figure 5.7. The experimental result clearly shows that the threshold level has a larger effect on the resolution of the centroid method than the resolution of squared-centroid method. Considering that the squared-centroid method emphasizes the main body of the target image where pixel values are much higher than the background, the background pixels have very small influence on the target location result. As a consequence, further removing the background by threshold processing does not significantly improve its resolution. On the other hand, with a cleaner background, the centroid method achieves better resolution. From Figure 5.7, it is found that the centroid method is able to achieve the same resolution as the squared-centroid when using a threshold to remove the ROI background.

In conclusion, the DSNU correction significantly increases the resolution of both centroid and squared-centroid methods. The threshold processing has a greater influence on the resolution of the centroid method than the squared-centroid method. By using thresholding to remove the background noise, the centroid method is able to achieve the same resolution as the squared-centroid method. Limited by test condition, the accuracy of the sub-pixel interpolation method cannot be experimentally characterized at this point. The effect of sub-pixel interpolation method on 3D reconstruction accuracy will be studied in later sections.

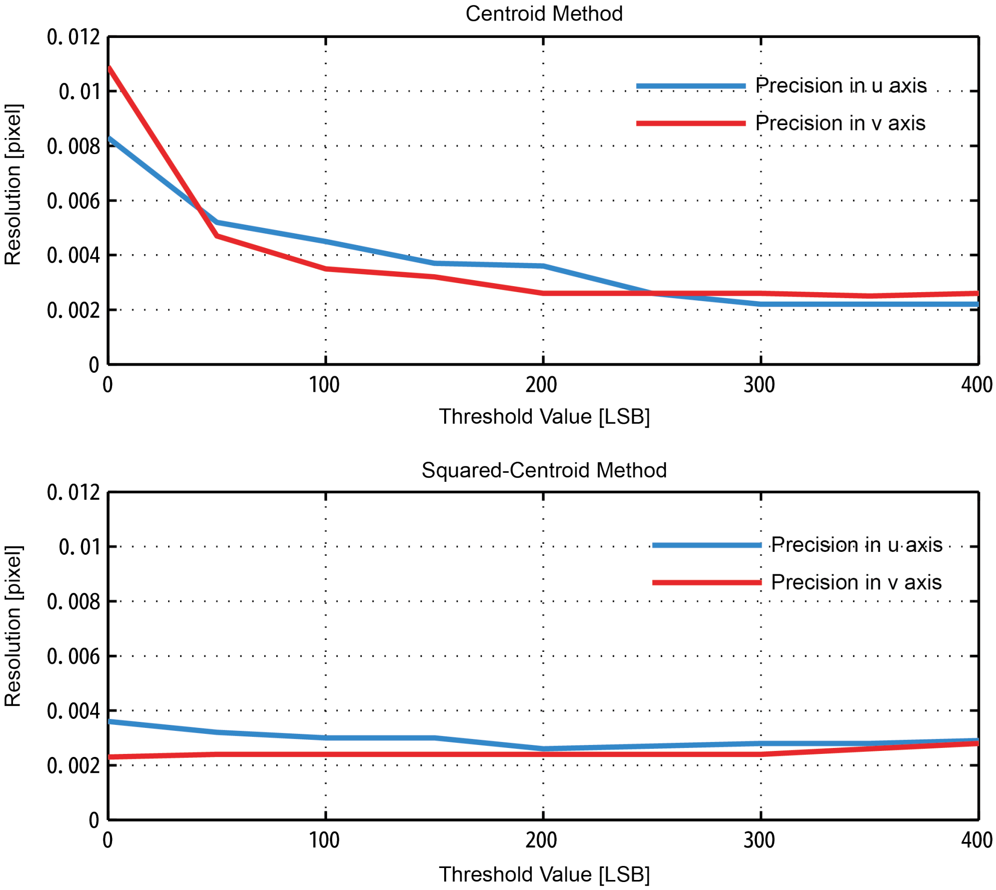


Figure 5.7: Threshold effect on sub-pixel interpolation resolution

5.4 Camera Calibration and Off-line 3D Reconstruction

Camera calibration is one of the key procedures in order to achieve high 3D position accuracy. In this section, the system is calibrated and the calibration performance of different methods is evaluated. The effects of sub-pixel interpolation, DSNU correction, average gain compensation, threshold level, distribution and quantity of test points, and lens distortion model are investigated.

In this section, the performance of the Heikkila camera and NK camera calibration method are compared. The camera model used in both camera calibration methods is described in Section 2.1. In order to make a fair comparison between these two calibration methods, the NK method uses the same lens distortion correction algorithm of the Heikkila method. This lens distortion correction considers the curvature and decentering distortion, and it is describe in Equation 5.5.

$$\begin{aligned} X_I &= X_D - \left[K_1 X_D \left(X_D^2 + Y_D^2\right) + K_2 X_D \left(X_D^2 + Y_D^2\right)^2 + P_1 \left(3X_D^2 + Y_D^2\right) + 2P_2 X_D Y_D\right] \\ Y_I &= Y_D - \left[K_1 Y_D \left(X_D^2 + Y_D^2\right) + K_2 Y_D \left(X_D^2 + Y_D^2\right)^2 + P_2 \left(X_D^2 + 3Y_D^2\right) + 2P_1 X_D Y_D\right] \end{aligned} \tag{5.5}$$

Unlike the simulation comparisons in Section 4.3.3, the true values of model parameters in the real system are unknown. Therefore, it is impossible to evaluate the camera calibration performance based on the accuracy of the estimated model parameters. Considering that better estimation of model parameters gives higher 3D position reconstruction accuracy, we used the off-line 3D reconstruction error to evaluate the calibration performance. The 3D reconstruction RMS error is calculated as

$$\varepsilon_{X_W} = \sqrt{\frac{1}{N}\sum\left(X_{W,est}(\hat{P}) - X_{W,CMM}\right)^2} \tag{5.6}$$

$$\varepsilon_{Y_W} = \sqrt{\frac{1}{N}\sum\left(Y_{W,est}(\hat{P}) - Y_{W,CMM}\right)^2} \tag{5.7}$$

$$\varepsilon_{Z_W} = \sqrt{\frac{1}{N}\sum\left(Z_{W,est}(\hat{P}) - Z_{W,CMM}\right)^2} \tag{5.8}$$

$$\varepsilon_{3D} = \sqrt{\frac{1}{N}\sum\left[\left(X_{W,est}(\hat{P}) - X_{W,CMM}\right)^2 + \left(Y_{W,est}(\hat{P}) - Y_{W,CMM}\right)^2 + \left(Z_{W,est}(\hat{P}) - Z_{W,CMM}\right)^2\right]} \tag{5.9}$$

where $\left(X_{W,est}(\hat{P}), Y_{W,est}(\hat{P}), Z_{W,est}(\hat{P})\right)$ is the reconstruction 3D position based on the estimated camera parameters $\hat{P}$, $(X_{W,CMM}, Y_{W,CMM}, Z_{W,CMM})$ is the 3D position given by CMM, ε_{X_W}, ε_{Y_W}, ε_{Z_W}, ε_{3D} are the reconstruction RMS errors in X_W, Y_W, Z_W and 3D, respectively, and N is the number of points.

In the experiments, the 3D position of the test points are given by a CMM. The CMM used in the experiments is Crysta-Apex C7106 manufactured by Mitutoyo. The specifications of this CMM are shown in Table 5.4. In Table 5.4, MPE_E is the maximum permissible error for length measurement, MPE_P is the maximum permissible error for probing and MPE_THP is the maximum permissible error for scanning probing. The detailed definition of these parameters can be found in the EN ISO 10360 [47].

Table 5.4: Specifications of Crysta-Apex C7106, data from [47]

Parameter		Specification
Range	X-axis	705mm
	Y-axis	1005mm
	Z-axis	605mm
Resolution		0.1μm
Accuracy	MPE_E	(1.7+3L/1000)μm
	MPE_P	1.7μm
	MPE_THP	2.3μm

The camera system is located approximately 960mm from the back of measurement volume to the front surface of camera body. The test points are distributed in a volume of 400mm×400mm×15mm. Test points are distributed in four different $X_W - Y_W$ planes. The illustration of the calibration setup is shown in Figure 5.8. Figure 5.9 shows the block diagram of the camera calibration and its performance evaluation. In order to fairly evaluate the performance of camera calibration method, the 3D test points used for camera calibration and 3D reconstruction accuracy evaluation are not the same.

The test points used for calibration and evaluation are interlaced (shown in Figure 5.10) and cover the measurement volume. Two hundred test points are used for camera calibration and a separate 200

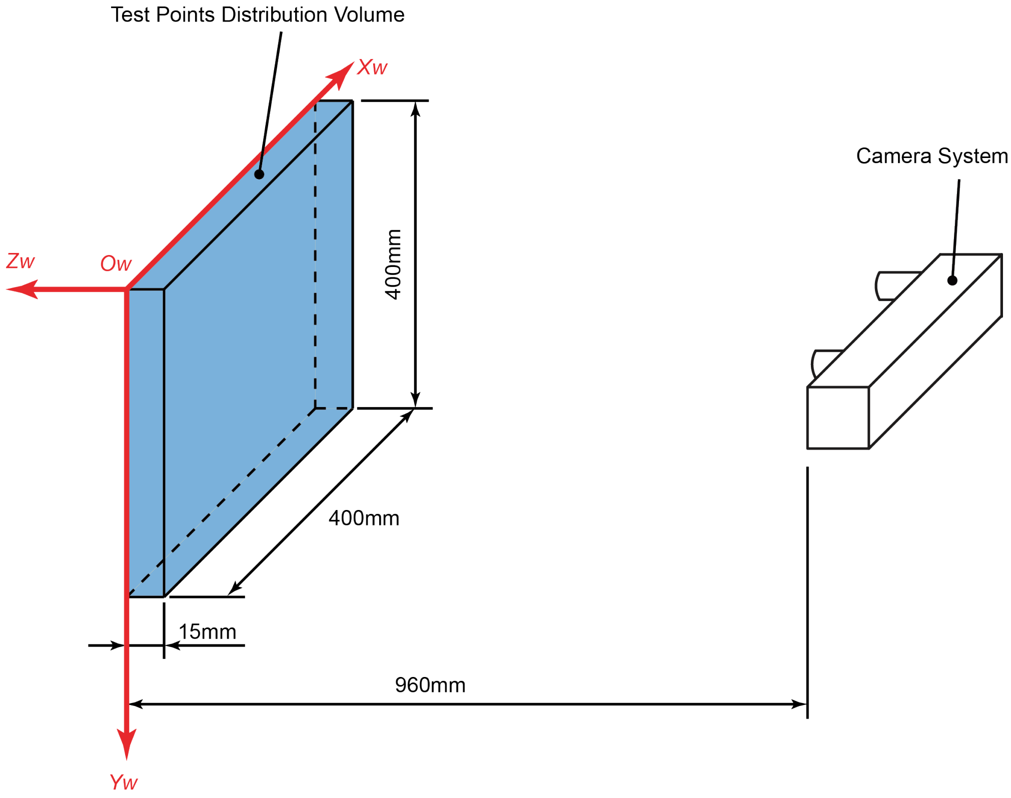


Figure 5.8: Illustration of camera calibration setup

test points are used for evaluation.

Effect of DSNU Correction

The effect of DSNU correction is investigated in this section. The 3D reconstruction results are shown in Figure 5.11.

From Figure 5.11, it is clear that DSNU correction significantly improves the off-line 3D reconstruction accuracy when using either centroid or squared-centroid methods to do the sub-pixel interpolation. Further, Figure 5.11 shows that higher 3D accuracy is achieved by using the squared-centroid method. At the same time, the proposed NK calibration method achieves better estimation of the camera parameters than the Heikkila method, which is reflected by higher 3D reconstruction accuracy when using NK method to calibrate the camera system. A significant improvement is found when using NK method, especially in the accuracy in depth direction (Z_W direction). Based on the theoretical analysis in Section 4.3.2, the NK method utilizes the 3D reconstruction error as the cost function in the optimization process, compared to a 2D image reprojection error used as the cost function in the Heikkila method. As a result, the Heikkila method is more sensitive to sub-pixel interpolation error, especially the larger error in the presence of no DSNU correction. If these estimated camera parameters are used to reconstruct the 3D position, large 3D position reconstruction error in the depth direction is introduced.

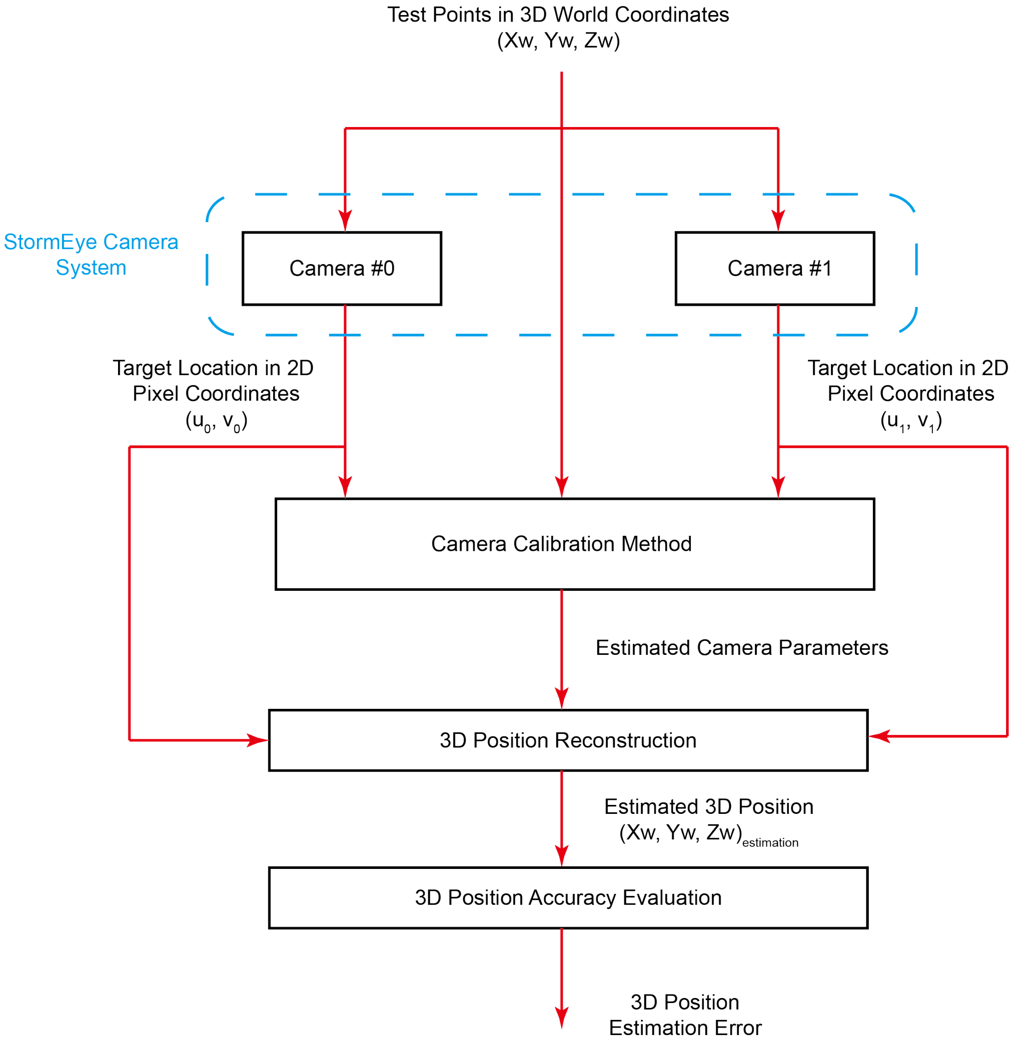

Test Points in 3D World Coordinates
(Xw, Yw, Zw)
StormEye Camera System
Camera #0
Camera #1
Target Location in 2D Pixel Coordinates (u_0, v_0)
Target Location in 2D Pixel Coordinates (u_1, v_1)
Camera Calibration Method
Estimated Camera Parameters
3D Position Reconstruction
Estimated 3D Position
$(Xw, Yw, Zw)_{estimation}$
3D Position Accuracy Evaluation
3D Position Estimation Error

Figure 5.9: Block diagram of camera calibration and its performance evaluation

Effect of Average Gain Compensation

Limited by test conditions, it is impossible to have a pixel-by-pixel PRNU correction for each individual pixel value. Considering that the even and odd pixel signals are amplified, conditioned and sampled through two independent analog channels, the average gain difference between even and odd pixels is measured. From the results in Section 5.1, the average gain compensation is able to attenuate the average gain difference between even and odd pixels to less than 0.1%. In this section, the effect of average gain compensation on the calibration and 3D reconstruction accuracy is investigated. At this point, all results shown in Figure 5.12 are DSNU corrected.

Figure 5.12 shows that the the gain compensation does not remarkably improve the 3D reconstruction accuracy. The experimental results in Table 5.1 shows that the average gain difference between even and odd pixels is about 2%. The gain compensation is based on the global measurement of full image sensor, and therefore gain compensation is unable to perfectly compensate the gain difference in every local region. As a result, the effect of gain compensation on 3D reconstruction accuracy is very small.

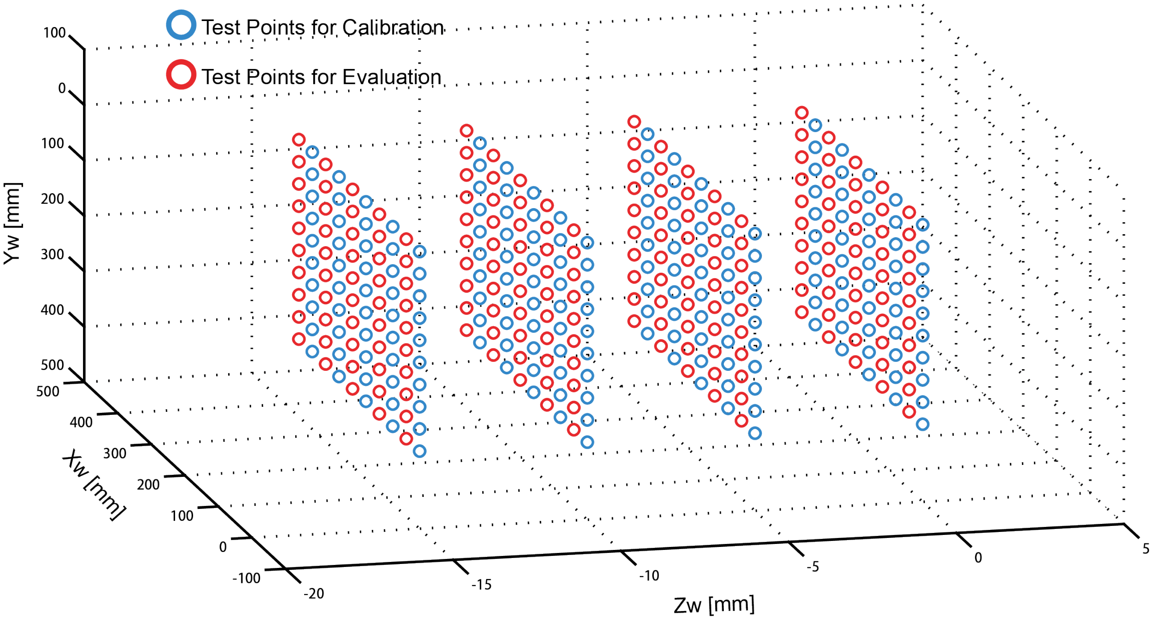


Figure 5.10: Test points distribution 1

Effect of Threshold Level

The DSNU correction is unable to fully remove the background noise in a ROI as a result of the temporal noise like ADC readout noise and sensor dark current shot noise. In order to remove the background noise, threshold processing is used (shown in Equation 5.4). The effect of threshold level on the resolution of sub-pixel target position interpolation method is investigated in Section 5.3. The experimental results show that threshold processing improves the resolution of the centroid method, while the resolution of the squared-centroid method benefits less from threshold processing. In this section, the effect of threshold processing on the calibration and 3D reconstruction accuracy is studied. The experimental results are shown in Figure 5.13.

Similar to the effect of average gain compensation, the threshold does not vastly improve the 3D reconstruction accuracy. Especially, the squared-centroid method emphasizes the main body of target image in the ROI. As a result, small background noise has very little influence on the target location result using the squared-centroid method.

Effect of Test Point Distribution

In the previous investigations, the test points for calibration and evaluation are evenly distributed in the measurement volume. However, it is desirable that the calibration results obtained from part of the volume can be effectively extended to the entire volume. This feature greatly improves the calibration efficiency, which is generally better when the measurement volume is large.

Based on the investigation in the previous section, we know that DSNU correction significantly improves the accuracy of 3D position reconstruction. In addition, average gain compensation and threshold processing are able to further improve the accuracy slightly. In order to make a fair comparison, the DSNU correction, average gain compensation and threshold processing are all enabled in this

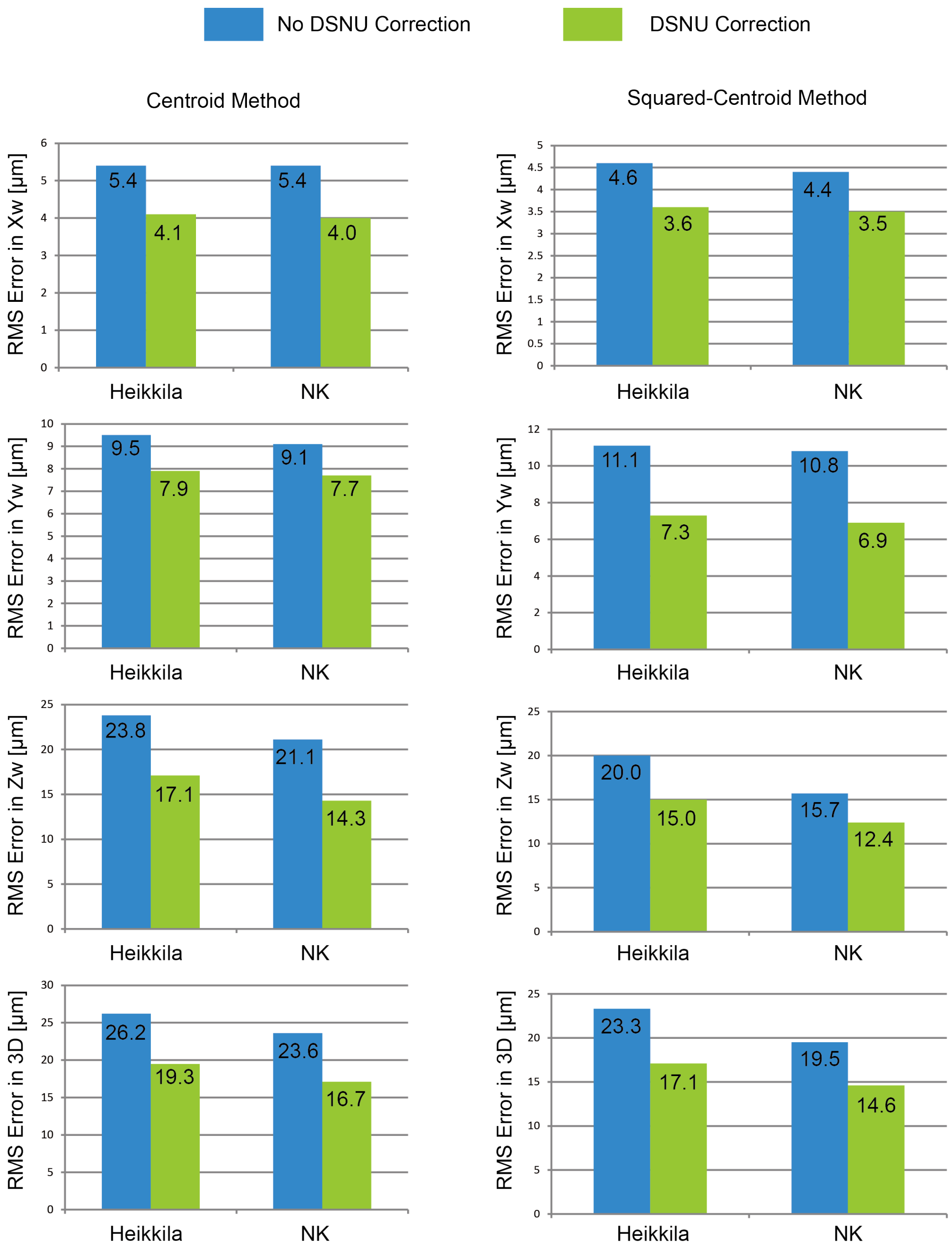


Figure 5.11: Effect of DSNU correction on calibration performance

section. Three types of test point distribution are studied.

The test point distribution shown in Figure 5.10 is denoted as test point distribution 1. For test point distribution 2, the test points located in $Z_W = 0$mm and $Z_W = 5$mm planes are used for calibration and test points in $Z_W = 10$mm and $Z_W = 15$mm planes are used for accuracy evaluation (shown in Figure 5.14). The other test point distribution is shown in Figure 5.15, and is denoted as test point distribution 3.

The 3D reconstruction accuracy of each test point distribution is summarized in Figure 5.16. It is clearly found that the 3D reconstruction accuracy slightly degrades when the calibration results are extended for the 3D reconstruction in other regions. However, this accuracy degradation is very small.

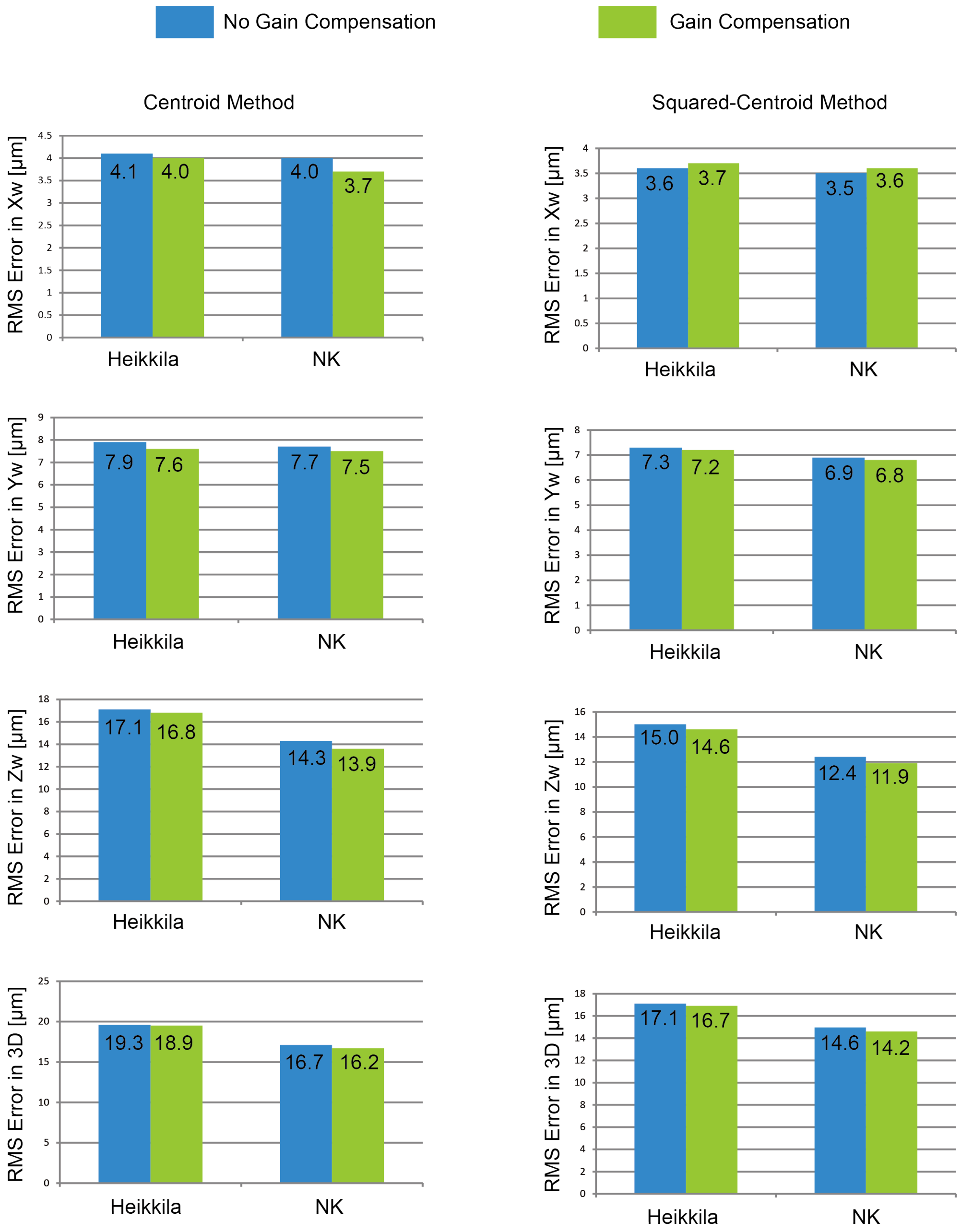


Figure 5.12: Effect of average gain compensation on calibration performance

Effect of Test Point Quantity

Four hundred test points in total are collected in the previous investigations: 200 test points for calibration and the other 200 test points for evaluation. Increasing the number of test points increases the calibration performance in the presence of sub-pixel interpolation errors. On the other hand, a large quantity of test points increases data collection time and computation time in off-line calibration, causing low efficiency. In this section, the effect of test point quantity is studied. An extensive test point data collection is conducted where 800 test points in the measurement volume are collected. Three different cases are compared: 1) 200 test points for calibration and 200 test points for evaluation; 2)

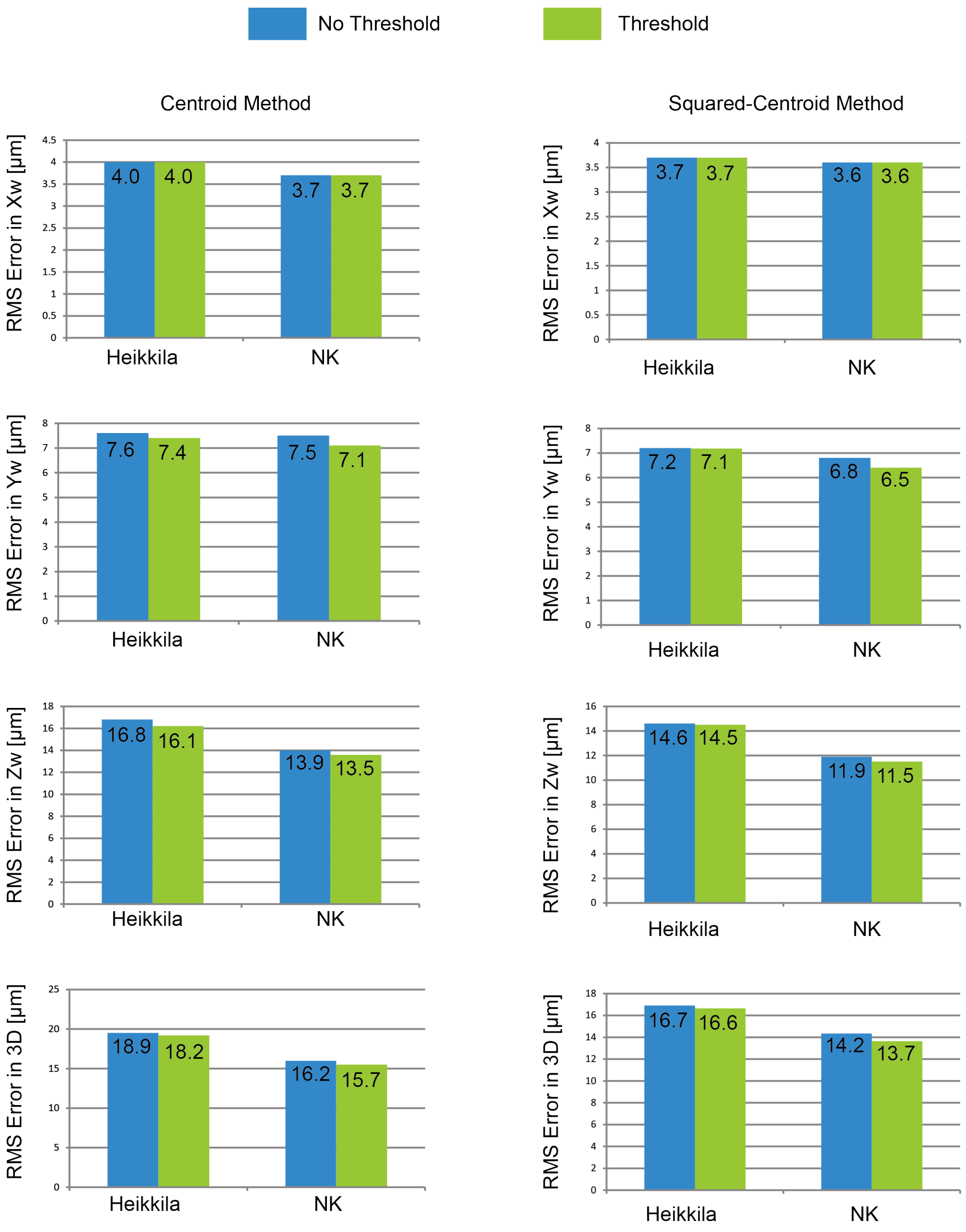


Figure 5.13: Effect of threshold on calibration performance

400 test points for calibration and 200 test points for evaluation; 3) 600 test points for calibration and 200 test points for evaluation. In these cases, calibration test points and evaluation test points are evenly distributed in the measurement volume similar to Figure 5.10. DSNU correction, average gain compensation and threshold processing are all enabled. The off-line 3D reconstruction results are compared in Figure 5.17. It is found that increasing the quantity of test points is able to improve the calibration performance (higher 3D reconstruction accuracy is achieved). However, it is clear that this improvement is not very significant.

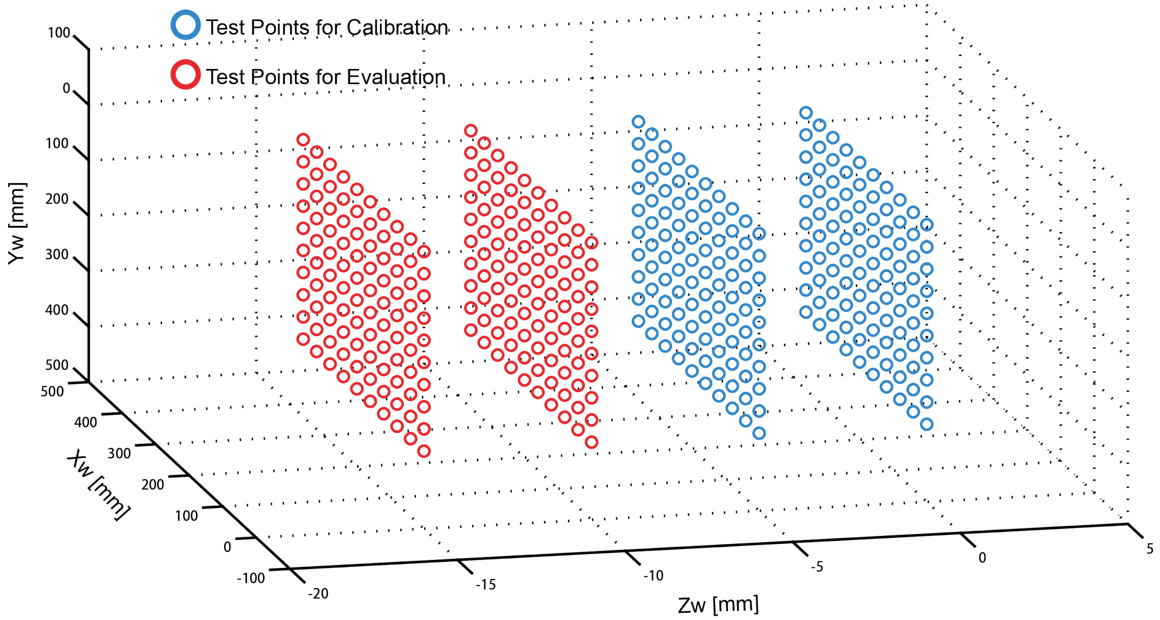


Figure 5.14: Test points distribution 2

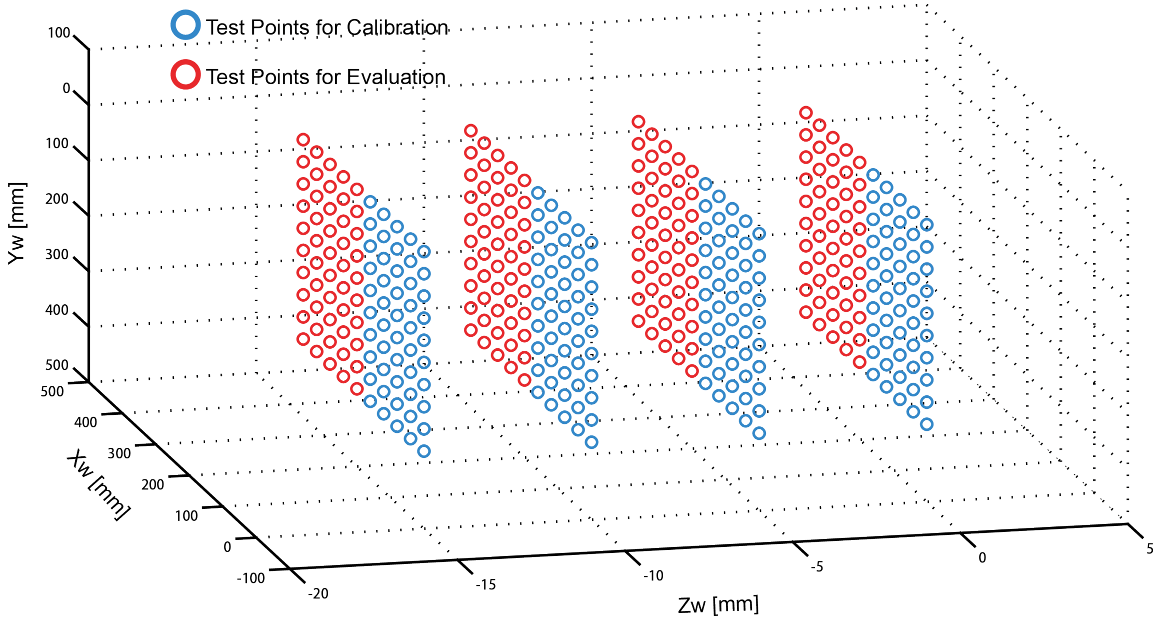


Figure 5.15: Test points distribution 3

Effect of Camera Model

The performance of camera calibration is not only decided by the calibration procedure, but also influenced by the camera model. As shown in Section 4.3.1, different researchers used different camera models in their calibration methods. The major difference in their camera models is the optical distortion of lens. In this part of discussion, the effect of the lens distortion model on the calibration performance of the NK method is investigated.

First, the effect of different types of lens distortion is studied. In the previous discussions, the

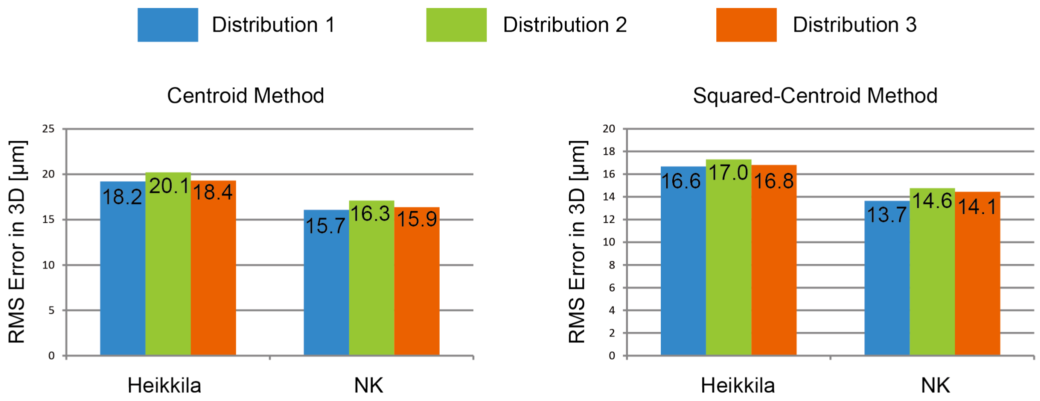


Figure 5.16: Effect of test point distribution on calibration performance

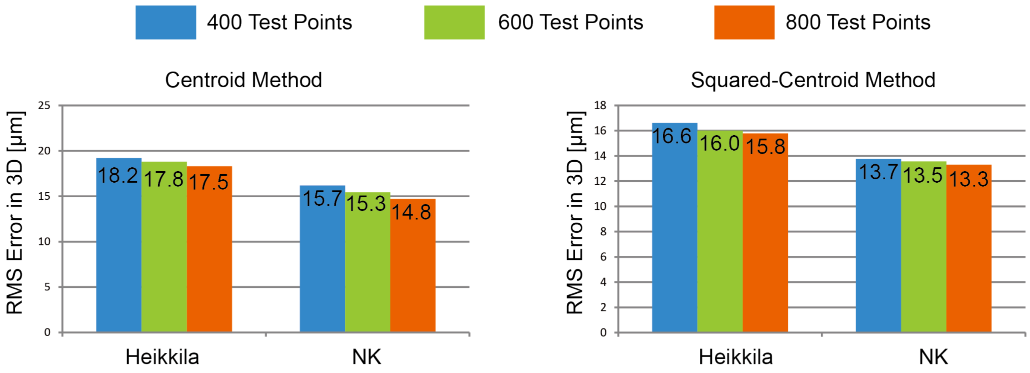


Figure 5.17: Effect of test point quantity on calibration performance

curvature distortion and decentering distortion are considered in the lens distortion correction. Shown in Section 2.1.3, there is a third type of lens distortion, thin prism distortion, that is considered by other researchers. The calibration performance of the NK method in three scenarios are compared. In the first scenario, only the curvature distortion is considered where the third and fifth order components of radial distortion are included (shown in Equation 5.10). The lens distortion model in the second scenario is the same as we used in previous discussions where the decentering distortion (shown in Equation 5.11) is added to the model. In the third scenario, the thin prism distortion is taken into account where its first order component is included (shown in Equation 5.12). DSNU correction, average gain compensation and threshold processing are enabled in all scenarios.

$$\begin{aligned}\delta_{X,r}(X_D,Y_D) &= K_1 X_D\left(X_D^2+Y_D^2\right)+K_2 X_D\left(X_D^2+Y_D^2\right)^2\\ \delta_{Y,r}(X_D,Y_D) &= K_1 Y_D\left(X_D^2+Y_D^2\right)+K_2 Y_D\left(X_D^2+Y_D^2\right)^2\end{aligned}\quad. \tag{5.10}$$

$$\begin{aligned}\delta_{X,d}(X_D,Y_D) &= P_1\left(3X_D^2+Y_D^2\right)+2P_2 X_D Y_D\\ \delta_{Y,d}(X_D,Y_D) &= P_2\left(X_D^2+3Y_D^2\right)+2P_1 X_D Y_D\end{aligned}\quad. \tag{5.11}$$

$$\begin{aligned}\delta_{X,p}(X_D,Y_D) &= S_1\left(X_D^2+Y_D^2\right)\\ \delta_{Y,p}(X_D,Y_D) &= S_2\left(X_D^2+Y_D^2\right)\end{aligned}\quad. \tag{5.12}$$

Shown in Table 5.5, it is clear that the curvature distortion is the dominant part in lens distortion, and incorporating the decentering distortion further improves the calibration performance. On the other hand, the effect of thin prism distortion is negligible in the lens we used. Therefore, the curvature and decentering distortion are the dominant distortion components in the lens we used.

Table 5.5: Effect of types of lens distortion on the calibration performance of the NK method

Distortion Types	Curvature	Curvature and Decentering	Curvature, Decentering and Thin Prism
3D Error RMS [μm]	16.5	13.4	13.4

Further, the effect of order of curvature distortion is investigated. It is known that the curvature distortion is the dominant distortion component of the lens used in this system. In the previous discussions, the third and fifth order components of curvature distortion are included in the model (shown in Equation 5.10). In this part, the seventh order component of curvature distortion is further incorporated (shown in Equation 5.13). At the same time, the decentering distortion is considered here. Shown in Table 5.6, the third and fifth order components are dominant in curvature distortion. However, the calibration performance does not benefit much by incorporating the seventh order component of curvature distortion.

$$\begin{array}{l} \delta_{X,r}\left(X_D,Y_D\right) = K_1X_D\left(X_D^2+Y_D^2\right)+K_2X_D\left(X_D^2+Y_D^2\right)^2+K_3X_D\left(X_D^2+Y_D^2\right)^3 \\ \delta_{Y,r}\left(X_D,Y_D\right) = K_1Y_D\left(X_D^2+Y_D^2\right)+K_2Y_D\left(X_D^2+Y_D^2\right)^2+K_3Y_D\left(X_D^2+Y_D^2\right)^3 \end{array}. \tag{5.13}$$

Table 5.6: Effect of order of curvature distortion on the calibration performance of the NK method

Order of Curvature Distortion	3rd order	3rd and 5th order	3rd, 5th and 7th order
3D Error RMS [μm]	109.6	13.4	13.3

In some camera calibration methods [15][19][18][17], the skew angle of image sensor (shown in Figure 2.4) is considered as 90°. To further verify this assumption, the calibration performance of the NK method is investigated under the situation with and without considering of this skew angle. The results are shown in Table 5.7. It is clear that incorporating the skewness of the image sensor into the system model does not significantly improve the 3D reconstruction accuracy.

Table 5.7: Effect of skew angle on the calibration performance of the NK method

	Model without Skew Angle	Model with Skew Angle
3D Error RMS [μm]	13.4	13.2

Based on investigation above, it is concluded that

- The accuracy of sub-pixel target position interpolation is the key to improve the calibration performance and 3D reconstruction performance. However, limited by the SNR of the target image, the sub-pixel interpolation is unavoidably contaminated with error. The proposed NK calibration method minimizes the 3D reconstruction error and is able to achieve better performance compared to the Heikkila method, especially in the reconstruction of depth information. On the other hand, since parameters of every camera module are simultaneously optimized in the NK method, the convergence speed of the NK method is much slower than the Heikkila method. As shown in the previous results, the Heikkila method is able to achieve close performance as the NK method does when the sub-pixel interpolation accuracy is improved, but the computation time of the Heikkila method is much shorter than the NK method.

- Generally speaking, the squared-centroid method achieves better performance compared to the centroid method. Especially in the presence of background noise, the squared-centroid method emphasizes the main body of the target image, therefore the background noise has insignificant influence on the target location result.
- DSNU correction is able to significantly improve calibration performance and 3D reconstruction accuracy. On the other hand, the effect of average gain compensation and threshold processing is not significant.
- It is desirable that the test points fill the entire test volume as much as possible. However, larger quantity and broader distribution of test points makes the calibration procedure time-consuming. According to the experimental results, when the camera model and calibration method are properly selected, the model parameters estimated from part of the test volume can be extended to the whole test volume for a small performance loss. In addition, when the quantity of test points is large enough, increasing the number of test points does not provide remarkable benefits to the calibration performance.
- The calibration performance is not only determined by the calibration procedure but also influenced by the camera model. For the selected wide-angle lens in this system, radial distortion and decentering distortion are the dominant components in lens distortion, and need to be identified in the calibration and corrected in 3D position reconstruction. On the other hand, the effect of thin prism distortion is negligible in this system. At the same time, experimental results show that the third and fifth order components of curvature distortion is dominant and need to be identified. However, this conclusion is not universally true for all optical lenses. Generally speaking, all types of distortion should be considered if their effect is not characterized.

To finalize the calibration results, an extensive calibration experiment is conducted. In this experiment, DSNU correction, average gain compensation and threshold processing are enabled. A CMM is used to move a single infrared LED accurately in a grid of reference positions throughout the test volume. Several samples are taken at each grid point and averaged in order to reduce the measurement noise. To guarantee that enough data are collected to determine the camera parameters with sufficient accuracy, 800 grid points are used. The camera parameter estimation results of the Heikkila method and the NK method are presented in Table 5.8 and Table 5.9. The camera extrinsic parameters (Euler angles and translations) are relative to the absolute world coordinates given by the CMM. Based on the experimental results above, curvature distortion with third and fifth order components and decentering distortion are incorporated in the lens distortion model. The skew angle of the image sensor is considered as 90°.

Table 5.8: Calibration results of the Heikkila Method

Camera Parameters	Camera 0 Parameters	Camera 1 Parameters
α [deg]	0.85	0.98
β [deg]	0.13	0.29
γ [deg]	-89.07	-90.02
T_x [mm]	381.52	-18.57
T_y [mm]	210.00	207.36
T_z [mm]	-905.37	-904.71
$f_x = dS_x$	2179.80	2178.63
$f_y = dS_y$	2180.05	2178.89
u_0 [pixel]	1021.03	1035.35
v_0 [pixel]	1019.38	1019.01
K_1 [mm^{-2}]	-1.32×10^{-4}	-1.31×10^{-4}
K_2 [mm^{-4}]	1.93×10^{-7}	1.91×10^{-7}
P_1 [mm^{-1}]	-5.34×10^{-6}	8.99×10^{-6}
P_2 [mm^{-1}]	-8.13×10^{-6}	1.17×10^{-6}

For calibration data obtained from grids of test points, the spatial errors at each test point is determined by comparing the reconstructed positions to their corresponding reference positions given by the CMM on a point-by-point basis. The root-mean-square (RMS) errors are obtained and given in Table 5.10.

5.5 Characterization of Real-time 3D Reconstruction

In Section 5.4, the stereo-vision system is calibrated with different calibration methods and the camera parameters are obtained. In this section, the 3D position reconstruction is implemented in the Tsunami real-time computer.

According to the investigation in Section 5.4, DSNU correction, average gain compensation and threshold processing are able to improve the position accuracy. Therefore, these image pre-processing functions are enabled in the investigations of this section. Secondly, different camera parameters are estimated based on the Heikkila and NK methods. The 3D position reconstruction accuracy using these two sets of camera parameters is compared. In addition, the effect of sub-pixel interpolation methods on the real-time 3D position reconstruction is further investigated in this section.

Resolution of Real-time 3D Reconstruction

The position of the infrared LED target is given by CMM and is fixed in the measurement volume. The system records 1000 targets' 3D positions continuously at the sampling frequency of 8kHz. It is found

Table 5.9: Calibration results of the NK Method

Camera Parameters	Camera 0 Parameters	Camera 1 Parameters
α [deg]	0.83	0.97
β [deg]	0.12	0.30
γ [deg]	-89.07	-90.02
T_x [mm]	381.70	-18.63
T_y [mm]	202.46	215.05
T_z [mm]	-905.89	-904.62
$f_x = dS_x$	2185.00	2174.71
$f_y = dS_y$	2181.12	2179.02
u_0 [pixel]	1016.87	1040.71
v_0 [pixel]	1019.22	1018.61
K_1 [mm^{-2}]	-1.32×10^{-4}	-1.32×10^{-4}
K_2 [mm^{-4}]	1.92×10^{-7}	1.94×10^{-7}
P_1 [mm^{-1}]	3.45×10^{-6}	7.95×10^{-6}
P_2 [mm^{-1}]	-8.69×10^{-6}	7.27×10^{-6}

Table 5.10: RMS error of calibration results

RMS Error	Heikkila Method	NK Method
X_W [μm]	4.0	4.0
Y_W [μm]	7.1	6.9
Z_W [μm]	12.3	10.6
3D [μm]	14.8	13.3

that the 3D reconstruction resolution obtained from the Heikkila camera parameters and NK camera parameters is the same: approximately 2μm resolution in $X_W - Y_W$ plane and 5μm resolution in Z_W.

Accuracy of Real-time 3D Reconstruction

To characterize the accuracy of the real-time 3D reconstruction, we use the CMM to move a single infrared LED target in a grid of reference positions filling the measurement volume. At each reference point, the real-time reconstruction result is compared with the result given by CMM, and the reconstruction error is recorded. The 3D reconstruction accuracy is evaluated by the RMS value of the reconstruction errors at every reference point. The reference test points used here are different from those test points used in calibration to ensure a fair characterization. The accuracy results are shown in Figure 5.19.

In the best case, a 3D position accuracy of 18.7μm is achieved by using squared-centroid method and model parameters from the NK method. However, the overall volume 3D RMS position error is

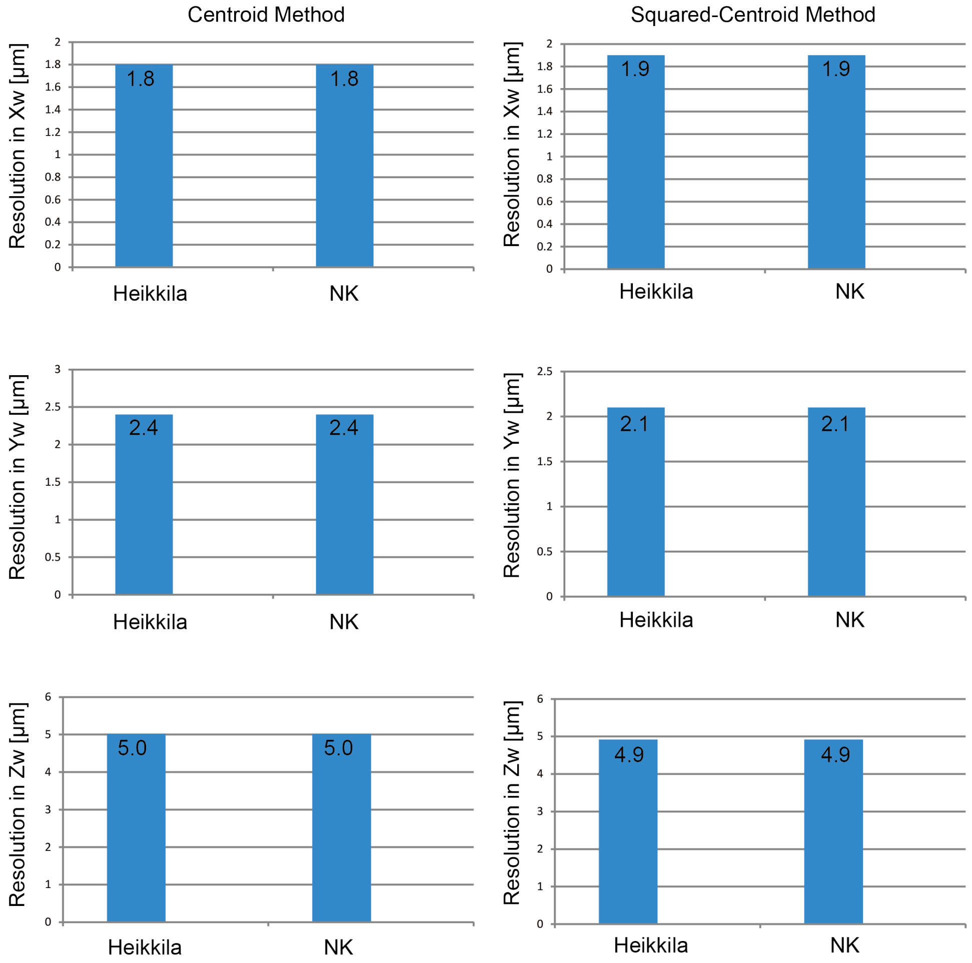


Figure 5.18: Real-time 3D reconstruction resolution at 8kHz sampling frequency

limited in its ability to represent the full accuracy specifications. Much of the underlying information that is necessary to assess the system is hidden. The overall volume RMS position error indicates the typical position error magnitude only in the ideal case where the position error is free of systematic bias, follows a normal distribution and is spatially distributed uniformly across the test volume. Indeed, position sensing systems may not meet these requirements because of the substantial systematic errors that do not satisfy a normal distribution and are not spatially distributed in a uniform way [12]. As a result, the accuracy specification must be carefully examined.

For instance, in the situation where 3D reconstruction is based on the squared-centroid method, the NK camera parameters and 8kHz position sampling frequency, we achieve a position RMS error of 7.4μm in X_W, 10.1μm in Y_W, 13.8μm in Z_W. In this case, the histogram of 3D reconstruction error in each direction is examined (shown in Figure 5.20). The error distributions in X_W, Y_W and Z_W clearly follow a normal distribution with zero mean error. However, it is found that the distribution of 3D position error is not normal and is skewed to higher errors, which is expected because the 3D position errors are defined positive: 3D Error $= \sqrt{\left(X_{W,est}(\hat{P}) - X_{W,CMM}\right)^2 + \left(Y_{W,est}(\hat{P}) - Y_{W,CMM}\right)^2 + \left(Z_{W,est}(\hat{P}) - Z_{W,CMM}\right)^2}$.

However, plotting the position errors histogram does not reflect the spatial distribution of 3D reconstruction errors, because the position error is spatially dependent. The position errors generally increase

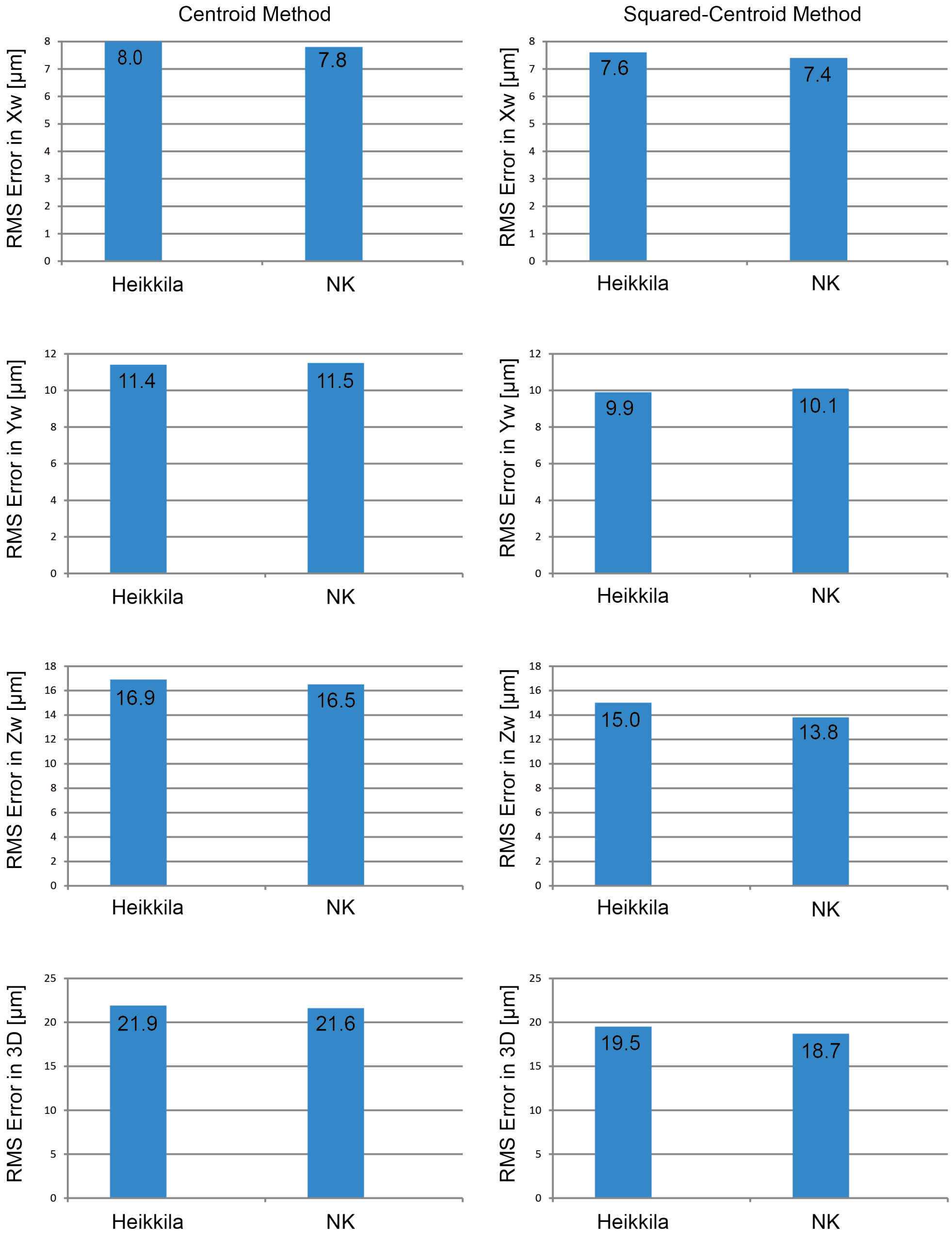


Figure 5.19: Real-time 3D reconstruction accuracy at 8kHz sampling frequency

with the distance from the camera (Z_W direction). Figure 5.21 plots the position errors as a 1D plot as a function of the sequence in which the test points were collected. In Figure 5.21, the vertical axis represents the 3D reconstruction error and the horizontal axis represents the test point index. Shown in Figure 5.8, the camera system is installed at Z_W =-960mm. In the test, 400 test points are collected starting at the back of test volume (Z_W=0mm plane), progressing through the same X_W-Y_W plane, and then moving forward to the next plane. Therefore, the test points with indices from 1 to 100 are located at Z_W=0mm plane; the test points with indices from 101 to 200 are located at Z_W=-5mm plane; the test points with indices from 201 to 300 are located at Z_W=-10mm plane; and the test points with index from 301 to 400 are located at Z_W=-15mm plane. The 3D reconstruction error (in RMS) is calculated

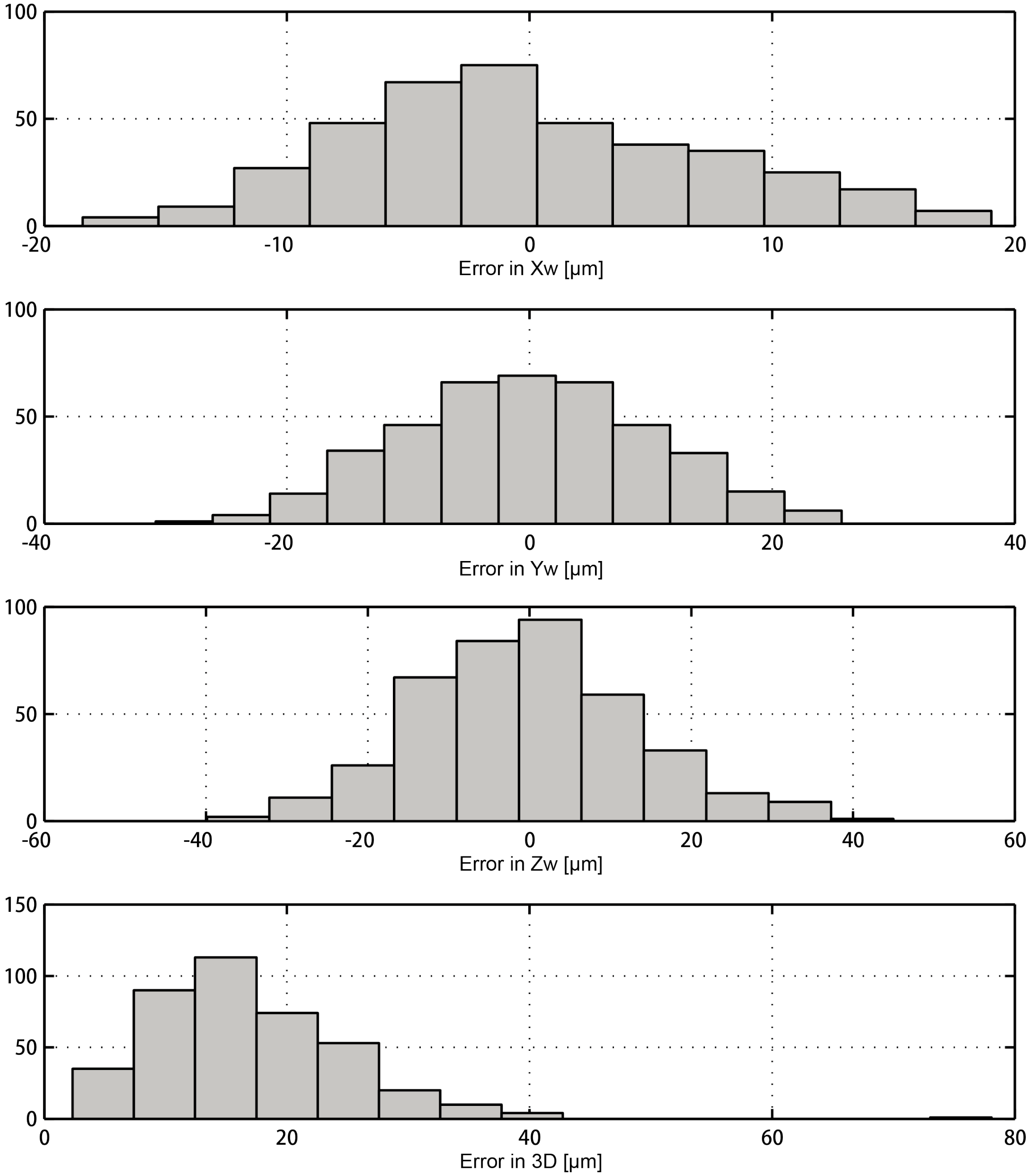


Figure 5.20: Histogram of 3D reconstruction error

for each plane respectively (shown in Figure 5.21). Since the measurement volume is approximately 960mm away from the camera system and the depth of measurement volume (Z_W direction) is very small (15mm), the 3D reconstruction accuracy is not significantly improved when the test point moves from Z_W=0mm plane to Z_W=-15mm plane.

5.6 Case Study: Long-Stroke Planar Motion Stage

Position-sensing systems have a large range of applications in motion control systems. High performance real-time motion control systems require a metrology solution with high position sampling frequency and positioning accuracy. X-Y motion stage systems are a typical motion control system

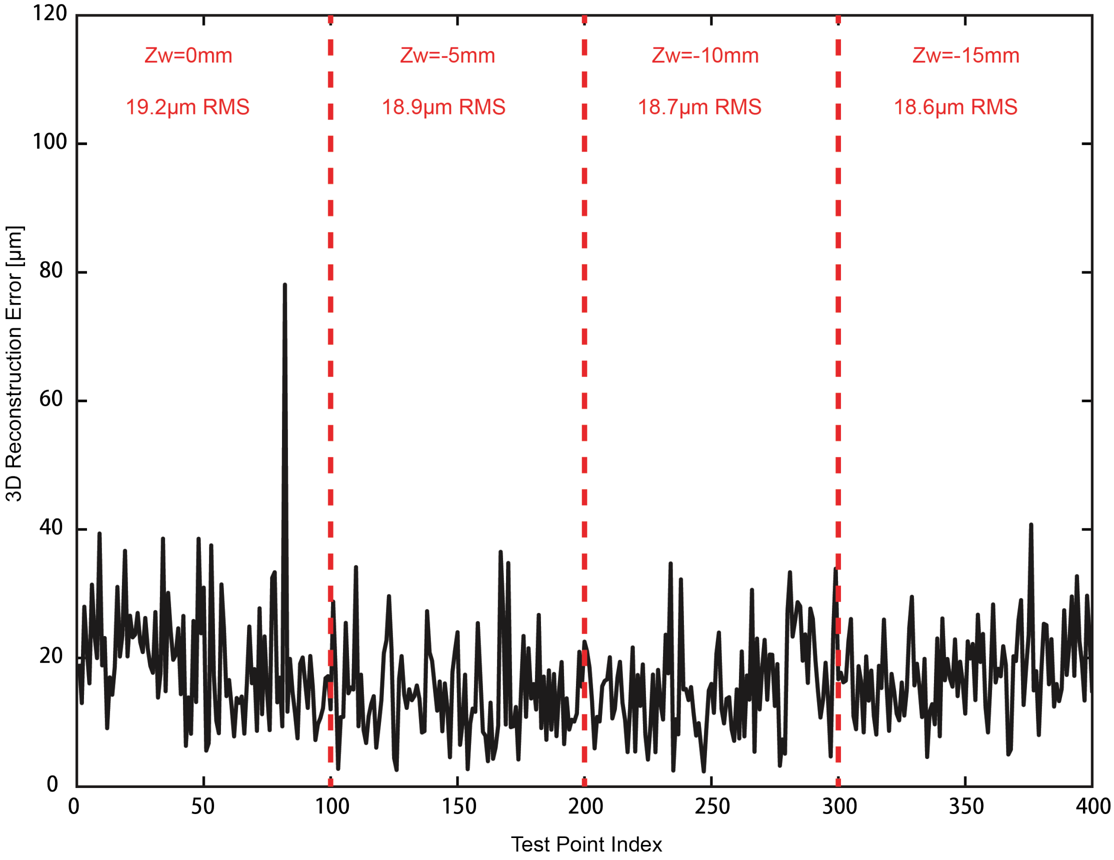


Figure 5.21: 3D reconstruction error plotted in test point sequence

widely used in manufacturing, assembly and inspection. In a planar motion stage system, multi-DOF motion is achieved on a moving body.

The purpose of this case study is to demonstrate the usability of the vision-based position sensing system in an application of real-time motion control. The planar motion stage used for this case study is an novel long-stroke magnetic planar stage developed by Xiaodong Lu and Irfan Usman. This planar motion stage achieves significant advantages over existing solutions, including frictionless 6-DOF actuation over meters-long stroke, and very low system complexity. To capture the 6-DOF motion of the moving stage in a large motion range, 4 infrared LED targets are mounted on the moving stage (shown in Figure 5.22). By sensing the 3D position of each target, the 6-DOF motion of the moving stage is recovered.

The prototype planar motion stage achieves an active planar motion range of 480mm by 270mm with a levitation motion range of 10mm, using a moving stage of 185mm by 185mm by 62mm. The prototype is capable of 5g continuous acceleration with a 2.3kg moving mass. Figure 5.23 shows the installation of the position sensing system for the planar motion stage.

Figure 5.24 shows the measurement resolution when sensing 6-DOF motion. In this experiment, the moving stage is fixed in the measurement volume, and 1000 samples are recorded at 8kHz sampling frequency. It shows that the prototype system achieves high resolution in 6-DOF measurement: 5μm in position and 0.0014 degrees in rotation. Limited by test conditions, the accuracy and repeatability of

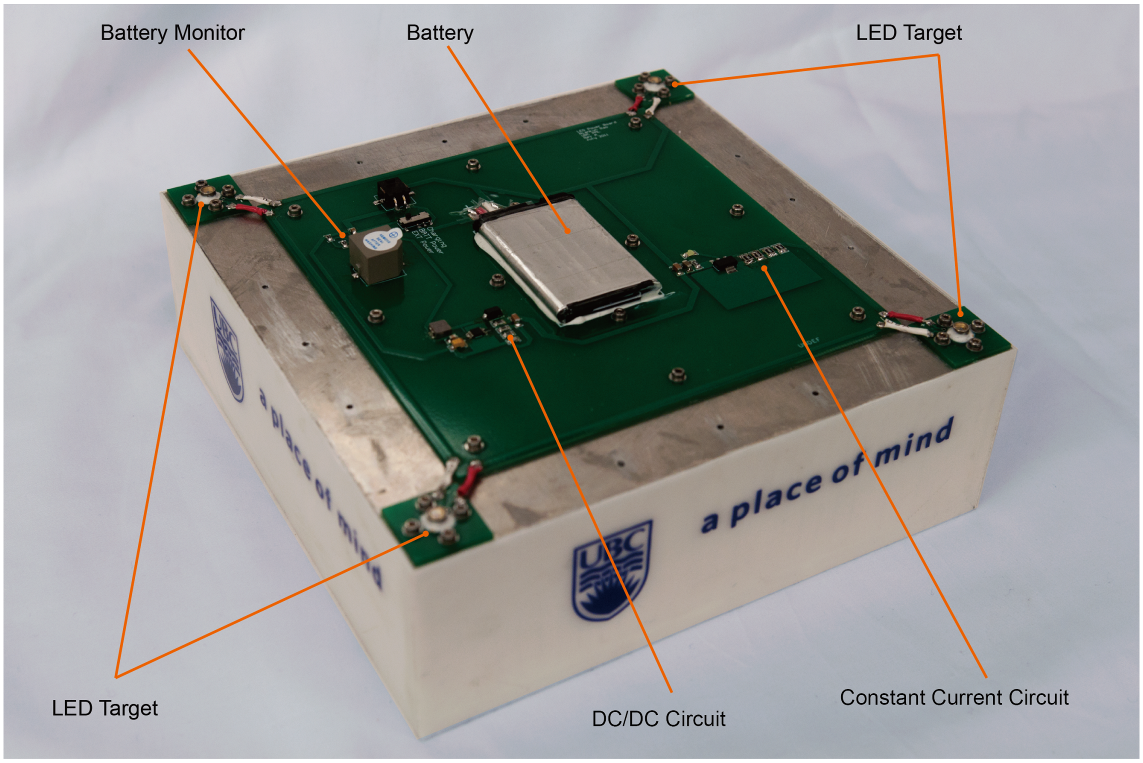


Figure 5.22: Infrared LED targets mounting on the moving stage

6-DOF measurement is not obtained.

A closed-loop motion bandwidth of 50Hz is achieved for the initial demonstration of this long-stroke planar motion stage. However, higher bandwidth is achievable because of high linearity of this actuator. Considering that the motion sensing system is working at 8kHz sampling frequency, the limitations on motion bandwidth do not come from the speed of motion feedback loop.

5.7 Summary

This chapter presents the experimental results of our system, including image sensor calibration and FPN correction, position sampling frequency characterization, investigation of sub-pixel target position interpolation methods, camera calibration experiments, and characterization of real-time 3D reconstruction performance.

The image sensor of each camera module is calibrated in Section 5.1. The DSNU maps are collected and analyzed. The experimental results show that DSNU correction is able to significantly suppress the non-uniformity in the black image. Limited by test conditions, pixel-by-pixel PRNU correction is unable to be implemented. Instead, average gain compensation between even and odd pixels are implemented.

In Section 5.2, the position sampling frequency with different target numbers and ROI sizes is studied. The system is able to track 8 targets at 8kHz position sampling frequency. Higher position sampling frequency is achievable when sensing fewer targets or using smaller ROIs.

Section 5.3 investigates the performance of two sub-pixel target position interpolation methods:

the centroid method and the squared-centroid method. The effect of DSNU correction, average gain compensation and threshold processing on the resolution of these two methods is studied. Experimental results show that a sub-pixel interpolation resolution less than 1% pixel is achieved.

In Section 5.4, the stereo-vision system is calibrated. The Heikkila calibration method and our proposed calibration method (the NK method) are compared. Meanwhile, the effects of DSNU correction, average gain compensation, test point quantity and distribution, and optical distortion model on the performance of both calibration methods are investigated. The experimental results clearly demonstrate that the proposed NK method achieves better performance than the Heikkila method does.

The resolution and accuracy of the real-time 3D reconstruction are characterized in Section 5.5. At the sampling frequency of 8kHz, this system achieves approximately 2μm resolution in X_W-Y_W plane and 5μm resolution in Z_W. Further, the 3D reconstruction accuracy is examined not only using RMS values but also investigated in histogram and spatial distribution. A real-time 3D reconstruction accuracy of 18.7μm RMS is achieved over a range of 400mm by 400mm by 15mm with 8kHz position sampling frequency.

A case study is conducted where the prototype system is integrated as the metrology solution of a novel long-stroke planar motion stage. The 6-DOF motion of the moving stage is obtained from the position sensing system in an active motion range of 480mm by 270mm by 10mm. When sensing the 6-DOF motion, the prototype system achieves high measurement resolution of 4μm in position and 0.0014 degrees in rotation. A closed-loop motion bandwidth of 50Hz is achieved for the initial demonstration of this long-stroke planar motion stage. Considering that the motion sensing system is able to work at 8kHz sampling frequency, the motion feedback loop is not the essential bottle neck to achieve higher motion bandwidth.

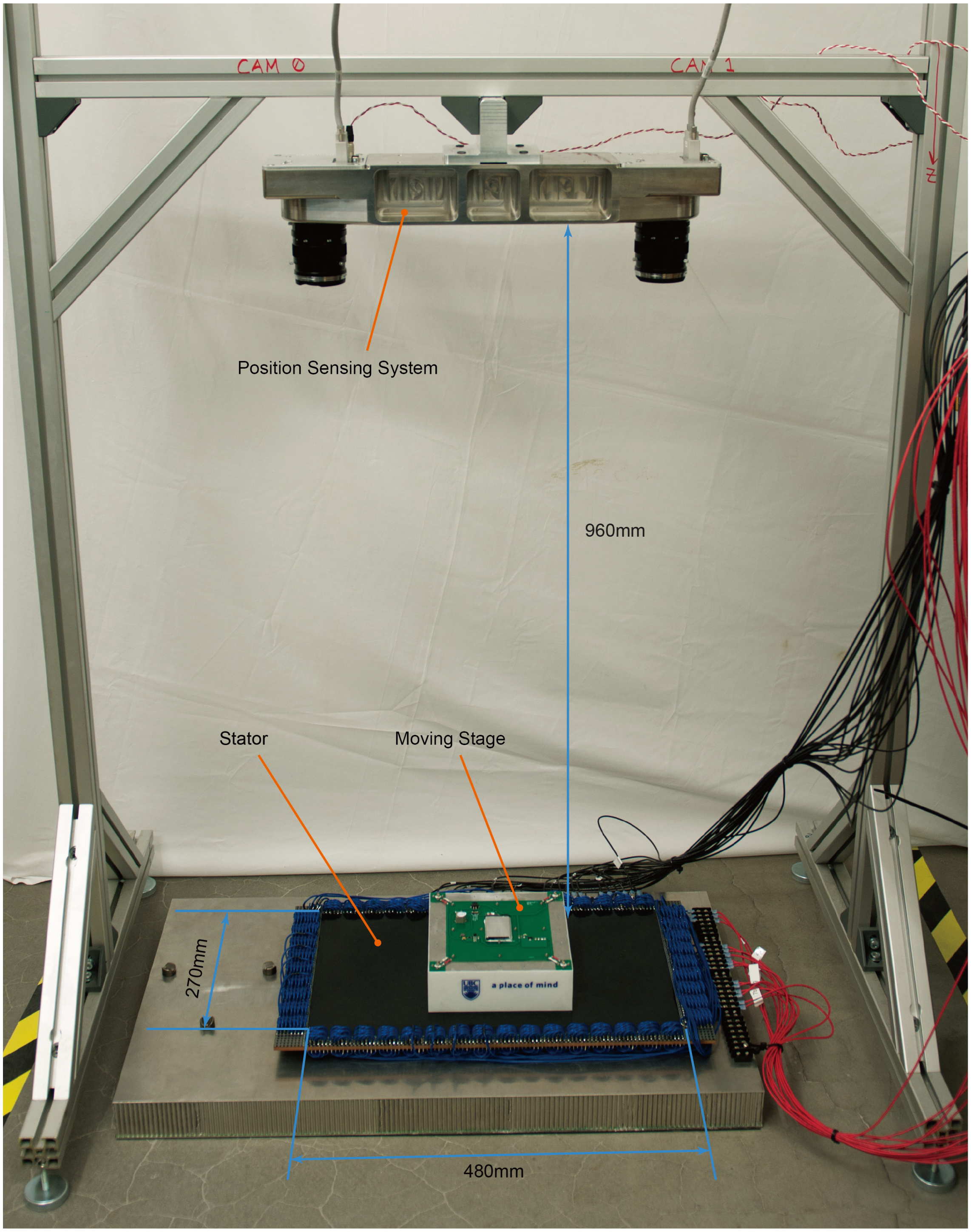


Figure 5.23: Installation of the position sensing system for a long-stroke planar motion stage

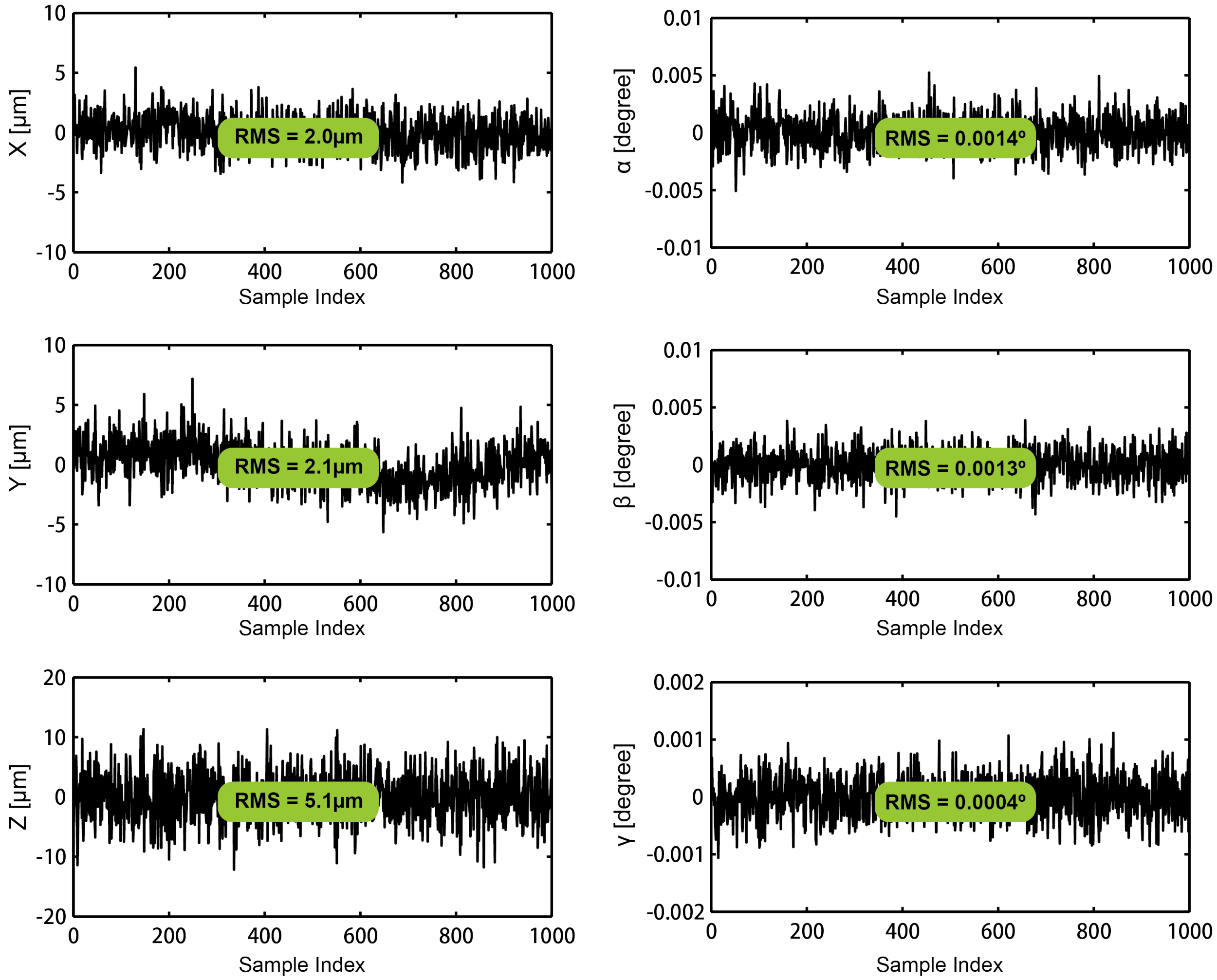


Figure 5.24: Resolution of 6-DOF measurement

Chapter 6

Conclusions and Future Work

6.1 Consclusions

This thesis presents the theory, design, implementation and calibration of a new high-speed stereo-vision system for real-time position sensing.

A novel stereo-vision system hardware prototype is designed and manufactured. The custom electronics is designed based on CMOS image sensor and FPGA architecture. It features high-performance image acquisition, high-speed camera interface and high-speed image processing. Taking advantage of the random pixel accessibility of the CMOS image sensor, image readout and processing based ROI is enabled. By reading small regions around each target image rather than the full frame data of image sensor, the frame rate and image processing speed are vastly increased. In addition, the system also integrates functionalities of FPN correction, threshold processing, and sub-pixel target position interpolation. A high-speed camera interface based on Camera Link technology provides the capability of fast data transmission between camera and image processing unit in real-time with low latency. Good design and implementation of analog and digital circuits provide a high SNR of target image, increasing the achievable resolution and accuracy of sub-pixel interpolation. High quality optical components, including lens and active targets, are carefully selected in order to satisfy the design objectives. To minize the effect of thermal expansion and external vibration, a solid camera body is designed and manufactured with optimized geometry, material and mechanical structure.

The calibration theory of stereo-vision system is investigated. The primary error sources of vision-based position sensing system include noise of target image, lens distortions and mechanical installation errors. Based on the modeling of FPN in the image sensor, the FPN is corrected by using DSNU and PRNU to linearly interpolate the pixel output. Further, the proposed calibration method for optical and mechanical parameters is presented. The proposed calibration method is designed based on multi-step calibration procedure where the optimization of linear and nonlinear optimization parameters is decoupled. In the nonlinear optimization, the 3D position reconstruction error rather than the 2D reprojection error in pixel coordinates is used as the cost function. Simulations are carried out to compare the performance of the proposed calibration method to the Heikkila method. The simulation results show that the proposed calibration method achieves better estimation of camera parameters than the Heikkila method, especially in the presence of large error in sub-pixel target position interpolation.

The performance of this prototype system is characterized by experiment. A multi-target position sensing experiment demonstrates that the system can measure 8 targets in 3-DOF at 8kHz sampling frequency. Higher sampling frequency is able to achieve with fewer targets and smaller ROIs. Further, the system is calibrated using the CMM. The proposed calibration method and the Heikkila method are compared. The experimental results show that the proposed calibration method achieves higher calibration performance compared to the Heikkila method. Third, the resolution and accuracy of real-time 3D position reconstruction is characterized. This system achieves a 3D measurement resolution of 1.9μm in X, 2.1μm in Y and 4.9μm in Z at 8kHz sampling frequency. Comparing the real-time 3D reconstruction results with the CMM, a 3D position accuracy of 18.7μm (RMS) is achieved over a range of 400mm by 400mm by 15mm.

A case study is conducted where the system is used as the metrology solution of a novel long-stroke planar motion stage. The 6-DOF motion of the moving stage is obtained from this position sensing system over an active motion range of 480mm by 270mm by 10mm. When sensing the 6-DOF motion, the prototype system achieves high measurement resolution of 5μm in position and 0.0014 degrees in rotation. A closed-loop motion bandwidth of 50Hz is achieved for the initial demonstration of this long-stroke planar motion stage. Considering that the motion sensing system is able to work at 8kHz sampling frequency, the motion feedback loop is not the essential bottle neck to achieve higher motion bandwidth.

Table 6.1 summarizes the specifications of the system.

Table 6.1: Specifications of the prototype system

Parameter	Specification
Position Sampling Frequency	8kHz for 8 Targets
Measurement Volume	400mm×400mm×15mm
Resolution	1.9μm (X), 2.1μm(Y), 4.9μm(Z)
Accuracy (RMS)	7.4μm(X), 10.1μm(Y), 13.8μm(Z), 18.7μm(3D)
Target Type	Active infrared LED
Dimensions	515×115×140mm
Weight	14kg
Camera Interface	Camera Link

This system also has its limitations. First, the experimental characterization results of the image sensor show that the DSNU, which is supposed to be temporally invariant, has a temporal standard deviation of 6 LSB in 13-bit ADC resolution. This number is higher than the noise floor of ADC circuits. Therefore, the increase of DSNU temporal deviation comes from the image sensor. The SNR of the target image is limited by the image sensor, which further limits the resolution and accuracy of sub-pixel target position interpolation and eventually limits the achievable 3D position reconstruction accuracy. Second, the experimental results show that the FPN is very sensitive to the operation conditions. To guarantee the effectiveness of the FPN correction, it must be ensured that the operation conditions when the FPN map is collected is the same as the raw image is taken. Third, the proposed calibration method for stereo-vision system utilizes the conventional gradient-descent optimization method in the nonlinear iteration. This kind of optimization technique suffers from susceptibility to be trapped in a local minimum. Therefore, camera model parameters obtained from the proposed calibration method do not guarantee a global optimization over the measurement volume. Meanwhile, because model parameters of two camera modules are simultaneously optimized in the nonlinear iteration, the convergence of the nonlinear least-square optimization is slow. Lack of a better synchronization mechanism between the CMM and the stereo-vision system, the test point data collection is time-consuming and labored. Finally, this system lacks a detection and prediction mechanism when the camera module loses the track of target. Therefore, the system requires a full recovery if any target is lost during the position sensing.

6.2 Future Work

Several threads for future work in vision-based position sensing system have emerged from this research.

One of the fundamental limitations in the vision-based position sensing system is the achievable resolution and accuracy of sub-pixel target position interpolation. The performance of sub-pixel position interpolation is significantly influenced by the SNR of target image. Based on the current imaging electronics, we achieve a target image SNR of 56.5dB after DSNU correction. Higher SNR is achievable with better electronics design. Meanwhile, limited by test conditions, pixel-by-pixel PRNU correction is not achievable. Instead, average gain compensation is used to compensate the gain difference between even and odd pixels. However, the effect of this average gain compensation is very small according to experimental results. Further investigations on the pixel-by-pixel PRNU correction are suggested in order to improve the SNR of target image.

The performance of camera calibration is affected by the combination of camera model and parameter estimation methods. In this thesis, we use the camera model from prior art. Higher calibration performance is able to achieve with a more accurate camera model. In the estimation of camera parameters, this thesis utilizes the cost function that minimizes the 3D reconstruction error in the nonlinear optimization iteration. However, the nonlinear iteration used in this thesis is a conventional gradient-descent optimization. This kind of optimization techniques have problems of poor convergence and susceptibility to be trapped in local minimum. Even though a good initial guess of model parameters is given in the first step of calibration, the selected optimization method still has the risk of local rather than global optimization. Research efforts are suggested to devote to the investigation of better optimization algorithms in order to guarantee a global optimization.

The prototype system is calibrated off-line using the CMM and then uses the estimated model parameters to realize the real-time 3D reconstruction. However, the camera parameters might change with operating environment. Especially for the extrinsic parameters of camera geometry, they are more sensitive to the temperature change and external vibrations. Once these camera parameters change as a result of variant operating environment, the system is required to be re-calibrated in order to guarantee the position accuracy. In the current design, the camera body is carefully designed to minimize the effect of thermal deformation and external vibrations. Current design is a passive solution that mitigates the change of model parameters. Meanwhile, materials with very good thermal stability are generally very expensive and hard to machine since the camera body structure is complicated. In order to solve this problem, online camera calibration should be implemented in the future.

Bibliography

[1] Yingling Huang. A High Accuracy CMOS Camera Based 3D Optical Tracking System Design. Master's thesis, University of British Columbia, 2010. → pages 2

[2] OSI Optoelectronics Inc. Lateral-effect Photodiodes. Application note accessed from http://www.osioptoelectronics.no/application-notes/AN-08-Lateral-Effect-Photodiodes.pdf on March 9 2011. → pages 2

[3] Greg Welch, Gary Bishop, Leandra Vicci, Stephen Brumback, Kurtis Keller, and D'nardo Colucci. High-Performance Wide-Area Optical Tracking: The HiBall Tracking System. *Presence: Teleoperators & Virtual Environments*, 10:1–21, 2001. → pages vi, 2, 3

[4] Dave Litwiller. CMOS vs. CCD: Maturing Technologies, Maturing Markets. *Photonics Spectra*, 2005. → pages 3, 4, 5, 25, 26

[5] Advanced Realtime Tracking Inc. ARTtrack optical tracking system datasheet. Datasheet accessed from http://www.ar-tracking.de/Optical-tracking.33.0.html online on Feb 28 2011. → pages 3

[6] Mathieu Herve. The Cyclope: A 6 DOF Optical Tracker Based on a Single Camera. In *2nd INTUITION International Workshop*, 2005. → pages 3

[7] Northern Digital Inc. Polaris Family of Optical Tracking Systems Datasheet. Datasheet accessed from http://www.ndigital.com/medical/polarisfamily.php online on Feb 28 2011. → pages 4

[8] V.Macellari. CoSTEL: a computer peripheral remote sensing device for 3 dimensional monitoring of human motion. *Medical and Biological Engineering and Computing*, 21(3):311–318, 1983. → pages vi, 4

[9] Northern Digital Inc. OPTOTRAK Portable Coordinate Measurement Machine Datasheet. Datasheet accessed from http://www.ndigital.com/industrial/optotrak.php online on Feb 28 2011. → pages 4

[10] Ulrich Muehlmann, Miguel Ribo, Peter Lang, and Axel Pinz. A New High Speed CMOS Camera for Real-Time Tracking Applications. In *Proceedings of the 2004 IEEE International Conference on Robotics & Automation*, volume 5, pages 5195–5200, April 2004. → pages vi, 5

[11] Crispin D. Lovell-Smith. A Prototype Optical Tracking System: Investigation and Development. Master's thesis, University of Canterbury, 2009. → pages vi, 6, 16, 19

[12] Andrew D. Wiles, David G. Thompson, and Donald D. Frantz. Accuracy Assessment and Interpretation for Optical Tracking Systems. In *Proceedings of SPIE*, volume 5367, pages 421–432, Feburary 2004. → pages 10, 91

[13] Max Born and Emil Wolf. *Principles of Optics*, chapter Geometry theory of optical imaging, page 142. Permagon Press, 1965. → pages 10, 13

[14] Chester Slama, Charles Theurer, and Soren Henriksen. *Mannual of Photogrammetry*, chapter Elements of Photogrammetric Optics, page 126. American Society of Photogrammetry, 1980. → pages vi, 10, 12, 13

[15] Y.I. Abdel-Aziz and H.M. Karara. Direct Linear Transformation into Object Space Coordinates in Close-Range Photogrammetry. In *Proceedings of Symposium on Close-Range Photogrammetry*, pages 1–18, Januarary 1971. → pages 10, 19, 21, 58, 59, 87

[16] D.C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing*, pages 444–462, May 1966. → pages 10, 13, 14

[17] Juyang Weng, Paul Cohen, and Marc Herniou. Camera Calibration with Distortion Models and Accuracy Evaluation. *IEEE Transactions on Pttern and Analysis and Machine Intelligence*, 14:965–980, October 1992. → pages 10, 12, 13, 20, 58, 59, 87

[18] Roger Y. Tsai. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, RA-3:323–344, August 1987. → pages vii, 10, 14, 20, 57, 58, 59, 60, 87

[19] J. Heikkila and O. Silven. A Four-step Camera Calibration Procedure with Implicit Image Correction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1997. → pages 10, 12, 14, 20, 58, 59, 61, 87

[20] W. Faig. Calibration of close-range photogrammetric systems: mathematical formualtion. *Photogrammetric Engineering and Remote Sensing*, 41:1479–1486, Decemeber 1975. → pages 14

[21] Zhenyou Zhang. A Flexible New Technique for Camera Calibratoin. *IEEE Transactions on Pttern and Analysis and Machine Intelligence*, 22:1330–1334, 2000. → pages 14

[22] M.R. Shortis, T.A. Clarke, and S. Robson. Practical testing of the precision and accuracy of target image centering alogrithms. In *Proceedings of SPIE, Videometrics IV*, volume 2598, pages 65–76, 1995. → pages 16, 17, 18

[23] M.R. Shortis, T.A. Clarke, and T. Short. Comparison of some techniqures for the subpixel location of discrete target images. In *Videometrics III. SPIE*, volume 2350, pages 239–250, 1994. → pages 16, 27

[24] J. Chen and T.A. Clark. The automatic location and identification of targets in digital photogrammetric engineering measurement. *Int. Arch. of Photogrammetry and Remote Sensing*, 29:686–693, 1992. → pages 18

[25] Nicholas A. Doudoumopoulos, Laruren Purcell, and Eric R. Fossum. CMOS Active Pixel Sensor Technology for High Performance Machine Vision Applications. In *SME Applied Machine Vision*, June 1996. → pages 25

[26] Dave Litwiller. CMOS vs. CCD: Facts and Fiction. *Photonics Spectra*, 2001. → pages vi, 26

[27] Nixon O. Applications Set Imager Choices. *Advanced Imaging*, 2008. → pages 26

[28] Dave Litwiller. CMOS vs. CCD: Machine Vision. *Sensors & Machine Vision*, 2007. → pages v, 26, 27

[29] DALSA Inc. Electronic Shuttering for High Speed CMOS Machine Vision. White paper accessed from www.dalsa.com/shared/content/pdfs/photonik_cmos_shuttering_english.pdf online on March 5 2011. → pages vii, 27, 56

[30] Cypress Semiconductor Inc. LUPA-4000 CMOS image sensor datasheet. Datasheet accessed from http://www.proscan.de/pdf-files/LUPA-4000 datasheet.pdf online on March 4 2011. → pages v, 28

[31] Jennifer Eyre and Jeff Bier. The Evolution of DSP Processors. *IEEE Signal Processing Magazine*, 17:43–51, 2000. → pages 29

[32] Linear Tecnology Inc. LTC2299 ADC datasheet. Datasheet accessed from http://cds.linear.com/docs/Datasheet/2299fa.pdf on March 9 2011. → pages v, 31

[33] National Instruments Inc. Choosing the Right Camera Bus. White paper accessed from http://zone.ni.com/devzone/cda/tut/p/id/5386? on March 13 2011. → pages v, 33, 34, 35, 37

[34] Darren Bessette. GigE Gains in Traction. *Advanced Imaging Magazine*, June 2010. → pages 35

[35] Camera Link. Specifications of the Camera Link Interface Standard for Digital Cameras and Frame Grabbers. Datasheet accessed from http://www.alacron.com/clientuploads/PDFs/forweb/CameraLinkSPEC.pdf on March 15 2011. → pages vi, 35, 36

[36] Kris Smeds and Xiaodong Lu. A multirate multiprocessor control platform for high-speed precision motion conrol. In *ASPE Annual General Meeting*, November 2010. → pages vi, 39

[37] Opnext Inc. HE8812SG infrared LED datasheet. Datasheet accessed from http://opnext.com/products/pdf/ode_2063_he8812sg.pdf on March 9 2011. → pages vii, 47

[38] Ann R. Thryft. Lens choices get more complicated. *Test and Measurement World*, pages 58–59, Janurary 2011. → pages 47

[39] Wikipedia. Fixed-pattern noise. Accessed from http://en.wikipedia.org/wiki/Fixed-pattern_noise on March 7 2011. → pages 55

[40] Kenji Irie, Alan E. McKinnon, Keith Unsworth, and Ian M. Woodhead. A Technique for Evaluation of CCD Video-Camera Noise. *IEEE Transactions on Circuits and Systems for Video Technology*, 18:280–284, Feburary 2008. → pages vii, 55, 56

[41] Verena Schneider. Fixed-pattern Correction of HDR Image Sensors. In *2005 PhD Research in Microeletronics and Electronics*, July 2005. → pages 55

[42] D. Joseph and S. Collins. Temperature Dependence of Fixed Pattern Noise in Logarithmic CMOS Image Sensors. *IEEE Transactions on Instrumentation and Measurement*, 58:2503–2511, August 2009. → pages 55

[43] Fabio Remondino and Clive Fraser. Digital camera calibration methods: considerations and comparisons. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI:266–272, 2006. → pages 57

[44] F. Hatze. High-precision three-dimensional photogrammetric calibraion and object space reconstruction using a modified DLT approach. *Journal of Biomechanics*, 21:533–538, 1988. → pages 58, 59

[45] D. Gazzani. Comparative assessment of two algorithms for calibraing stereophotogrammetric systems. *Journal of Biomechanics*, 26:1449–1454, 1993. → pages 58, 59

[46] O.D. Faugeras and M. Hebert. Calibration problem for stereo. In *Proceedings of International Conference Computer Vision Pattern Recognization*, pages 15–20, June 1986. → pages 58

[47] Mitutoyo Inc. Crysta-Apex C standard CNC CMM datasheet. Datasheet accessed from http://www.mitutoyo.com/pdf/1809Crysta-ApexC.pdf on March 9 2011. → pages v, 78