A HIGH ACCURACY CMOS CAMERA BASED

3D OPTICAL TRACKING SYSTEM DESIGN

by

Yingling Huang

B.Sc., Harbin Institute of Technology, 2006

M.Sc., Harbin Institute of Technology, 2008

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)

October 2010

Abstract

This thesis presents the design and implementation of a high accuracy optical tracking system. The primary contributions of this thesis include the integration of the 3D optical tracking systems, noise correction algorithms, target location algorithms, and camera calibration algorithms with lens distortion identification. With the best combination of these algorithms, the achieved 3D RMS error for the 3D position estimation is 41 μm for a measuring range of 500 mm by 500 mm on XY plane and 10 mm in Z.

For most machines and robots, MDOF motion sensing in real time is required. Current optical tracking systems for MDOF motion measurement are either low accuracy or low measurement rate. For instance, none of these systems can provide over 10 kHz measurement rate for real time sensing. The optical tracking system presented in this thesis can potentially achieve high measurement rate up to 10 kHz and micron level accuracy.

Target location algorithms are widely studied. However, to the best of the author's knowledge, no experimental comparison of these different algorithms in 3D optical tracking systems was found in the literature. In addition, no report on the accuracy comparison of different lens distortion calibration methods was presented in the literature. Actually, both target location and lens calibration methods have great impact on the accuracy of the 3D optical tracking. Therefore, this thesis also investigates on the performance evaluation of target location algorithms in the 3D tracking system and identifies the lens distortion model for the lenses.

Table of Contents

List of Tables

List of Figures

viii

Nomenclature

Symbols

A_0 Aperture area of the lens in [m²]

A_i Target image area on the image sensor in [m²]

α_c Camera field of view in [degree]

c Speed of light in the air, $c = 2.998 \times 10^8$ m/s

d Distance from lens optical center to the image sensor in [m]

d_A Aperture diameter of the lens in [m]

$E_e(\lambda)$ Irradiance on the lens in [W/m²]

f Focus length of the lens in [m]

F F-number of the lens, $F = f/d_A$

G_c Conversion gain from electrons to voltage in [V/e⁻]

G Target moving range along X axis in [m]

H Target moving range along Z axis in [m]

h Planck's constant, $h = 6.626 \times 10^{-34}$ J·s

$\boldsymbol{I_0}$ Matrix for raw images

$\boldsymbol{I_{blk}}$ Matrix for black images

$\boldsymbol{\bar{I}_{blk}}$ Matrix for averaged black images

$\boldsymbol{I_c}$ Matrix for corrected images

$\boldsymbol{I_{ij}}$ Intensity matrix for images

(i,j) Pixel location on the image

K_1, K_2, K_3	Radial lens distortion coefficients
L	The length of the image sensor along X axis in [m]
$L_s(\lambda)$	Radiance of the light source at wavelength of λ in [W/sr]
M	Maximum blurred image size in [m]
m	Total column number for region of calculation in target location algorithms
n	Total row number for region of calculation in target location algorithms
P_0	Optical power output of the light source in [Watt]
P_1, P_2	Decentering distortion coefficients
q	Elementary charge, $q \approx 1.6 \times 10^{-19}$ C
λ	Wavelength of the light in [m]
$\boldsymbol{R}$	Rotation vector for rigid body transformation
S_1, S_2	Thin prism distortion coefficients
R_1	Distance from the tracking target to the lens in [m]
R_2	Distance from the lens to the image sensor in [m]
$R_s(\lambda)$	Responsivity of the image sensor in [V/(J/m^2)]
$R_{sa}(\lambda)$	Responsivity of the image sensor in [A/W]
T_{atm}	Transmittance of the atmosphere [%]
$T_f(\lambda)$	Transmittance of the filter in [%]
$T_l(\lambda)$	Transmittance of the lens in [%]
$\boldsymbol{T}$	Translation vector for rigid body transformation
(t_1, t_2, t_3)	Translation parameters for rigid body transformation
t_{exp}	Exposure time of the image capture in [s]

V	Voltage output from the image sensor in [Volt]
$W_s(\lambda)$	Energy on the image sensor in [Joule]
$W_p(\lambda)$	Energy intensity on the image sensor in [J/m^2]
(x_c, y_c)	Target location on the image
(x_i, y_i)	Ideal image coordinates on the image sensor
(x_d, y_d)	Distorted image coordinates on the image sensor
(X_w, Y_w, Z_w)	3D position in a world coordinate system
(X_c, Y_c, Z_c)	3D position in a camera coordinate system
α, β, γ	Rotation angle around X axis (α), Y axis (β) and Z axis (γ)
Ω	Solid angle of the light source in [sr]
δ	Distance between the shape image location to the image sensor location in [m]
θ	Viewing angle from the light source in [degree]
$\eta(\lambda)$	Power distribution coefficient at wavelength of λ in [%]
η_{QE}	Quantum efficiency of the image sensor in [%]

Acronyms

2D	Two dimensional
3D	Three dimensional
ADC	Analog to digital converter
CCD	Charge coupled device
CMOS	Complementary metal oxide semiconductor
CMM	Coordinate measuring machine
CCS	Camera coordinate system

DOF   Degree of freedom

DLT   Direct linear transformation

FPN   Fixed pattern noise

LED   Light emitting diode

IR LED   Infrared light emitting diode

MDOF   Multiple degree of freedom

NDLT   Proposed nonlinear optimization method based on DLT

PRNU   Pixel response non uniformity

RMS   Root mean square

WCS   World coordinate system

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Professor Xiaodong Lu for all his support, guidance and help on my research work. I have learned a lot from him through taking courses that he taught, discussions in weekly meetings and lab presentations. His critical thinking helps me to establish a solid foundation for all my knowledge and deep understanding to all the problems. His passion on doing world-class research always triggers me to do better. Working with him for more than 2 years has made me more professional and more confident.

I would like to thank Professor Yusuf Altintas. He taught me a very practical course, machine tools and vibration (MECH 592), and also provided the access to the UBC Metrology Laboratory where I have accomplished all of my experiments based on the CMM. I would also like to thank Professor Ryozo Nagamune and Professor Hsi-Yung (Steve) Feng for being in my thesis committee. Besides, Professor Ryozo Nagamune taught me a useful course on controller design and gave me some valuable suggestions on nonlinear optimization algorithms.

My implementation could not be succeeded without the assistant of Kris Smeds, one of my colleagues at UBC. It was him who found out CMOS solution and it was him who helped me select the CMOS camera which is one of the most important parts for my hardware. His suggestions and comments on my research benefit me a lot. Niankun Rao, another colleague at UBC, also gave me help on tracking algorithms design and some useful comments on my research. I also appreciate the help, comments and suggestions from my other colleagues, including Matthew Paone, Richard Graetz, Irfan Usman and Arash Jamalian. It was great experience for me to work with those smart people. Especially, I would like to thank Matthew

For Mom, Dad, Brother, Sister in Law and Shunshun

献给妈妈，爸爸，哥哥，嫂子和顺顺

Chapter 1

Introduction

1.1 Motivation

Most machines and robots are multiple-degree-of-freedom (MDOF) systems which generally require MDOF motion sensors for real time motion control. The simple but widely used idea for MDOF sensing is to measure each DOF by a separated sensor and combine the measurements to obtain MDOF displacements. The main disadvantage of this solution is that the accuracy of the motion sensing highly relies on the performance of the supporting bearings. One illustrative example is the X-Y table, which provides large X and Y movements by employing a stack of two linear motors [1] mounted on roller bearings. If there is a part P installed on the X stage, as shown in Figure 1.1, this part has 2DOF motion.
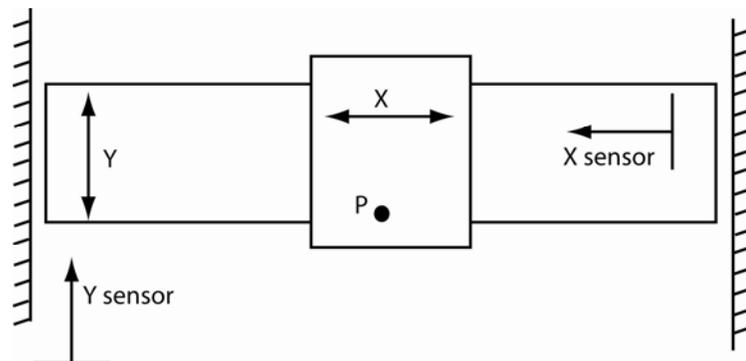
Figure 1.1: XY sensing for a part P installed on an X-Y table.

In the traditional method, the motion of the part P can be derived by an X sensor and a Y sensor. The obtained results can be very accurate only if the Y stage moves along Y direction without any lateral or parasitic yaw motion and the X stage moves along X direction only. The limitation for this method is that high accuracy sensing requires high quality bearings. To solve this problem, an integrated 2D sensor, which can achieve high accuracy without high quality bearings, is demanding. Another example for MDOF sensing is robot arms. It is required to obtain high accuracy of MDOF sensing for precise control of the robot arms. So the demand on MDOF sensing all in one sensor with high accuracy motivates the research in this thesis. The requirement for the accuracy of this work is in micrometer range and the measurement range is in meters using a single sensor.

1.2   Existing solutions

For MDOF sensing with high accuracy, optical tracking systems turn out to be a very attractive solution. Optical tracking systems can cover a wide measuring range using optical sensors and have the potential of achieving high accuracy and high measurement rate. Furthermore, optical tracking systems have already been widely used in real time robotics guidance [2], 3DOF or 6DOF measurement [3] [4], handheld 3D scanners [5], image guiding surgery [6] [7], virtual reality/augmented reality [8] [9] [10] and so on. There are a variety of optical tracking systems available in the world. A review of existing optical tracking systems is given as follows.

1.2.1 Laser Tracker

The Laser Tracker is a 3DOF motion measurement system based on a combination of two techniques. One technique is applying a laser interferometer to measure relative distance from the laser source to the tracking target (a steering mirror). The other is applying optical encoders to measure azimuth (α) and elevation of the tracking target (β). The principle of laser interferometer for distance measurement is shown in Figure 1.2 (a). The laser interferometer splits the laser source into two beams using the beam splitter. One beam is used as a reference while the other beam is reflected back from a mirror at some distance. It is then merged with the reference beam to produce interference. As the wavelength of the laser is known, the distance can be calculated by counting the number of interference fringes. With the distance (d) measured from the laser interferometer and two angles (α, β) measured from optical encoders, 3D Cartesian coordinates of the tracking target P(x, y, z) can be calculated based on triangulation, as shown in Figure 1.2 (b).

(a)   (b)
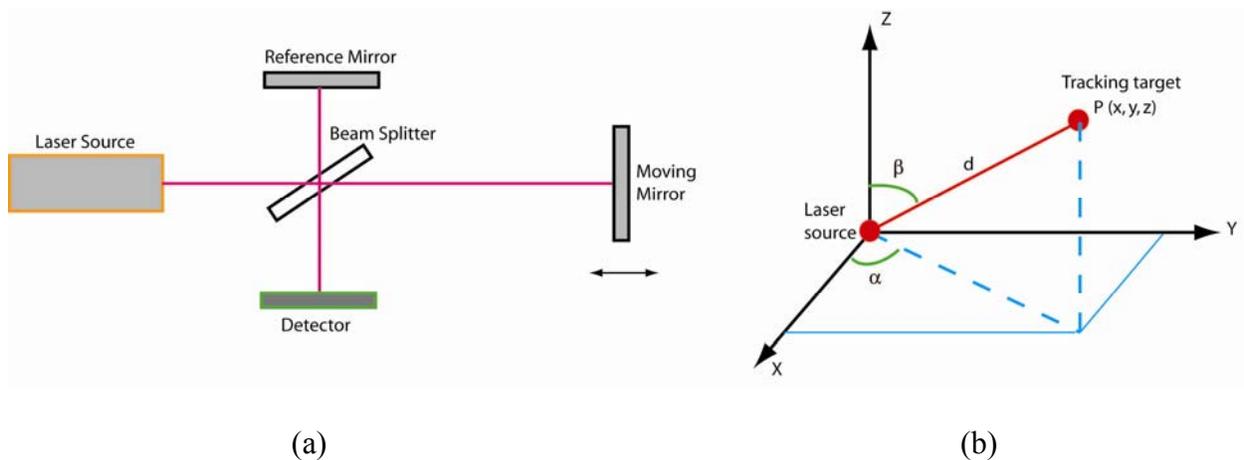
Figure 1.2: (a) Distance measurement principle for a laser interferometer; (b) Polar coordinates converted to Cartesian coordinates based on triangulation for the Laser Tracker.

There are several advantages for the Laser Tracker, such as large measurement range (up to tens of meters) [11] and highly accurate in controlled laboratories or stable working environments [12]. For instance, 3D Laser Tracker LTD500 from Leica [13] can achieve measurement accuracy of 20 μm within 1 meter range. However, this method needs to re-establish tracking each time the laser beam is broken. Besides, this method requires extensive accessories to perform 6DOF measurements, and has heavy mechanical parts which increases the potential maintenance cost [12].

1.2.2 HiBall tracking system

The HiBall tracking system [14] is a 6DOF tracking system using a HiBall optical sensor and ceiling LED arrays, as shown in Figure 1.3. The HiBall optical sensor is composed of 6 linear effect photo diodes (LEPD) and 6 lenses [15] arranged to view LED arrays mounted on the ceiling. Based on some geometric relation of these observed LEDs, this system can fulfill 6DOF measurement for the motion of the HiBall optical sensor relative to the ceiling.
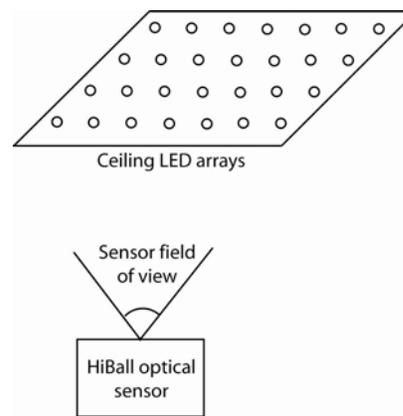
Figure 1.3: Simple structure for the HiBall tracking system.

This tracking system covers a very wide measuring area (5 m by 5 m) and achieves a measurement rate up to 2000 Hz. However, its measurement accuracy is very low as the position noise is 0.5 mm and rotation noise is less than 0.03˚.

1.2.3 Cyclope Tracker

The Cyclope Tracker [3] is a 6DOF optical tracking system using one CCD camera and several retro-reflectors, as shown in Figure 1.4. The retro-reflectors have a pre-defined pattern and reflect infrared light from the infrared LEDs mounted on the lens. The camera is used to capture images for these retro-reflectors. Based on one image and known geometric relation of these retro-reflectors, 6DOF motion of the retro-reflectors relative to the camera can be calculated.
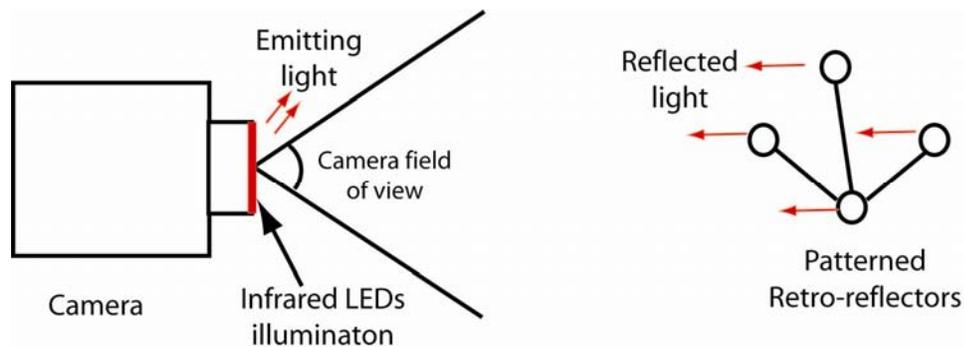
Figure 1.4: Schematic of the Cyclope Tracker using one camera and patterned retro-reflectors.

The Cyclope Tracker is a solution for 6DOF sensing and it also a wide field of view. However, this tracking system offers a low measurement rate of 60 Hz and low accuracy of 1 mm in translation and 0.3˚ in rotation with a measuring depth of 1.5 m [3].

1.2.4   ARTtrack

The ARTtrack [16] is a 6DOF optical tracking system using 2 CCD cameras and several retro-reflectors. Similar to the Cyclope Tracker, the ARTtrack also uses infrared LEDs mounted on the lens to provide illumination for the retro-reflectors. Two cameras are used to capture images for the retro-reflectors. Based on two images from two different cameras and triangulation method, 3D motion measurement of each retro-reflector can be calculated. Using 3D information of 3 retro-reflectors, 6DOF motion measurement can be obtained.
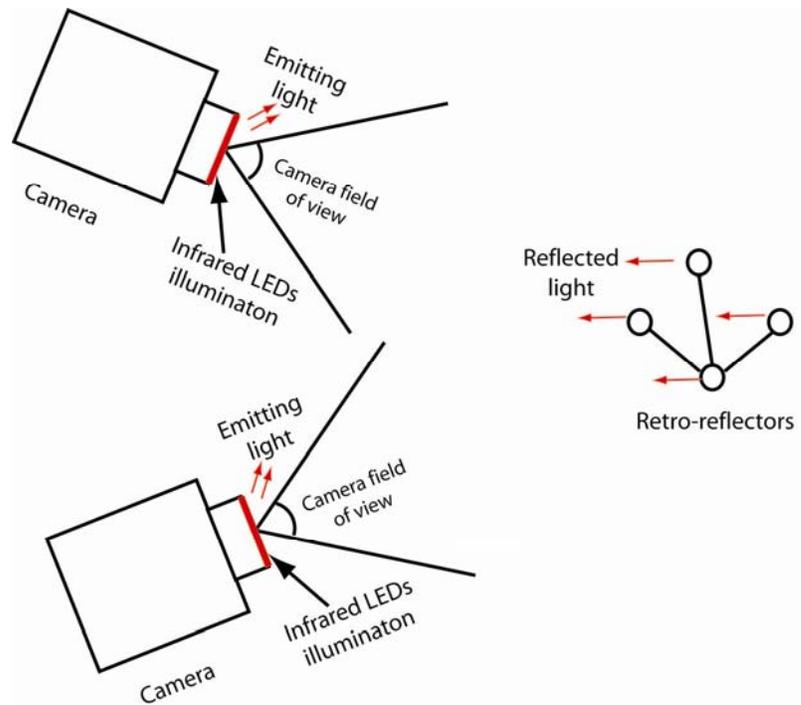
Figure 1.5: Schematic of the ARTtrack using 2 cameras and retro-reflectors.

The advantage of the ARTtrack is a large measuring depth (up to 6 meters). The disadvantages of this system are a low measurement rate of 60 Hz and low accuracy of 0.04 pixels.

1.2.5 OPTOTRAK

The OPTOTRAK is a 6DOF motion measurement system from Northern Digital Inc. It has three linear CCDs and uses infrared LEDs as the tracking targets, as shown in Figure 1.6. Three linear CCDs are aligned in different directions in order to cover each single DOF of the tracking target using one linear CCD. For instance, X motion of the tracking target P can be obtained using linear CCD I, Y motion of P can be obtained using linear CCD II, and Z motion of P can be obtained using linear CCD III. Then three linear CCDs combined together provides 3D position measurement for the tracking target and 6DOF motion measurement can be achieved by using 3 infrared LEDs simultaneously.
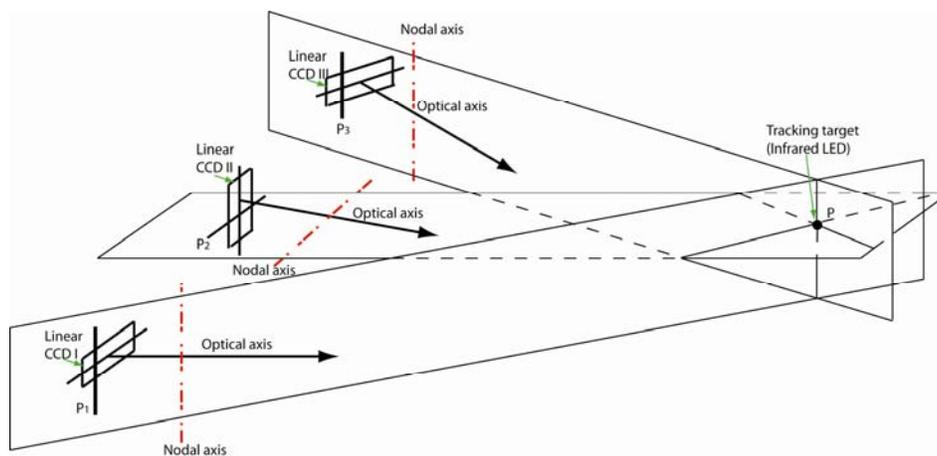
Figure 1.6: Schematic of OPTOTRAK using 3 linear CCDs and infrared LEDs [17].

The advantages of OPTOTRAK are that it can achieve high accuracy for 6DOF motion measurement and have a large tracking volume. For instance, the measurement accuracy is 15 μm in X, 15 μm in Y and 50 μm in Z with a measuring depth of 1.5 m. However, the measurement rate of this optical tracking system is low (less than 1 KHz for non-contact sensing).

1.2.6   Summary of the review

To summarize the reviewed optical tracking systems, a comparison table for these systems is shown in Table 1.1. It is concluded that none of the optical tracking system can provide 6DOF motion measurement with high accuracy and high speed.

Table 1.1: Comparison of existing optical tracking systems.

Name	MDOF	Measuring Rate(Hz)	Measuring Depth (m)	Measurement accuracy
LTD500 Laser Tracker (Leica)	3DOF	20,000	> 10 m	20 μm within 1 m
Hiball-3100 tracker (UNC)	6DOF	2000	Up to 5 m	< 0.2 mm (position) < 0.01° (rotation)
The Cyclope Tracker (Immersion Company)	6DOF	60	Up to 1.5 m	< 1 mm (position) < 0.3° (rotation)
A.R.T tracking system (ART)	6DOF	60	Up to 6 m	0.04 pixel resolution
OPTOTRAK (Northern Digital Inc.)	6DOF	1069 (wired) 847 (wireless)	1.5 m to 3 m	15 μm in X and Y; 50 μm in Z

1.3   Proposed solution

To achieve 6DOF motion measurement with high accuracy (<50 μm) and high measurement rate (>10 kHz), the proposed solution is to use two CMOS cameras and 3 infrared LEDs. The schematic for the proposed solution is shown in Figure 1.7. Based on two target locations on the left and right image sensors and triangulation methods, the 3D position of the target (LED) can be calculated. The 6DOF motion measurement can be obtained using 3 LEDs.
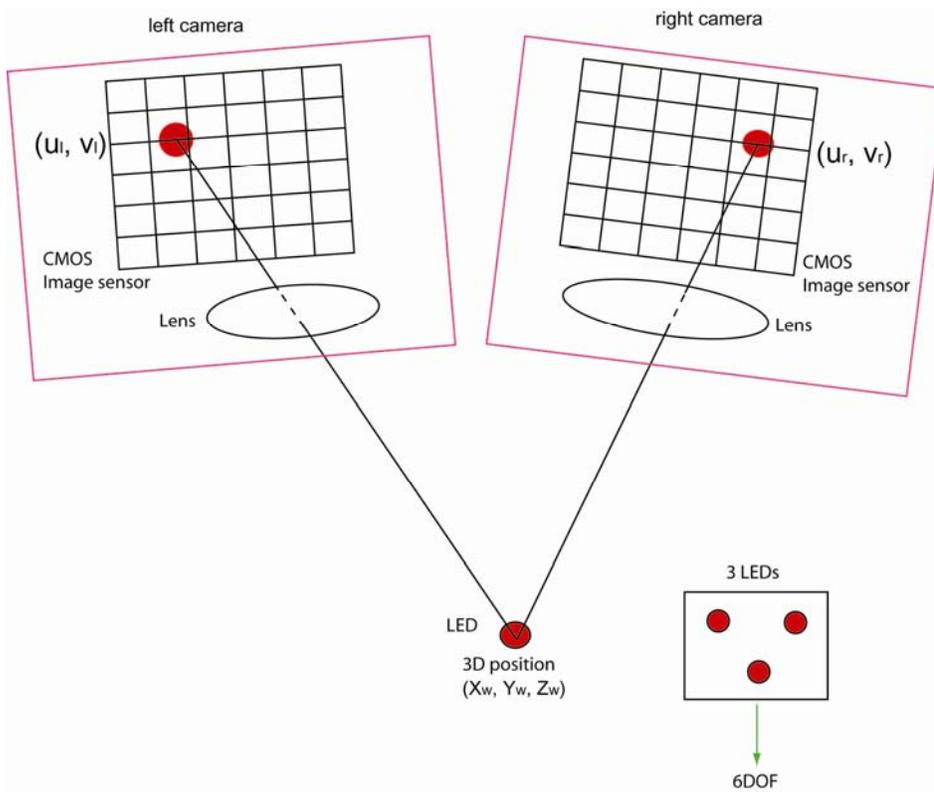
Figure 1.7: Schematic for the proposed optical tracking system.

The proposed solution of using CMOS image sensors instead of CCD image sensors for the 6DOF optical tracking system design has the potential of achieving a high frame rate and

high accuracy. For instance, LUPA 4000 from Cypress can achieve a high frame rate up to 10 kHz with a region of interest (ROI) of 25 pixels x 25 pixels.

Compared to all the available optical tracking systems, this solution has the potential of achieving high speed and high accuracy for the 6DOF motion measurement.

Due to limited time, this thesis mainly tackles the challenging of fulfilling the potential of high accuracy (<50 μm) for the proposed solution using CMOS cameras, which have low image quality and high noise level compared to CCD cameras. Besides, this thesis mainly focuses on 3D motion measurement with one tracking target while 6DOF motion measurement can be easily obtained using 3 tracking targets.

1.4   Thesis overview

This thesis is concerned with the design and integration of a CMOS camera based 3D optical tracking system with high accuracy. It consists 6 chapters as follows:

Chapter 2 introduces the optical tracking system hardware. In this chapter an overview of the hardware for the 3D optical tracking systems is given. The hardware includes tracking targets, image sensors, filters, lenses, frame grabbers and a computer, as shown in Figure 1.8. Model analysis for the optical tracking systems is explained later with formulas for calculating light intensity and voltage output. All the hardware component selections are presented and the effect of the ambient light is also analyzed to verify the feasibility of the hardware selection.
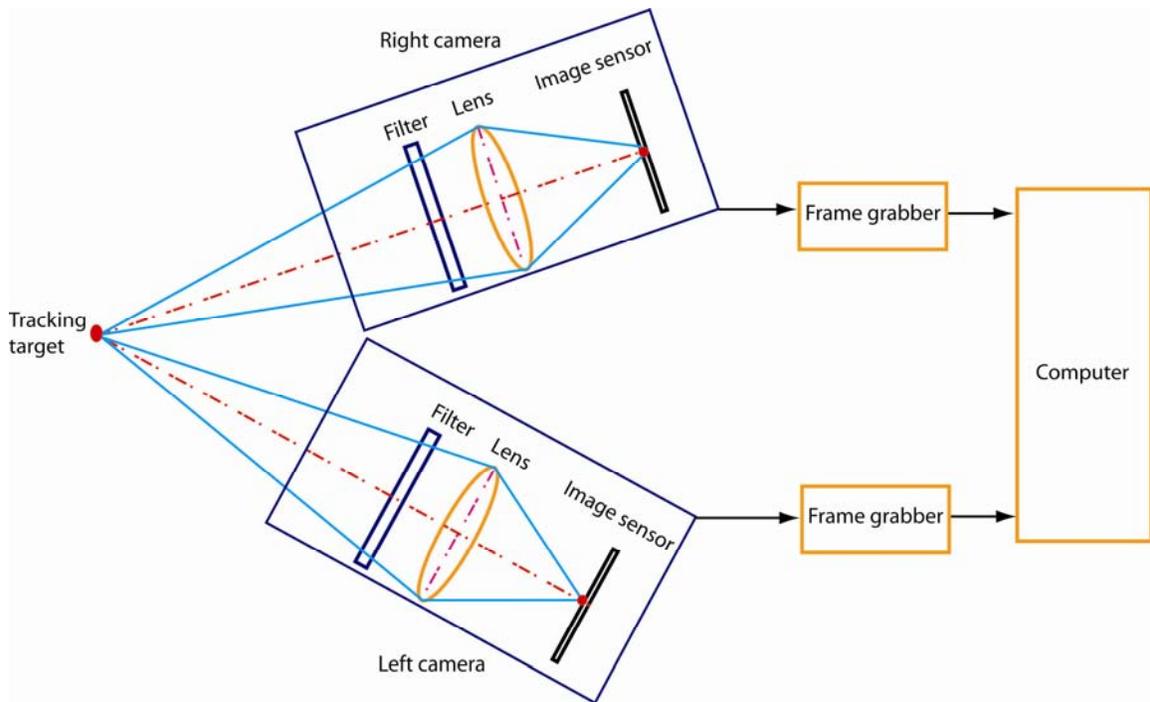
Figure 1.8: Hardware structure for 3D optical tracking systems.

Chapter 3 describes the 3D optical tracking algorithms. An overview for all the algorithms is given and Figure 1.9 shows the flow chart for the 3D tracking algorithms. Noise correction algorithms for processing raw images are presented first and followed by the introduction of target location algorithms. Simulation investigation for target location algorithms are also presented in this chapter.
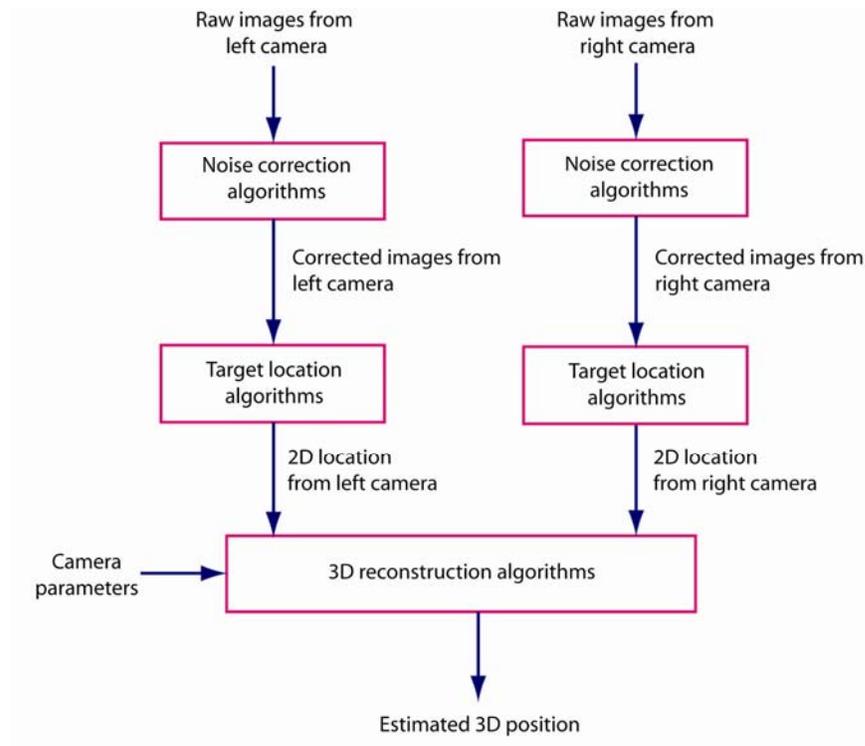
Figure 1.9: Flow chart of the 3D optical tracking algorithms.

Chapter 4 presents camera calibration and 3D reconstruction algorithms. The flow chart of camera calibration is seen in Figure 1.10. The camera model with the definition of coordinate frames and a detail explanation for the four steps of image formation process as well as full camera model are described. This chapter also reviews most of the previous work related camera calibration with advantages and disadvantages. Then the proposed solution for camera calibration, involving direct linear transformation (DLT) algorithms for an initial guess and nonlinear least square optimization for a full scale optimization, is given. The simulation investigation for the DLT method and the proposed solution under different conditions are also discussed.
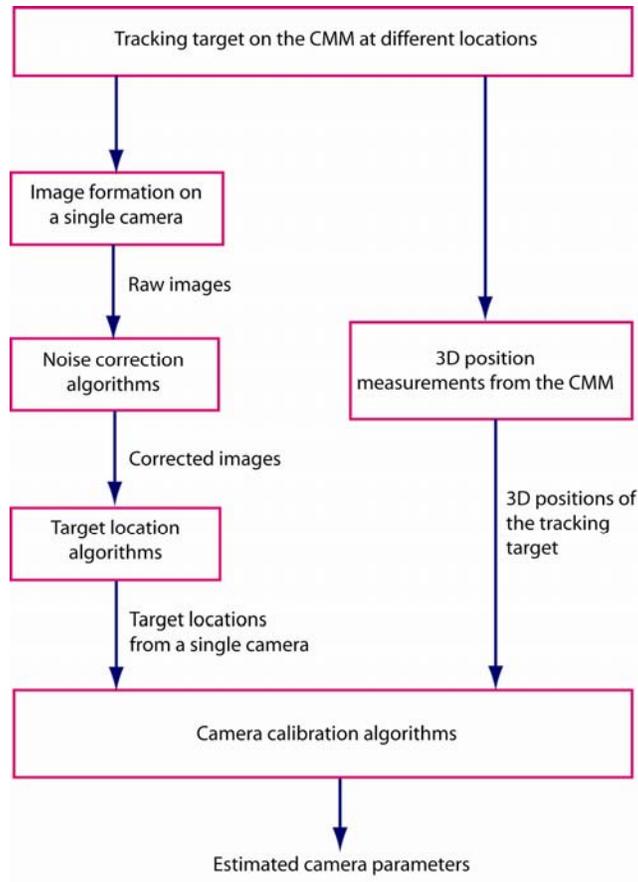
Figure 1.10: Flow chart of camera calibration for a single camera.

Chapter 5 explains the experimental results and discussion. A study of the implemented system performance is presented through several examples. Comparison tests on IR LEDs are explained in order to prove the best performance of the selected IR LED. Experimental results of the FPN correction, PRNU correction, repeatability tests, and tests for evaluating the designed 3D optical tracking system are all demonstrated in this chapter.

Chapter 6 describes the conclusions and future work. The contribution of the thesis, along with some suggestions for the future work, is also summarized in this chapter.

Chapter 2

3D Optical Tracking System Hardware

As a fundamental part of the optical tracking system, the hardware of the 3D Optical tracking system determines the performance for the 3D tracking, such as accuracy, speed and volume. In addition, the hardware is often the critical limitation for the performance improvement once the algorithms of the system have been well designed and implemented.

For a 3D optical tracking system, the hardware consists of image sensors, lenses, filters, tracking targets, frame grabbers and a computer. These components have great influence on the performance of the tracking system. For instance, the resolution of the image sensors determines the achievable highest tracking accuracy, the viewing angle of the lenses decides the tracking volume, the transmittance of the optical filters affects the signal to noise ratio of the tracking system, and the stability of the tracking targets has large impact on the tracking repeatability. As the hardware is crucial for all the optical tracking systems, it is required that considerable attention should be paid to each component selection.

In this chapter, the details for the system hardware selection and analysis are presented. More specifically, Section 2.1 gives an overview of the system hardware and Section 2.2 introduces model analysis for the system signal transmission. Section 2.3 presents considerations and selection of the hardware components, including image sensors, optical lenses, tracking targets, and optical filters. Section 2.4 describes the analysis of the ambient light effect on the

tracking system and verifies the feasibility of the system. Finally, this chapter is closed with a summary in Section 2.5.

2.1   Tracking system hardware overview

The hardware structure for the 3D optical tracking system is shown in Figure 2.1. Here, a light source is used as a tracking target which can be easily seen from left and right image sensors and form images on the image sensors with optical lenses. Before the light reaches the optical lens for each camera, optical filters are required to pass all the light from the light source and block the ambient light to reduce the background noise. After the image formation on the image sensors, frame grabbers are used to convert the analog image data to the digital image data and transfer all the digital images from image sensors to a computer. Based on the computer, all the algorithms of the 3D optical tracking system such as 3D tracking algorithms and calibration algorithms are implemented and evaluated using all the raw images from the frame grabbers.

In the following sections, model analysis for the optical tracking systems is presented, followed by the selection of each hardware component.
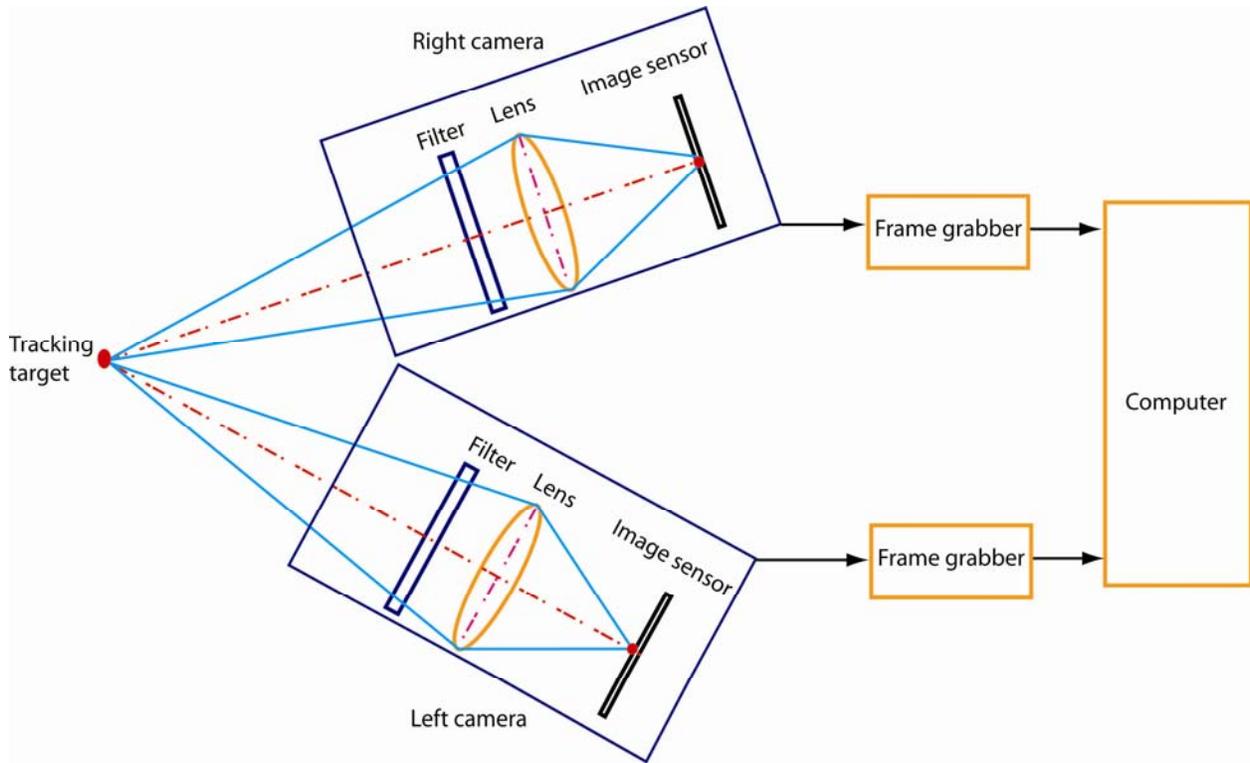
Figure 2.1: Hardware structure for 3D optical tracking systems.

2.2 Model analysis for the 3D optical tracking systems

The model analysis for 3D optical tracking systems is to model the signal transmission process from the light source to the image sensor and calculate the voltage output of each pixel on the image sensor with a specific optical power input. The signal transmission process is to convert the optical power of the light source into a measurable voltage on the image sensor. Figure 2.2 shows a camera directly viewing the light source. It is seen that the viewing angle of the light source is 2θ, R_1 is the distance from the light source to the lens, and R_2 is the distance from the lens to the image sensor. In this section, mathematical models for light source radiance calculation and voltage output calculation are presented.
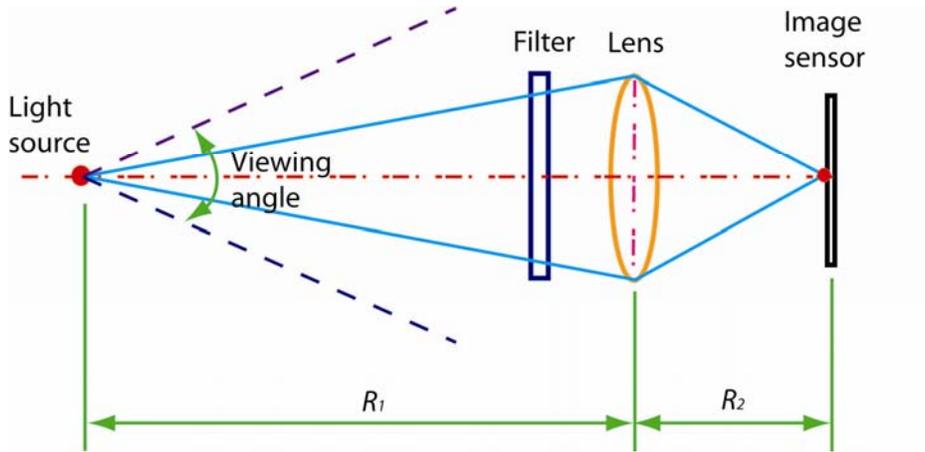
Figure 2.2: Light transmission from the light source to the image sensor.

2.2.1 Radiance calculation for the light source

Assume that the light source has an optical power P_0 with the units of Watt as the optical power output and the viewing angle of the light source is 2θ in degree, shown in Figure 2.3. The solid angle of the light source with the unit of steradian (sr) can be expressed as

$$\Omega = 2\pi(1 - \cos\theta). \tag{2.1}$$
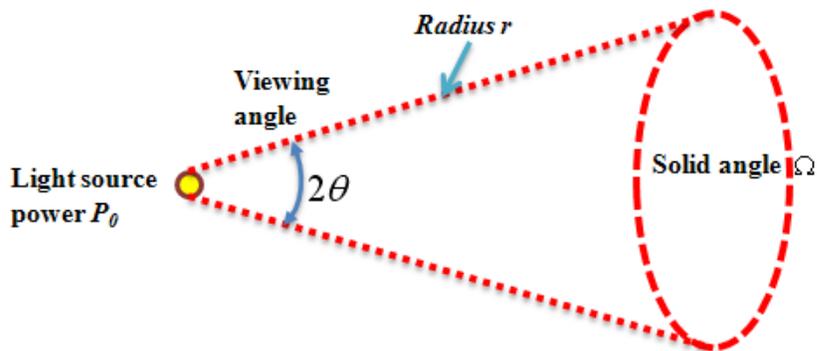
Figure 2.3: Light emission from the light source.

17

The light sources are not a single wavelength of light. For example, white light is a wide range of wavelengths with red, blue and green colors, and even laser light sources contain more than a single wavelength. As a result, it is natural to assume that the light source in the 3D optical tracking system contains a certain range of wavelengths. The units of the wavelengths are meters, with the terms microns (denoted μm and equal to $10^{-6} m$) and nanometers ($10^{-9} m$) being used just as frequently [18]. The irradiance of a certain wavelength λ from the light source is related to the optical power of this specific wavelength, while the power distribution of the light source determines the power output for a specific wavelength. So the expression for the radiance of wavelength λ of the light source is represented by

$$L_s(\lambda) = \frac{P_0 \eta(\lambda)}{\Omega} ,$$ (2.2)

where $L_s(\lambda)$ is the radiant intensity of the light source at wavelength λ with unit of W/sr, and $\eta(\lambda)$ is the power distribution coefficient for wavelength λ.

2.2.2   Pixel voltage calculation

The block diagram shown in Figure 2.4 illustrates the signal transmission from the optical power output to the pixel voltage output on the image sensor. Assuming that the distance of the light source to the lens is R_1 with units of meters and the aperture size of the lens is A_0 with units of m^2, the energy received by the image sensor $W_s(\lambda)$ with unit of Joule can be calculated by

$$W_s(\lambda) = L_s(\lambda) \frac{A_0}{R_1^2} T_{atm} T_f(\lambda) T_l(\lambda) t_{exp} ,$$ (2.3)

where $L_s(\lambda)$ is the radiant intensity of the light source at wavelength λ, T_{atm} is the intervening atmospheric transmittance which is close to 1 in air, $T_f(\lambda)$ is the transmittance of the filter, and $T_l(\lambda)$ is the transmittance of the lens for wavelength λ. Eqn. (2.3) is based on the small angle approximation (valid when $R_1^2 \gg A_0$) [19].
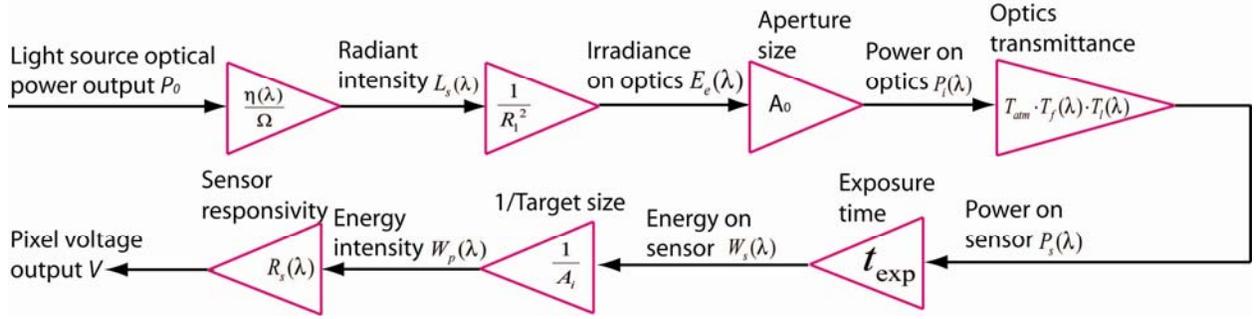
Figure 2.4: Block diagram of signal transmission from the light source to the image sensor.

As shown in Figure 2.4, the image size of the light source forming on the image sensor is A_i with unit of m^2. To simplify the calculation, the energy of the light source is assumed to be uniformly distributed within the image area of the light source. Then the energy intensity with unit of J/m^2 can be obtained as

$$W_p(\lambda) = \frac{W_s(\lambda)}{A_i}. \tag{2.4}$$

In summary, the pixel voltage output on the image sensor is

$$
\begin{aligned}
V &= \int_{\lambda_1}^{\lambda_2} W_p(\lambda) \cdot R_s(\lambda) d\lambda \\
&= \int_{\lambda_1}^{\lambda_2} L_s(\lambda) \cdot \frac{1}{R_1^2} \cdot A_0 \cdot T_{atm} \cdot T_f(\lambda) \cdot T_l(\lambda) \cdot t_{exp} \cdot \frac{1}{A_i} \cdot R_s(\lambda) d\lambda
\end{aligned}
\tag{2.5}
$$

where $R_s(\lambda)$ is the responsivity of the image sensor with unit of $V/(J/m^2)$, λ_1 and λ_2 is the wavelength range for the light source. The responsivity of the image sensor $R_s(\lambda)$ is related to the light source spectrum, quantum efficiency of the image sensor, conversion gain and image size on the image sensor. It can be represented by

$$R_s(\lambda) = \eta_{QE}\frac{\lambda}{hc}G_cA_i,$$
(2.6)

where η_{QE} is the quantum efficiency of the image sensor; λ is the wavelength of the light source with unit of meter, h is the Planck's constant ($h = 6.626\times10^{-34}$ J·s), c is the speed of light in the air ($c = 2.998\times10^8$ m/s), q is the elementary charge ($q \approx 1.6\times10^{-19}$ C), and G_c is the conversion gain from electrons to voltage output with unit of V/e⁻.

However, there is another definition for the responsivity of the image sensor with unit of A/W. In this case, the responsivity of the image sensor only depends on the wavelength of the light source and quantum efficiency of the image sensor, which can be expressed as

$$R_{sa}(\lambda) = \eta_{QE}\frac{\lambda}{hc}q.$$
(2.7)

The conversion from the responsivity $R_{sa}(\lambda)$ with unit of A/W in the above equation into the responsivity $R_s(\lambda)$ with unit of $V/(J/m^2)$ can be represented by

$$R_s(\lambda) = G_cA_i\frac{R_{sa}(\lambda)}{q}.$$
(2.8)

Combining Eqn. (2.5) and (2.8), the voltage output with responsivity $R_{sa}(\lambda)$ is provided as

$$V = \int_{\lambda_1}^{\lambda_2}W_p(\lambda)\cdot R_s(\lambda)d\lambda$$
$$= \int_{\lambda_1}^{\lambda_2}L_s(\lambda)\cdot\frac{1}{R_1^2}\cdot A_0\cdot T_f(\lambda)\cdot T_l(\lambda)\cdot t_{exp}\cdot G\cdot\frac{R_{sa}(\lambda)}{q}d\lambda$$
(2.9)

So the voltage output for the system signal transmission can be calculated using Eqn. (2.5) or Eqn. (2.9) depending on the image sensor selected and given responsivity of the image sensor provided by the manufactures.

The voltage output from the image sensor is then amplified and processed using signal amplifiers and ADCs on the electronics for sensor readout. The final data output from the cameras are expressed in images with digital intensity values of 0~255 for 8 bits ADC or 0~4095 for 12 bits ADC.

2.3   Tracking system hardware selection

2.3.1   Image sensor selection

Image sensors are devices that convert light into electrons. In general, there are two types of image sensors, charge-coupled device (CCD) and complementary metal oxide semiconductor (CMOS). In a CCD image sensor, every pixel's charge is transferred through a very limited number of output nodes (often just one) to be converted to voltage, buffered, and sent off-chip as an analog signal. All the pixels can be devoted to light capture, and the output uniformity (a key factor in image quality) is high. In a CMOS image sensor, each pixel has its own charge-to-voltage conversion, and the sensor also includes amplifiers and digitization circuits to output digital bits. These other functions of CMOS sensors increase the design complexity and reduce the light to charge area available for light capture [20]. The structures of both CCD and CMOS image sensors are shown in Figure 2.5. It is seen that the CCD image sensors move photon generated charge from pixel to pixel and convert it to voltage at an output node. The CMOS

image sensors convert charge to voltage inside each pixel and transfer the signal out through line or column amplifiers [21].

(a) (b)
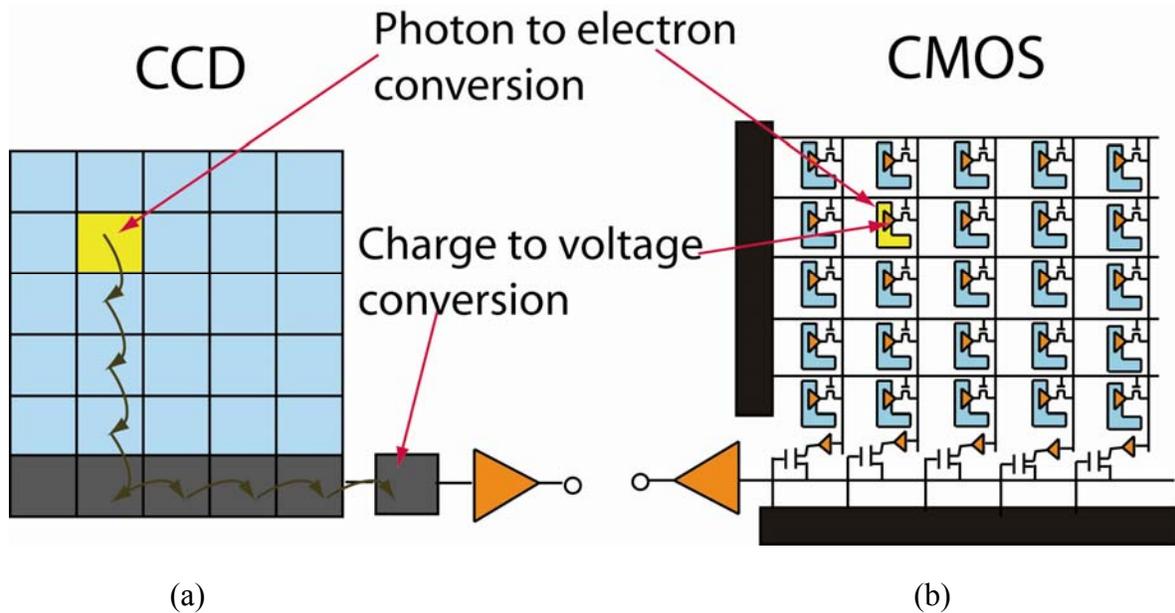
Figure 2.5: (a) Structure of CCD image sensors; (b) Structure of CMOS image sensors [20].

Due to these differences, the CCD image sensors are now widely used in cameras that focus on high quality images since they have high pixel count and excellent light sensitivity [22]. The CMOS image sensors, on the other hand, are more often used in high speed video cameras for their unique ability of programmable windowing which can achieve over 10,000 frame rate per second with small region of interest (ROI), such as 25x25 pixels. Because of the intense research on the CMOS image sensors, the CMOS image sensors can produce image qualities that are comparable to the CCD image sensors.

For the 3D optical tracking system design, high frame rate over 10,000 frames per second is required. So the CMOS image sensor is considered as it can achieve high frame rate with small ROI, while CCD are not capable of doing this. The LUPA 4000 made by Cypress Semiconductor

Corporation, which allows high frame rate (>10,000 Hz) with small ROI is selected for the system. It has 2048 x 2048 pixels with pixel size of 12 μm x 12 μm and supports random programmable windowing at high frame rate. However, as the LUPA 4000 is only an image sensor, it requires additional power supplies and circuitry to be functional. Later SI-4000F MegaCamera[TM] from Silicon Imaging with imbedded LUPA 4000 is found. This CMOS camera can output 12 bits monochrome image data and has high speed widowing ability of over 10,000 fps with ROI of 25 x 25 pixels and 6 tracking targets. The exposure time is also controllable from 69 μs to ~150 ms @ 30 MHz pixel rate with 12 bits image data output. The SI-4000F MegaCamera[TM] with embedded LUPA 4000 are shown in Figure 2.6.
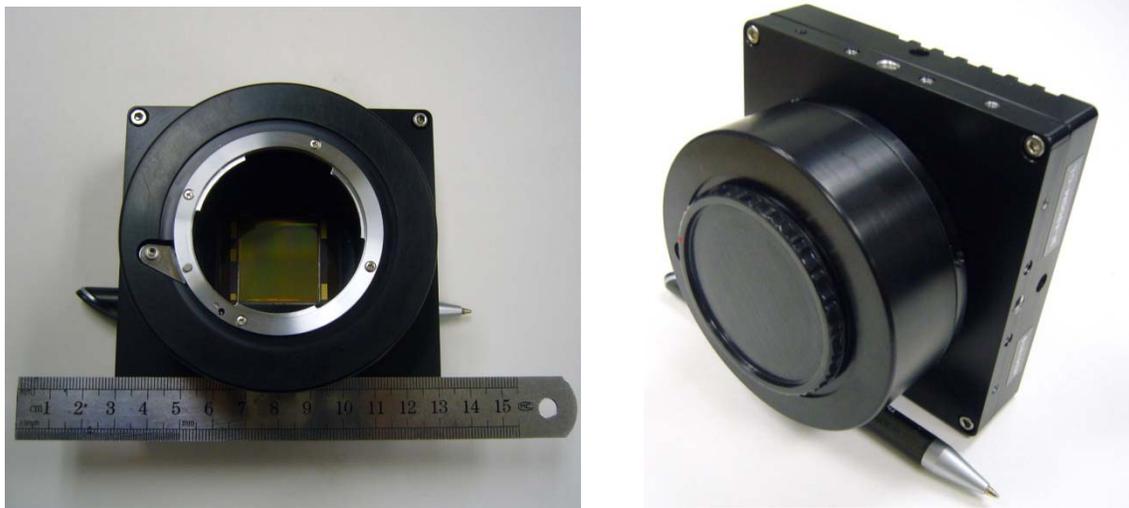
Figure 2.6: The CMOS camera (SI-4000F MegaCamera[TM]) with image sensor of LUPA 4000.

Once the camera is selected, the frame grabber and camera link cables can also be decided based on the camera data output. The frame grabber chosen is the PIXCI® EL1DB, from Epix Inc, shown in Figure 2.7 (a). The frame grabber is a dual base camera link frame grabber in one PCI Express x1 Expansion Slot. It has 250MB/sec data transfer rate and 2 camera control

ports [23]. The camera link cables for camera and frame grabber connection are also from Epix Inc and it is shown in Figure 2.7 (b).

<center>(a)   (b)</center>

Figure 2.7: (a) Frame grabber (PIXCI® EL1DB from EPIX); (b) Camera link cable.

2.3.2 Lens selection

It is well known that a complete camera includes an image sensor and a lens. The lens affects the performance of the 3D optical tracking system because it is the fundamental component for the image formation process. Figure 2.8 shows the basic principle for image formation using a lens. R_1 and R_2 represent the distance from the object to the lens and the distance from the lens to the image sensor respectively, while f is the focal length of the lens. The units of these quantities are in meters.

<center>24</center>
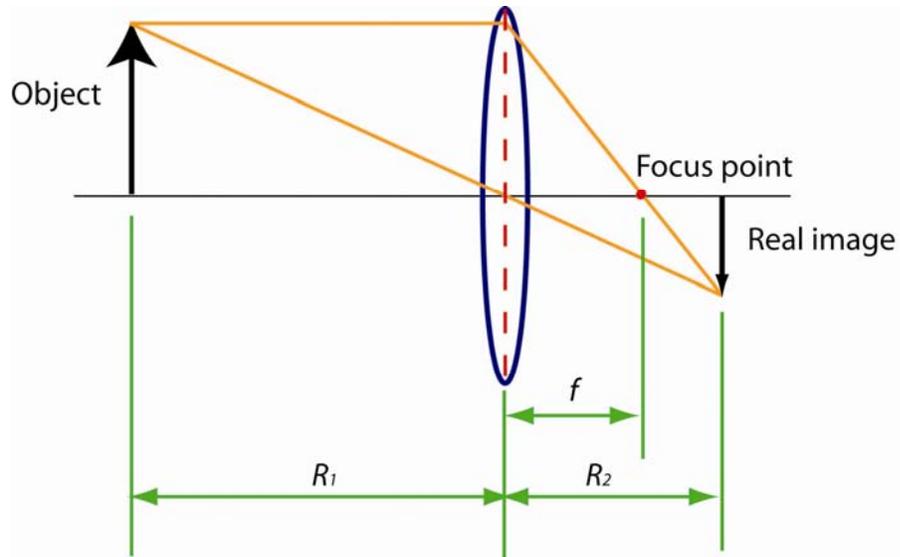
Figure 2.8: Image formation of a thin lens.

Assuming that the lens is a thin lens with negligible thickness, the relationship of these three parameters can be expressed using thin lens formula:

$$\frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{f}. \tag{2.10}$$

For the lens selection of the optical tracking system, two main considerations discussed as follows are focal length and aperture size.

(a) Focus length

The selected CMOS image sensor has a sensing area of 24.6 mm by 24.6 mm. Meanwhile, the required tracking range is 1 m by 1 m. This means that the focal length of the lens should be carefully selected so that the selected CMOS image sensor can fully cover the tracking range. The demonstration for this consideration is shown in Figure 2.9.
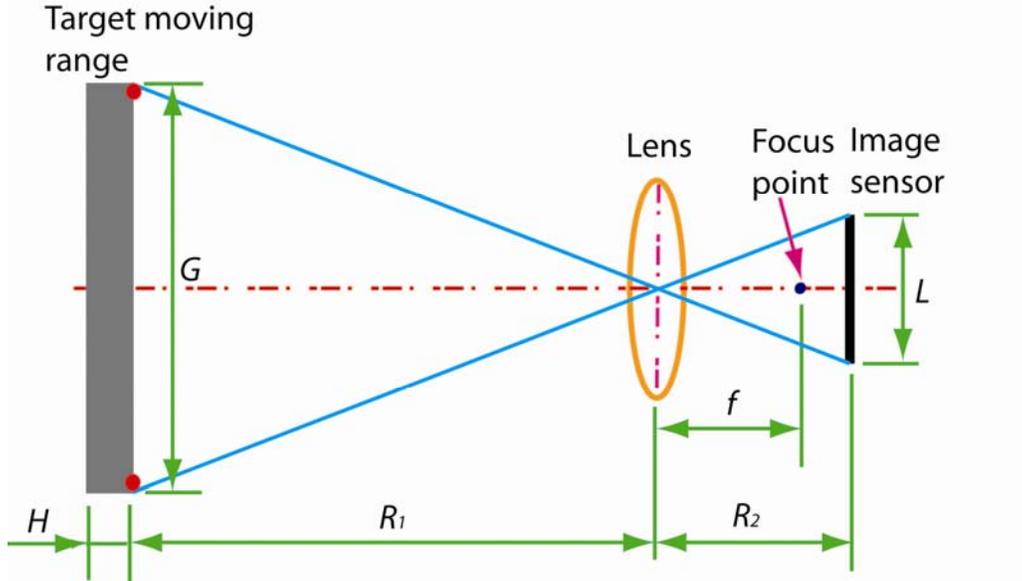
Figure 2.9: Image sensor fully covers target moving range.

Here, R_1, R_2 and f have the same meaning as in Figure 2.8. L is the length size of the CMOS image sensor, G is the target moving range along X axis, and H is the target moving range along Z axis. So the triangle relation for L and G is

$$\frac{L}{R_2} = \frac{G}{R_1}.$$
(2.11)

By combining Eqn. (2.10) and Eqn. (2.11), the focal length f can be calculated as

$$f = \frac{R_1 L}{G + L}.$$
(2.12)

As the image sensor length L is known to be 24.6 mm, the target moving range G is 1 m, and the target to the lens distance R_1 is about 1 m, the focal length f is calculated to be 25 mm.

(b) Aperture size

Aperture size is another important consideration for the lens. The size of the lens aperture determines how much light passes through the lens and reaches the image sensor. It also affects

the depth of field for the tracking system, as shown in Figure 2.10. d_A represents the aperture diameter of the lens. It is assumed that the tracking target is a point light source.

(a)

(b)
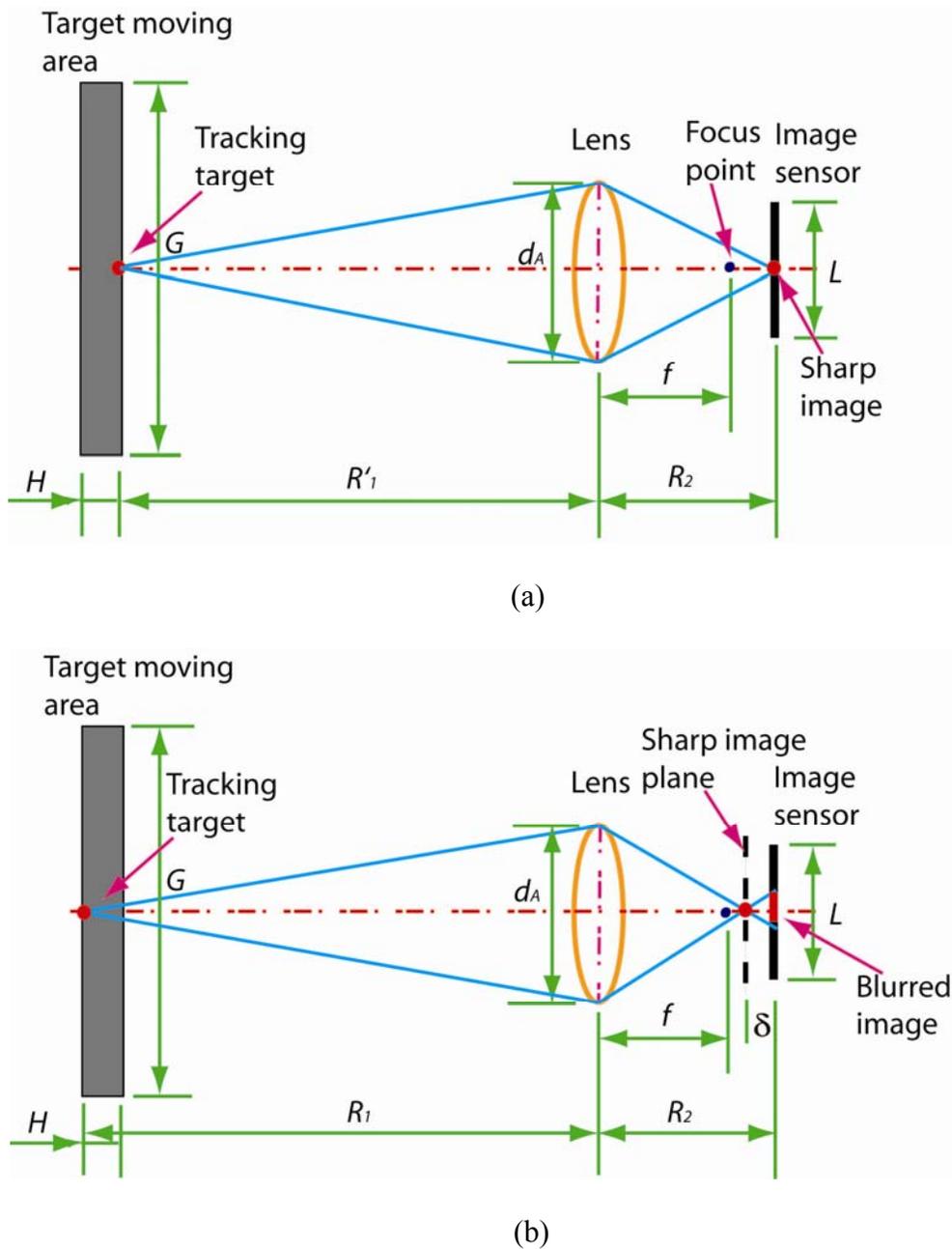
Figure 2.10: (a) Image formation for the tracking target on the near end with sensor fixed on the sharp image plane; (b) Image formation on the far end obtaining blurred image.

Figure 2.10 (a) shows the image formation of the tracking target on the image sensor. When the image sensor is placed on the sharp image plane for the tracking target, the image of the tracking target is a sharp dot on the image sensor. However, precise focus is only possible at one distance. Figure 2.10 (b) shows the defocused case of the image formation process. When the tracking target is moving to the far end along Z axis, the distance between the target and the lens R_1 is varying as well. The sharp image plane is also shifting along Z axis accordingly. For a camera with fixed focus lens, the image sensor is fixed at a certain distance from the lens (within the range of $f \sim 2f$), the target image on the image sensor will be a relatively blurred spot image, whose shape is affected by the lens aperture. When this circular spot is sufficiently small, it is indistinguishable from a point, and appears to be in focus; it is rendered as 'acceptably sharp'. The diameter of the circle increases with distance from the point of focus; the largest circle that is indistinguishable from a point is known as the 'circle of confusion' [24]. From the triangulation relation Figure 2.10 in the distance from the sharp image plane to the image sensor δ can be expressed as

$$\delta = \frac{R_1{}' \cdot L}{G} - \frac{(R_1{}' + H) \cdot \dfrac{R_1{}' \cdot L}{G+L}}{(R_1{}' + H) - \dfrac{R_1{}' \cdot L}{G+L}}, \tag{2.13}$$

where $R_1{}'$ is the closest distance from the tracking target to the lens. By simplifying the above equation, the following is obtained as

$$\delta = \frac{R_1{}' \cdot L}{G} \cdot \frac{1}{1 + \dfrac{(R_1{}' + H)}{H} \cdot \dfrac{G}{L}}. \tag{2.14}$$

The maximum blurred image size M can be calculated as

28

$$M = \frac{\delta}{2} \cdot \frac{d_A}{f} = \frac{R_1' \cdot L}{G} \frac{1}{(1+(\frac{R_1'+H}{H}) \cdot \frac{G}{L})} \cdot \frac{1}{2F} , \qquad (2.15)$$

where F is the F-number of the lens and $F= f/d_A$ (focal length/aperture diameter). Based on the above equations, F can be also represented as

$$F = \frac{\delta}{2M} = \frac{R_1' \cdot L}{G} \cdot \frac{1}{1+\frac{(R_1'+H)}{H} \cdot \frac{G}{L}} \cdot \frac{1}{2M} . \qquad (2.16)$$

For the optical tracking system, the ideal maximum blurred size should be less than 1 pixel with the depth of field more than 0.2 m, so F should be:

$$F > \frac{R_1' \cdot L}{G} \cdot \frac{1}{1+\frac{(R_1'+H)}{H} \cdot \frac{G}{L}} \cdot \frac{1}{2M} = \frac{1 \times 24.6 \times 10^{-3}}{1} \times \frac{1}{1+\frac{(1+0.2)}{0.2} \cdot \frac{1}{24.6 \times 10^{-3}}} \times \frac{1}{2 \times 1 \times 12 \times 10^{-6}} = 4.18 \cdot \qquad (2.17)$$

So the F-number for the lens larger than 4.18 is required to obtain maximum blurred size less than 1 pixel with the depth of field of 0.2 m.

When the parameters of the lens are determined, a selection can be made based on the focal length and aperture size requirements. To be compatible with the selected CMOS image sensor, this lens should be F-mount compatible. Keeping all these considerations in mind, the lens chosen for the tracking system is Distagon T2.8/25 ZF from Carl Zeiss, as shown in Figure 2.11. This lens has fixed focal length of 25.7 mm and its F-number is adjustable from 2.8 to 22, which allows a large depth of field without gaining more image blur.

Figure 2.11: The selected lens (Distagon T2.8/25 ZF from Carl Zeiss) with f=25.7 mm, F-number F=2.8~22.

Notice that an infrared version of this type of lens is selected for good transmittance for infrared (wavelengths from 800 nm to 3000 nm) light while regular lens does not have high transmittance for infrared light, seen in Figure 2.12. The main difference between the infrared lens and the regular lens is that the infrared version has special infrared coatings to pass through infrared light while the regular lens does not. This unique feature of the infrared version allows applying the infrared light as the light source of the 3D tracking system.
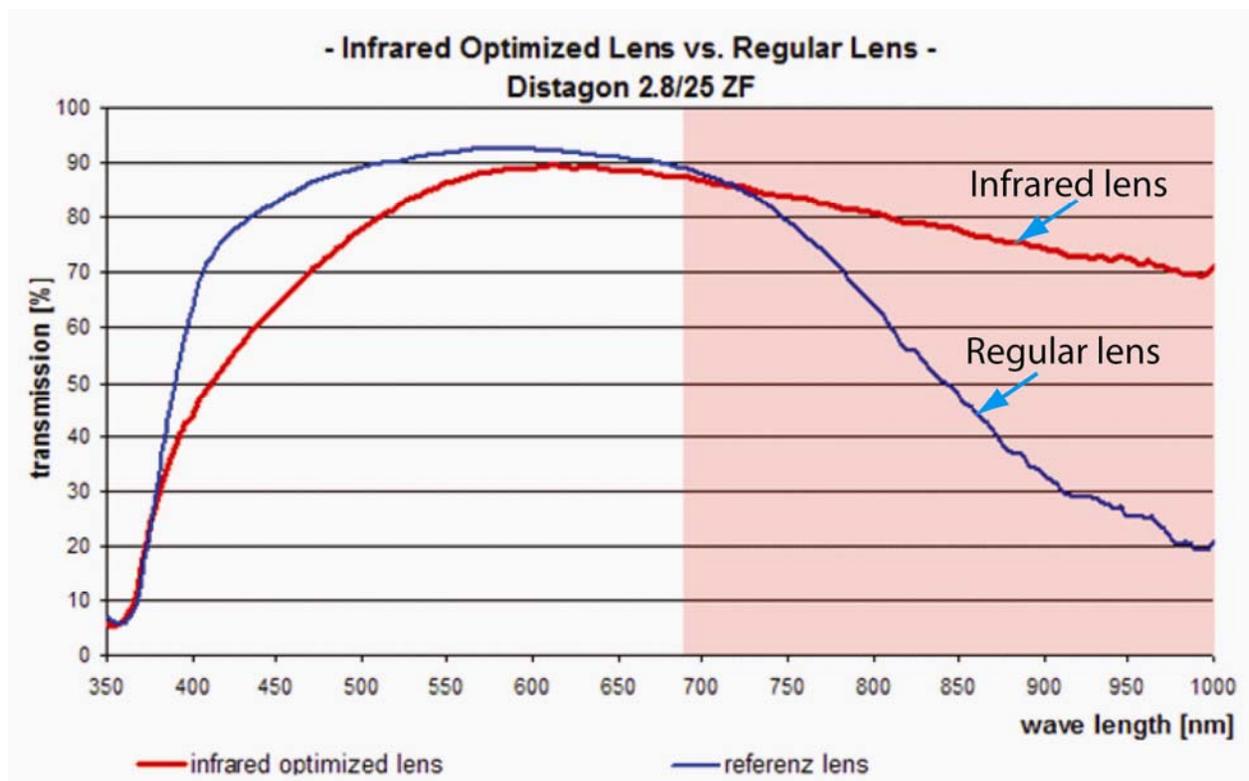
Figure 2.12: Transmittance comparison of two types of lenses:
infrared lens (red) and regular lens (blue) [25].

2.3.3 Tracking target selection

There are two main types of tracking targets for optical tracking systems. One is 'active'
tracking targets with a light source. The other type is 'passive' tracking targets with retro-
reflectors reflecting light from the light source mounted around the camera. Passive tracking
targets require a strong illumination source for it to reflect enough light for the camera to capture,
which makes the system more complicated to handle in terms of energy consumption.
Furthermore, the passive tracking target may change their appearance much when viewed from
different directions, while active tracking targets do not.

Due to the disadvantages of passive tracking targets, it is reasonable to design active tracking targets for the 3D optical tracking system. This subsection discusses the considerations, selection of the active tracking targets, and the safety check for these targets.

1) Considerations

The tracking targets of the 3D optical tracking system should satisfy the requirements as follows:

(a) Large viewing angle

To make sure that the camera can capture the tracking target within the moving range on the target plane, the viewing angle from the light source should be larger than the camera field of view, as shown in Figure 2.13.
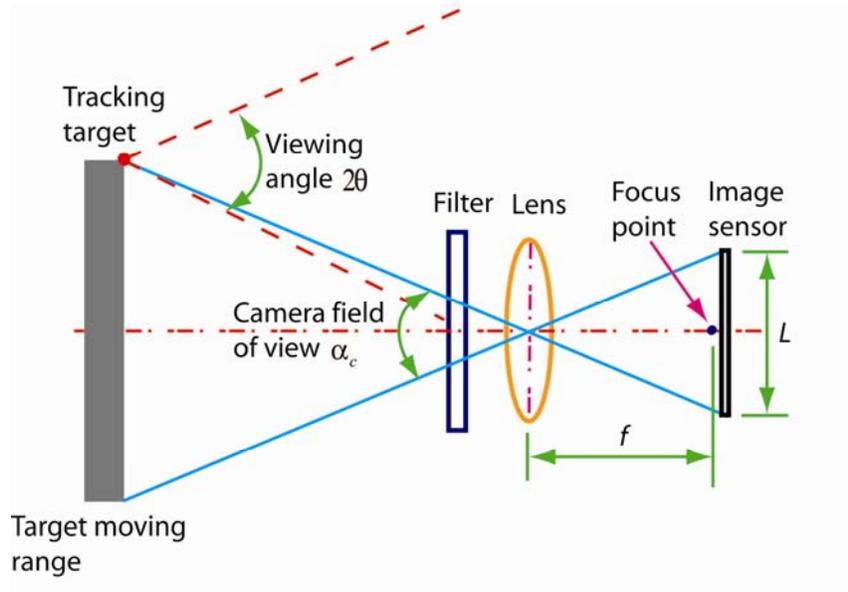
Figure 2.13: The camera viewing the tracking target with viewing angle of 2θ.

The camera field of view α_c can be approximate as:

$$\alpha_c = 2 \times \tan^{-1}(\frac{L}{2f}), \qquad\qquad (2.18)$$

where L is the length of the image sensor, f is the focal length of the lens. Based on the selected image sensor and optical lens, the camera field of view is $53°$. So the viewing angle for the tracking target should be larger than $53°$ in order to capture the tracking target even when the tracking target is on the edge of the moving range.

Laser light source is a widely used light source due to its high intensity with high power output in many applications such as laser interferometers, laser Doppler vibrometers and so on. However, the laser light is not applicable to the 3D optical tracking system because its viewing angle is very small (typically less than $20°$). When the laser light is moving within the tracking range, the image sensor could not observe the laser light in most of the cases.

The light emitting diode (LED) can provide a wide viewing angle (up to $180°$), which satisfies the requirements. So the LEDs are considered as the tracking targets for the 3D tracking system.

(b) Uniform intensity distribution

Another consideration for the tracking target is the uniformity of the intensity distribution. The reason is that the image of the tracking target can remain the same with uniform intensity distribution, even when viewing from different angles. If the intensity distribution is not uniform, the image of the LED will vary from angle to angle. Then the locations of the tracking target will vary as well, which causes more error for the 3D position estimation. In the ideal case, the intensity distribution of the tracking target is uniform with viewing angle of $180°$.

(c) Wavelength range

The third consideration for the tracking target is the wavelength range. In order to reduce the noise from the visible ambient light within the wavelength range of 380 nm to 750 nm [26], infrared LEDs (IR LEDs) with wavelength range of 750 nm to 3000 nm [26] are considered. Meanwhile, long pass filters are applied to block the visible ambient light and pass all the light from the tracking target. In addition, the responsivity of CMOS image sensor is relatively low for the wavelength above 950 nm (less than 25% of the peak responsivity and quantum efficiency is about 5%) as shown in Figure 2.14. So the effective wavelength range for the LEDs is 800 nm to 900 nm.
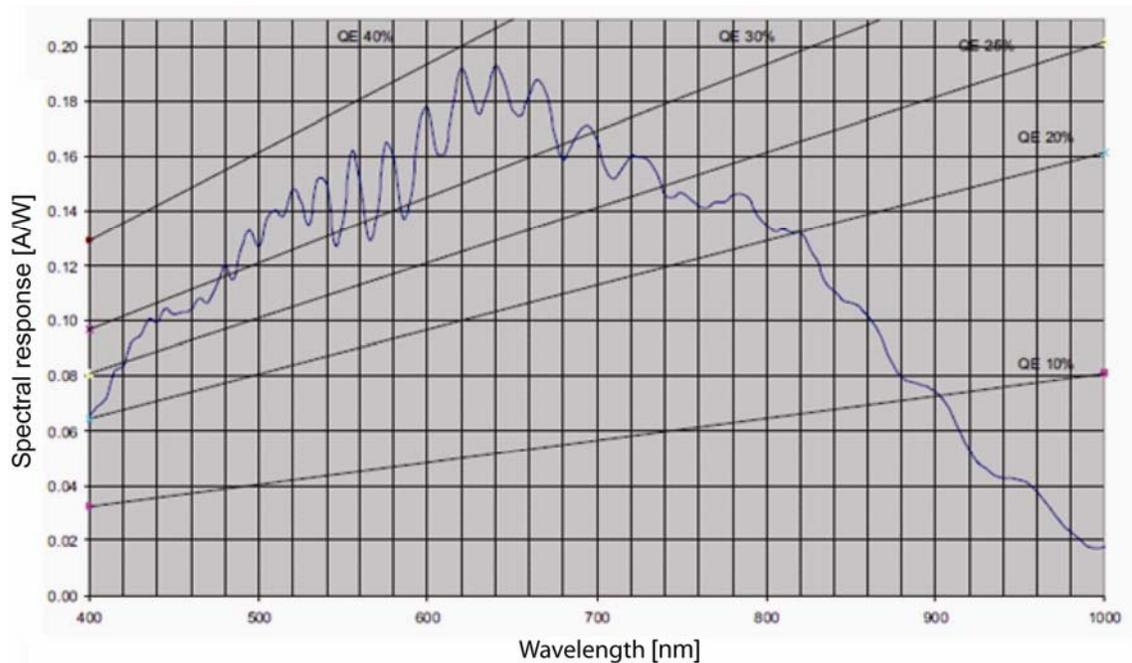
Figure 2.14: Responsivity and QE (quantum efficiency) chart for the selected CMOS sensor [27].

(d) High power efficiency

The fourth consideration for LEDs is the high optical power output with high power efficiency. High optical power output allows the tracking target to have much higher intensity

34

than the ambient light and easy to filter the ambient noise on the image. High power efficiency is very useful once batteries are used to power the tracking targets and can save energy with more tracking targets.

To summarize the considerations discussed above, the tracking target should satisfy: 1) viewing angle the larger, the better (> 53° at least); 2) uniform intensity distribution; 3) wavelength range within 800 nm~900 nm; 4) high optical power output with high power efficiency.

2). Selection

Since the uniform intensity distribution of the IR LEDs with viewing angle of ±90° is not practical in reality, an IR LED HE8812SG from Opnext whose light intensity is uniformly distributed within the angle of ±40° best meets the requirements. The viewing angle for this LED is 100°. The shape of the LED and its intensity distribution is shown in Figure 2.15. The wavelength of the HE8812SG is 870 nm with minimum power output of 40 mW.

(a)   (b)
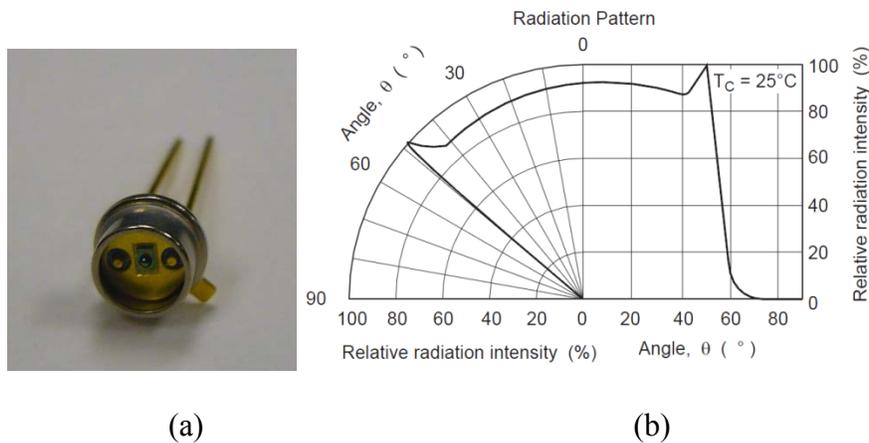
Figure 2.15: (a) Image for the HE8812SG; (b) Angular intensity distribution for the HE8812SG [28].

To verify that the HE8812SG has the best performance for the system, two other IR LEDs are selected for comparison. One is the SE3470 from Honeywell with 880 nm wavelength, as shown in Figure 2.16. This IR LED has relatively uniform intensity distribution within ±20˚, which is less than the camera field of view. The other one is the LN152 from Panasonic with wavelength of 950 nm, and optical power output of 10 mW, as shown in Figure 2.17. The LN152 has very wide viewing angle of ±80˚ with uniform intensity distribution within ±40˚ but the optical power output is relative low. Besides, the image sensor is not very sensitive to wavelength of 950 nm. The performance comparison for these three IR LEDs is presented in Chapter 5.
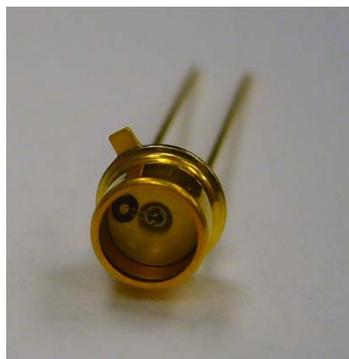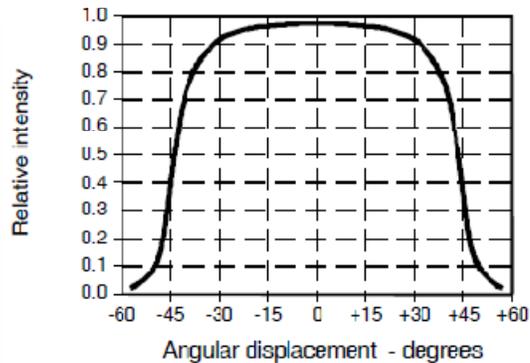
(a)   (b)

Figure 2.16: (a) Image for the SE3470; (b) Angular intensity distribution for the SE3470 [29].
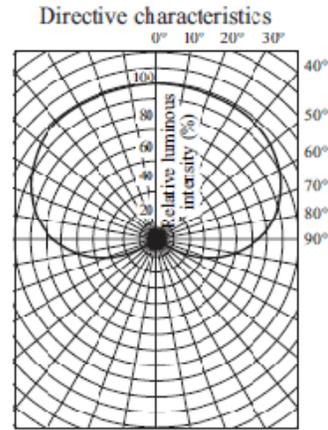
(a)   (b)

Figure 2.17: (a) Image for the LN152; (b) Angular intensity distribution for the LN152 [30].

3). Safety consideration

It is well known that strong visible light, such as laser light or sun, can be harmful to human eyes. In fact, the human eye has an exposure limit before it becomes damaged. The harm from the strong visible light is well known. However, human eyes are not sensitive to infrared light at all. This does not mean that strong infrared light cannot damage the human eye. On the contrary, it can actually burn the retina of human eyes when it is too strong. To keep the experimental environment safe, care must be paid with the selected IR LEDs. Questions have arisen as to how strong the chosen IR LEDs are and whether it is necessary to apply exposure limits to them.

With the development of high power LEDs including visible LEDs and IR LEDs, there has been much effort in developing LEDs safety standards [31] [32]. International Commission on Non-Ionizing Radiation Protection (ICNIRP) put forth guidelines for LEDs safety which has become the worldwide standard. In their statement, they listed 5 principal types of hazards to the eyes from the light sources [31]:

37

(a). Ultra-Violet photochemical injury to the cornea (photo-keratitis) and lens (cataract) of the eye (180 nm to 400 nm).

(b). Blue-light photochemical injury to the retina of the eye (principally light wavelength is from 400 nm to 550 nm, unless aphakic, 310 nm to 550 nm).

(c). Retinal thermal injury of the eye (400 nm to 1400 nm).

(d). Near IR thermal hazards to the lens (approximately 800 nm to 3000 nm).

(e). Thermal injury or burns of the cornea of the eye (approximately 1400 nm to 1mm).

For the selected IR LEDs (800 nm to 1000 nm), only aspect (c) and (d) are remotely relevant, since aspects (a) and (b) can only occur from short wavelength light and UV, and thermal injury occurs with longer wavelength larger than 1000 nm. Therefore, only the relevant potential hazards need to be evaluated. Figure 2.18 shows the image formation on human eyes.
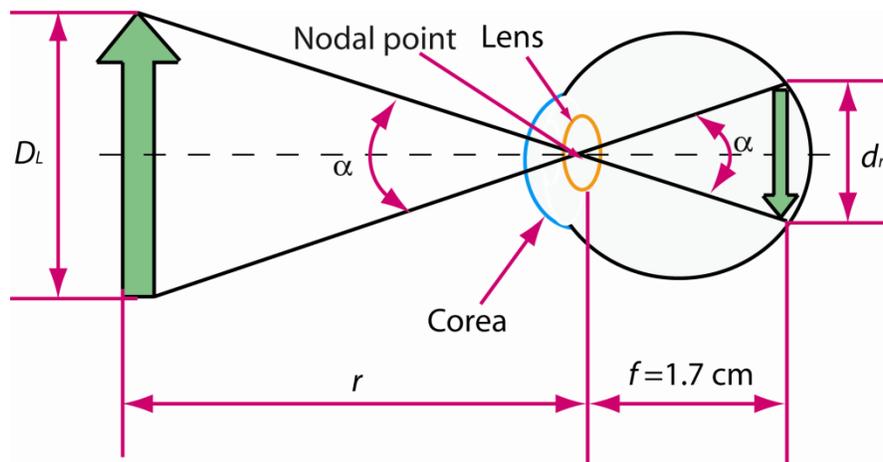
Figure 2.18: Image formation on human eyes [32]: r is the distance between the object and lens of the eye, $f = 1.7$ cm is the focus length of the eye, d_r is the image size of the object, D_L is the object size, α is the viewing angle of the eye.

Here, the irradiance on the retina from IR LEDs can be calculated based on the model analysis for the tracking systems in Section 2.2. The retinal irradiance E_e with unit of W/cm^2 can be expressed as

$$E_e = \frac{\pi L_s \cdot \tau \cdot d_e^2}{4f^2}, \tag{2.19}$$

where L_s with unit of W/(cm^2 sr) is the radiance of the IR LED, d_e is the pupil size of human eyes with unit of cm, f is the effective focal length of eyes with unit of cm, and τ (unit-less, $\tau \approx 0.9$) is the transmittance of the ocular media of human eyes. For adults, the effective focal length of the eyes is about 1.7 cm. Then the Eqn. (2.19) can be simplified to be

$$E_e = 0.27 \cdot L_s \cdot \tau \cdot d_e^2. \tag{2.20}$$

Based on this equation, the irradiance on the retina from the selected IR LEDs can be estimated as follows:

a) HE8812SG from Opnext

The radiance with unit of mW/(cm^{2}sr) for HE8812SG with maximum forward current I_F =200 mA is expressed as:

$$L_{s1} = \frac{P_o}{A_s \cdot \Omega} = \frac{40}{\frac{\pi}{4} \times 0.25^2 \times 2\pi(1 - \cos 50^\circ)} = 360 \, [\text{mW/(cm}^2\text{sr)}], \tag{2.21}$$

where P_o is the output optical power of this IR LED in Watt, A_s is the effective light emitting area in cm^2 and Ω is the solid angle of IR LED emission with unit of sr. The retinal irradiance from this IR LED can be calculated as:

$$E_{e1} = 0.27 \cdot L_{s1} \cdot \tau \cdot d_e^2 = 0.27 \times 360 \times 0.9 \times (0.3)^2 = 8 \, [\text{mW/cm}^2], \tag{2.22}$$

where the pupil diameter d_e is approximate to be 0.3 cm (average pupil diameter for human eyes is 0.2 ~ 0.4 cm) and the transmittance of the ocular media $\tau = 0.9$.

b) SE3470 from Honeywell

The radiance for SE3470 with maximum forward current $I_F=100$ mA can be calculated as:

$$L_{s2} = \frac{P_o}{A_s \cdot \Omega} = \frac{10}{\frac{\pi}{4} \times 0.25^2 \times 2\pi(1-\cos 45°)} = 107 \text{ [mW/(cm}^2\text{sr)]},$$
(2.23)

Similar to the above IR LED, the retinal irradiance from SE3470 can be obtained as:

$$E_{e2} = 0.27 \cdot L_{s2} \cdot \tau \cdot d_e^2 = 0.27 \times 107 \times 0.9 \times (0.3)^2 = 2.34 \text{ [m W/cm}^2\text{]}.$$
(2.24)

c) LN152 from Panasonic

The radiance for LN152 with maximum forward current $I_F=100$ mA can be calculated as:

$$L_{s3} = \frac{P_o}{A_s \cdot \Omega} = \frac{10}{\frac{\pi}{4} \times 0.25^2 \times 2\pi(1-\cos 90°)} = 31.4 \text{ [mW/(cm}^2\text{sr)]}.$$
(2.25)

Then the retinal irradiance from LN152 is:

$$E_{e3} = 0.27 \cdot L_{s2} \cdot \tau \cdot d_e^2 = 0.27 \times 31.4 \times 0.9 \times (0.3)^2 = 0.69 \text{ [m W/cm}^2\text{]}.$$
(2.26)

According to ICNIRP's safety standard [32], the retinal irradiance without infrared cornea-lens thermal hazard should be less than 100 mW/cm^2 when directly viewing the IR LEDs for more than 10 minutes. The safety check for all the IR LEDs is shown in Table 2.1. It is concluded that all the selected IR LEDs are safe to use.

Table 2.1: Safety check for selected IR LEDs.

Parameters	HE8812SG (Opnext)	SE3470 (Honeywell)	LN152 (Panasonic)
Radiance from IR LED L_s (mW/(cm^{2}sr))	360	107	31.4
Retinal irradiance E_e (m W/cm^2)	8	2.34	0.69

2.3.4 Optical filter selection

After the selection for the tracking targets, it is clear that the wavelengths of the light source should be within the range of 850 nm to 1000 nm. Therefore, the filter should only pass the wavelengths of the light source and block all the visible ambient light. The ideal case for the filter is that the transmittance for the desired wavelengths of light is 1 and the transmittance for the undesired wavelengths of light is 0. However, this is not realizable. Real optical filters have transit period from high transmittance for desired light to low transmittance for undesired light.

For the tracking system, the desired wavelengths range should cover the light spectrum of selected IR LEDs. Figure 2.19 shows the spectral distribution of HE8812SG. From this figure, it indicates that with the peak intensity at wavelength 870 nm, HE8812SG contains a range of wavelengths starting from 820nm and extending to more than 920 nm. So the designed optical filter should pass as much wavelengths of this IR LED as possible to ensure over 90% of the light from the tracking target can pass.
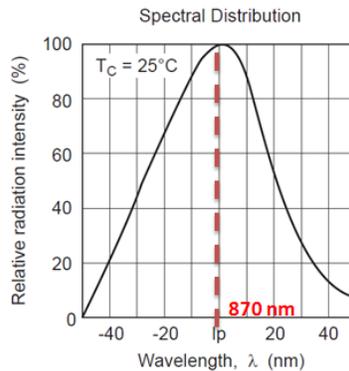
Figure 2.19: Spectral distribution for the HE8812SG [28].

Compared all the commercial optical filters, the best fit for the tracking system is LP830 from MidOpt. The LP830 has over 96% transmittance at wavelength $\geq$ 870 nm, and less than 1% transmittance at wavelength $\leq$ 740 nm. The transmission performance of LP830 is shown in Figure 2.20. To ensure the compatibility with the selected optical lens, the thread diameter selected is 58mm and the pitch for the filter is 0.75 mm.
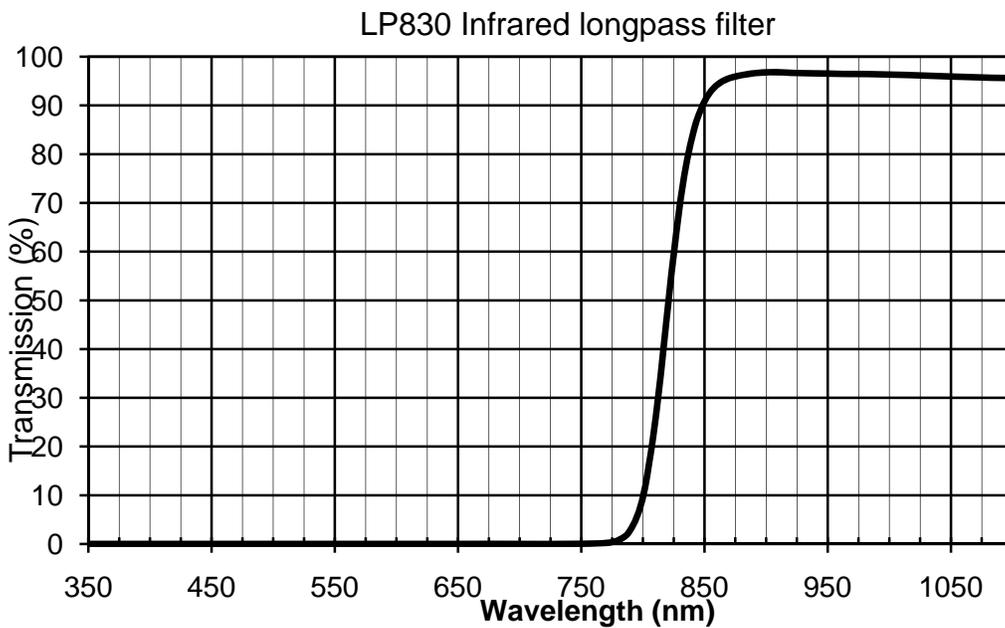
Figure 2.20: Transmission performance of the LP830.

2.4   Ambient light effect

After all the hardware components are selected, the next step for the 3DOF optical tracking system design is to analyze the ambient light effect and check the feasibility of the designed system. As all image processing algorithms and machine vision algorithms are based on the image data of the tracking target, ambient light effect analysis, which affects the extraction of

tracking target image data, is very important for the system. It will determine whether all the hardware components selection for the system is feasible.

The ambient light effect analysis is to compare the output voltages from the selected IR LED and from the ambient light. The ambient light is assumed to be fluorescent lamps as they are widely used for illumination in the labs and offices. Typically, the ambient noise reaching the CMOS cameras is reflective light from surfaces and walls. Here, the worst case in which the camera directly viewing the fluorescent lamp is considered. So the distance from the selected IR LED to the optical lens is the same as the distance from the fluorescent lamps to the optical lens, as shown in Figure 2.21. As the voltage output from the ambient light in the other cases can not be as high as in the worst case, the hardware components selection can be proved to be feasible if the voltage output ratio of the IR LED to ambient light for the worst case is still good (>10).
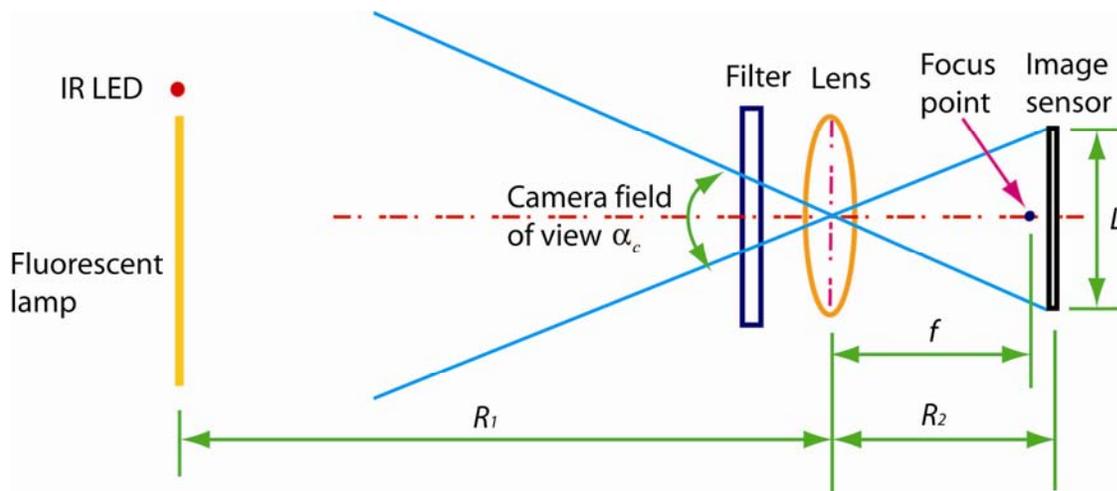
Figure 2.21: Camera directly viewing IR LED and fluorescent light: R_1 is the distance from the light sources to the lens, R_2 is the distance from the lens to the image sensor, f is the focus length of the lens and L is the length of the image sensor.

The fluorescent lamps in the lab are F32T8/SPx35/ECO from General Electrical (GE). This type of lamps has 32 Watt power output and size of 1.2 m (length) x 25 mm (diameter). It

emits white visible light and contains a wide range of wavelengths from 300 nm to 1000 nm. To simplify the analysis, the light spectrum distribution for this type of fluorescent lamps is assumed to be uniform and its relative radiant intensity distribution is shown in Figure 2.22.
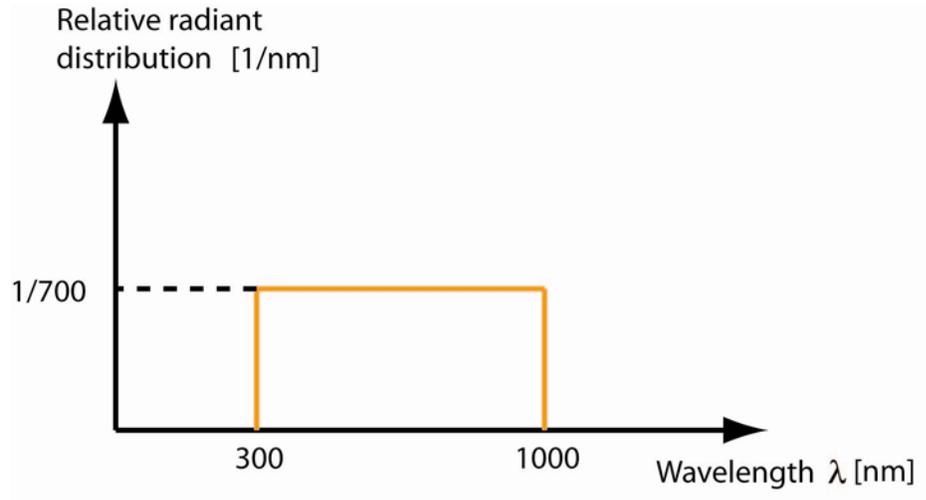
Figure 2.22: Relative radiant intensity distribution for the fluorescent lamp.

As the optical power output efficiency of the ambient light is about 40%, the total radiance from the fluorescent lamp E with unit of $W/(m^2 sr)$ can be obtained as:

$$E = \frac{P_0}{\Omega} = \frac{32 \times 0.4}{2\pi} = 2.04 \,[\text{W/sr}], \qquad (2.27)$$

where $\Omega = 2\pi$ is the solid angle for the ambient light in sr. So the radiance from the ambient light with the relative intensity distribution shown in Figure 2.22 can be calculated as:

$$L_s(\lambda) = \int_{\lambda_1 = 300}^{\lambda_2 = 1000} \frac{E}{700} d\lambda. \qquad (2.28)$$

Based on the model analysis in the previous section and responsivity data shown in Figure 2.14, we can calculate the voltage output for the fluorescent lamp using the following expression:

44

$$V = \frac{1}{R_1^2} \cdot A_0 \cdot T_l(\lambda) \cdot t_{exp} \cdot \frac{G_c}{q} \cdot \int_{\lambda_1}^{\lambda_2} L_s(\lambda) \cdot T_f(\lambda) \cdot R_{sa}(\lambda) d\lambda \,, \qquad (2.29)$$

where the distance from light source to lens R_l=1 m, aperture size $A_0 = \frac{\pi}{4} d_A^2 = 4.3$ mm^2

(aperture diameter is $d_A = f/F = 2.34$ mm with F-number F=11, f=25.7 mm), G_c is 13.5 µV/e$^-$ for

the selected CMOS image sensor, q is 1.6×10^{-19} C and the exposure time t_{exp} is assumed to be 80

µs . The voltage output of the IR LED HE8812SG can also be calculated using Eqn. (2.29) with

radiance L_s of 0.01 W/sr when forward current I is 100 mA. However, the transmittance of the

long pass filter $T_f(\lambda)$ and responsivity $R_{sa}(\lambda)$ are quite different for the IR LED and the ambient

light, which ensures higher voltage output from the IR LED. The voltage output calculation

results for both IR LED and ambient light are shown in Table 2.2.

Table 2.2: Voltage outputs for the IR LED and the ambient light.

Light source	t_{exp} (µs)	Aperture Size d_A (mm)	Radiance L_s (W/sr)	Pixel Voltage Output V (V)
Ambient light (GE F32T8)	80	2.34	2.04	0.0053
IR LED (HE8812SG)	80	2.34	0.010	0.8797
V(IR LED)/V(Ambient light) = 156:1				

From the calculated results above, it is concluded that the hardware selection is effective

and feasible for the 3D optical tracking system. The noise from the ambient light is very low

compared to the signal from the selected IR LED, even in the worst case.

2.5   Summary

This chapter presented the system hardware selection and analysis, including model analysis for the signal transmission, system hardware components selection and ambient light effect analysis. An overview for the system hardware was given in Section 2.1. In Section 2.2, details of the model analysis of the 3D optical tracking system were explained. The formula for the light source intensity calculation and voltage output calculation were derived. Section 2.3 discussed all the hardware components selection, involving image sensor selection, lens selection, tracking target selection and safety check for the selected IR LEDs, and filter selection. In Section 2.4, the ambient light effect was analyzed for the system. The results show that the noise from the ambient light is very low compared to the effective signal from the tracking target.

Chapter 3

Noise Correction and Target Location Algorithms

To achieve a high accuracy for the 3D optical tracking system, the tracking algorithms are very important. The tracking algorithms include noise correction, target location, and 3D reconstruction, as shown in Figure 3.1. Noise correction algorithms are implemented to correct the raw images captured from left and right cameras. Based on these corrected images, target location algorithms are applied to obtain accurate sub-pixel locations for the tracking target. Once the camera parameters are known from camera calibration algorithms, 3D reconstruction algorithms can be employed to estimate the 3D position of the tracking target using target locations from the left and right cameras.

In this chapter, only noise correction and target location algorithms for the 3D tracking are discussed. Camera calibration and 3D reconstruction algorithms will be presented in Chapter 4. Section 3.1 introduces the noise correction algorithms for processing raw images. Section 3.2 describes the target location algorithms and simulation investigation on these algorithms. This chapter is closed with a summary in Section 3.3.
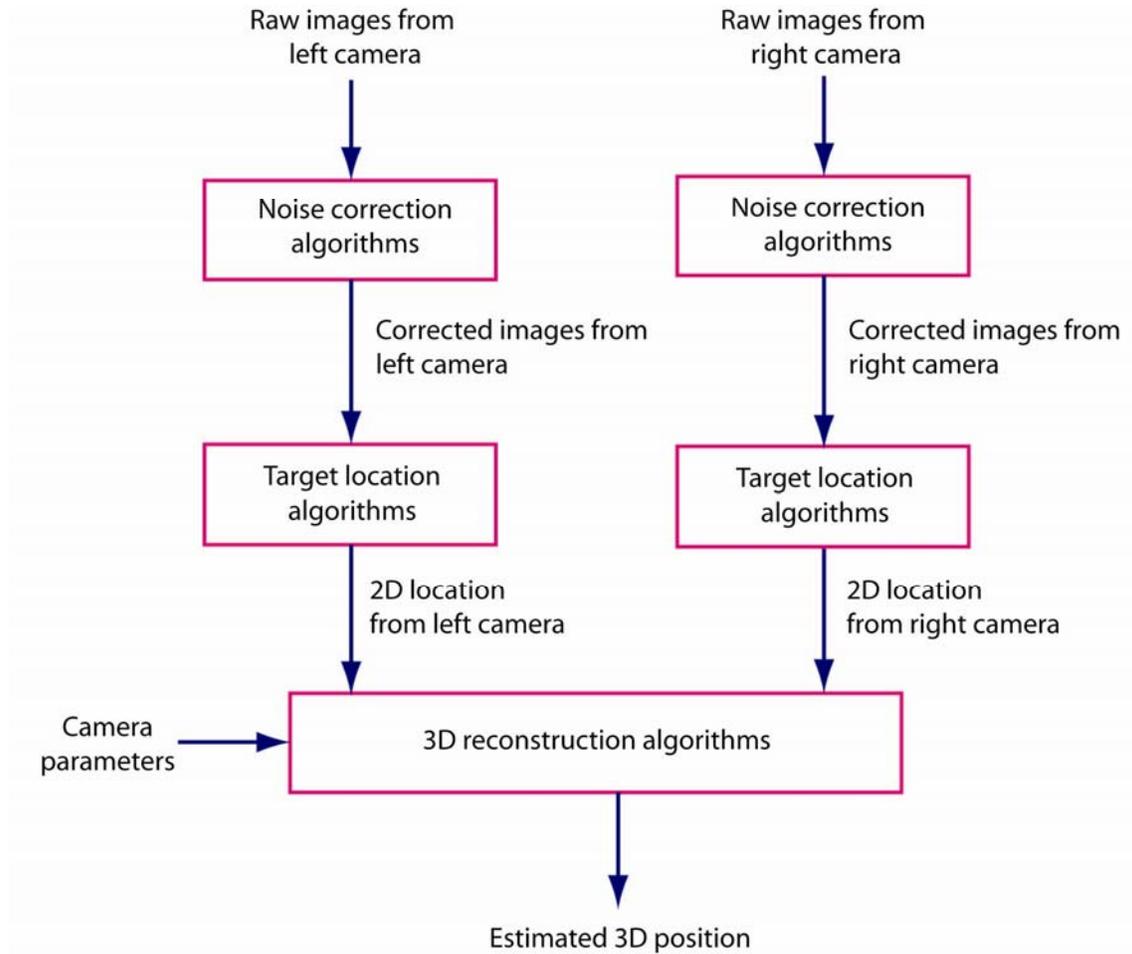
Figure 3.1: Flow chart of algorithms for the 3D optical tracking.

3.1 Noise correction algorithms

The actual images of the tracking target obtained are always the superposition of the

tracking target and the background noise. The background noise is mainly caused by ambient

illumination within the scene [33], ADC quantization errors, fixed pattern noise (FPN) due to

dark current, and pixel response non uniformity (PRNU) in CMOS image sensors.

For the 3D optical tracking system, the exposure time of the image sensor is required to

be less than 80 μs for in order to achieve the targeted frame rate (>10 KHz). Based on the

48

previous ambient light effect analysis in Chapter 2, it is concluded that the noise from the ambient light is very low and can be neglected compared to the signal from the tracking target with this exposure time. So the main noise sources for the images here are ADC quantization error, FPN and PRNU from CMOS image sensors. However, as the ADC quantization error is fixed and is negligible for the selected CMOS sensor with 12 bits ADC, the noise correction algorithms for the FPN and the PRNU are mainly discussed here.

3.1.1 FPN correction

The presence of the FPN, caused by the processing and the pixel design of the CMOS sensor, results in all the pixels with small offset differences relative to one another [34]. This can be corrected by subtracting a dark image, which is an image with no light reaching the image sensor by covering the lens [34]. This type of noise can be corrected using the following equation:

$$\boldsymbol{I}_c = \boldsymbol{I}_0 - \overline{\boldsymbol{I}}_{blk},$$
(3.1)

where $\boldsymbol{I}_0$ is the raw image data for the tracking target, $\overline{\boldsymbol{I}}_{blk}$ is the averaged image of 10 black images, and $\boldsymbol{I}_c$ is the image data after the FPN correction. The correction process is shown in Figure 3.2.
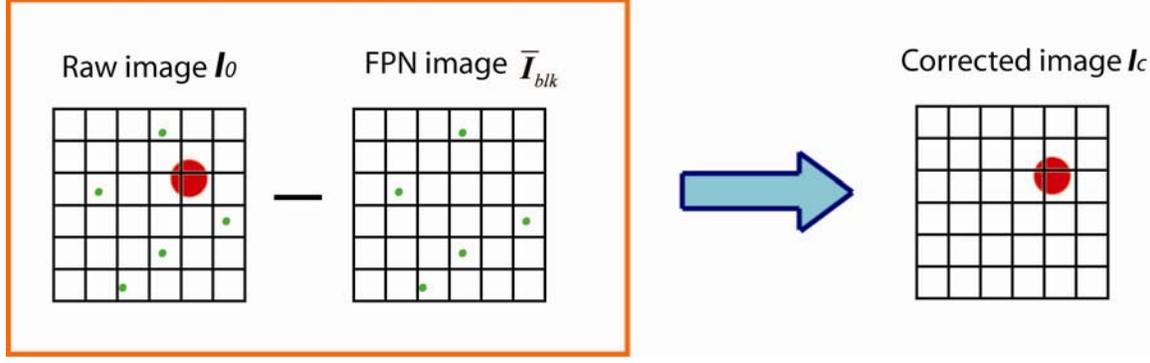
Figure 3.2: Illustration for the FPN correction.

The above FPN correction algorithm can fully eliminate FPN effect on the target images. Notice that both the raw image data for the tracking target and averaged black image data should be obtained under the same exposure time and at the same temperature of the image sensor, as the FPN varies with the exposure time and the sensor temperature.

3.1.2   PRNU correction

In this subsection, the correction of the PRNU noise, which is caused by the difference in the gain from pixel to pixel, will be dealt with. The correction algorithm using a grey image within 70% ~ 100% of the saturation intensity based on uniform illumination can be expressed as:

$$\boldsymbol{I}_{gc} = m_{Igray} \cdot (\boldsymbol{I}_0 - \overline{\boldsymbol{I}}_{blk}) . / (\overline{\boldsymbol{I}}_{gray} - \overline{\boldsymbol{I}}_{blk}) , \tag{3.2}$$

where $\boldsymbol{I}_{gc}$ is the image data after the PRNU correction, $\overline{\boldsymbol{I}}_{gray}$ is the averaged image of 10 gray images with intensity between 70% and 100% of saturation intensity, $m_{Igray} = \dfrac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{\boldsymbol{I}}_{gray}(i,j)$ is the mean intensity of the averaged gray image. The correction for the PRNU can increase gain uniformity from pixel to pixel and improve the accuracy for the target location estimation.

50

3.2   Target location algorithms

After the noise correction for both FPN and PRNU, a small region of calculation (ROC) can be selected by first finding the location of the peak intensity of the tracking target and then selecting the pixels around the peak for the target location estimation. As the selected IR LED covers a image area of 5x5 to 10x10 pixels on the image sensor, the ROC of 50x50 pixels (± 25 pixels around the peak in both X and Y direction) is selected to fully cover the target image.

There are quite a few target location algorithms for estimating the positions of the tracking target on a image, such as two dimensional (2D) Gaussian fitting method, grayscale centroid method, squared grayscale centroid method, binary centroid method, ellipse fitting and average perimeter fitting methods [33]. However, the estimation accuracy of binary centroid method, ellipse fitting and average perimeter fitting methods is relatively low compared to 2D Gaussian fitting, grayscale centroid and squared grayscale centroid. For high accuracy position estimation applications with noisy data, 2D Gaussian Fitting, grayscale centroid and squared grayscale centroid algorithms are more reliable and widely used for target location estimation. Here, these three algorithms are briefly discussed.

3.2.1   2D Gaussian fitting method

Gaussian functions are widely used in statistics and image processing where 2D Gaussians are used for intensity curve fitting. As shown in Figure 3.3, 2D Gaussian distribution is an extension from 1D Gaussian distribution, where 2D Gaussian is for both x and y axis.
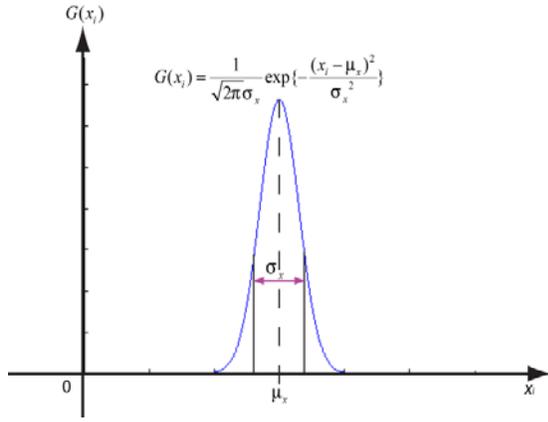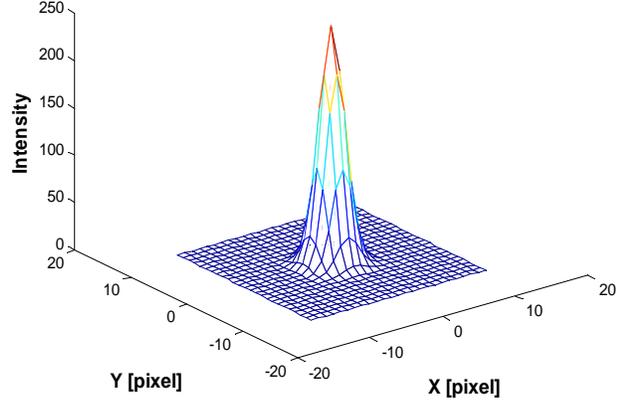
Figure 3.3: (a) 1D Gaussian distribution; (b) 2D Gaussian distribution.

The formula for 2D Gaussian distribution can be given by [33]:

$$G(x_i, y_i) = \frac{K}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_i-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x_i-\mu_x)}{\sigma_x}\frac{(y_i-\mu_y)}{\sigma_y} + \frac{(y_i-\mu_y)^2}{\sigma_y^2}\right]\right\}, \quad (3.3)$$

where x_i and y_i are the pixel coordinates of the tracking target on x and y axis, K is the scaling parameter for peak intensity output, ρ is the correlation coefficient for x and y position of tracking target, μ_x and μ_y are center location of the tracking target along x axis and y axis respectively, σ_x and σ_y are standard deviations from the center location.

Typically, the intensity distribution of a point light source can be well approximated by a 2D Gaussian function. This is also the main reason why 2D Gaussian fitting is widely used for high accuracy particle tracking [35]. For 2D Gaussian fitting, nonlinear least square estimator is applied to estimate the target location.

52

3.2.2 Grayscale centroid method

Grayscale centroid method is also one of the most widely used target location algorithms. This method combines simplicity and high accuracy at the same time. The expression for grayscale centroid method can be given by [36]:

$$x_c = \frac{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} j I_{ij}}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} I_{ij}}, \quad y_c = \frac{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} i I_{ij}}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} I_{ij}}, \tag{3.4}$$

where x_c and y_c are the location of the tracking target in the digitized pixel coordinate systems using ROC for calculation. As shown in Figure 3.4, I_{ij} is the intensity value at i, jth pixel location, m is the total column number for ROC and n is the total row number for ROC.
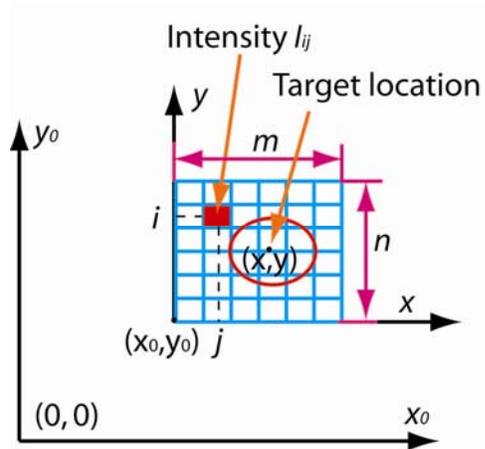
Figure 3.4: Demonstration of the grayscale centroid method.

3.2.3   Squared grayscale centroid method

The expression for squared grayscale centroid method is almost the same as Eqn. (3.4), except that the intensity value are squared. So the equation for squared grayscale centroid method is given by:

$$x_c = \frac{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} jI_{ij}^{2}}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} I_{ij}^{2}}, \quad y_c = \frac{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} iI_{ij}^{2}}{\sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m} I_{ij}^{2}}.$$

(3.5)

The squared gray scale centroid method emphasizes the main body of the tracking target where the intensity values are highest [37].

3.2.4   Comparison for target location algorithms

In order to compare the performance of the above algorithms and see which algorithm provides the best accuracy for target locations, simulation tests are done based on synthetic tracking target with an ideal Gaussian shape ($\sigma_x = 2$, $\sigma_y = 2$, $\mu_x = 0$, $\mu_y = 0$, $\rho = 0$ and peak intensity of 4095). The location of the target is moved by known sub-pixel amounts and the target is quantized at all the locations. Note that the saturation intensity for all the images is 4095 with 12 bits data output.

Target location algorithms such as 2D Gaussian fitting, grayscale centroid and squared grayscale centroid algorithms are employed to estimate the center location of the synthetic target image, and the calculation error which is the difference between the estimated target location and known target location is analyzed. For better comparison and evaluation for these algorithms, the

54

target is moved to 400 different locations and the root mean square (RMS) error for these locations is computed and plotted under different conditions.

These three algorithms are compared under different conditions, such as varying random noise, varying DC offset, varying fill factor and varying saturation levels. The comparison is shown as follows:

a). Random noise effect on target location accuracy

The synthetic target has sigma of 2 in both x and y axis and peak intensity of 4095 (the saturation intensity is 4095 with 12 bits). The ROC is 50 x 50 pixels with random noise added to the ideal image. The noise peak varies from 0 to 10% of the saturation intensity for 12 bits image data. The target is placed at 400 known positions and the RMS error is calculated for each method using the same target image in each case. The results of the RMS error are shown in Figure 3.5. The results show that grayscale centroid method is very sensitive to the random noise added to the target image and the RMS error increases a lot with noise level increasing. Meanwhile, the RMS error of squared grayscale centroid method is as low as 2D Gaussian fitting when the noise peak is less than 1% of the saturation level and increases dramatically when the noise peak is larger. The tracking accuracy of grayscale centroid is 10 times worse than the tracking accuracy of 2D Gaussian fitting.
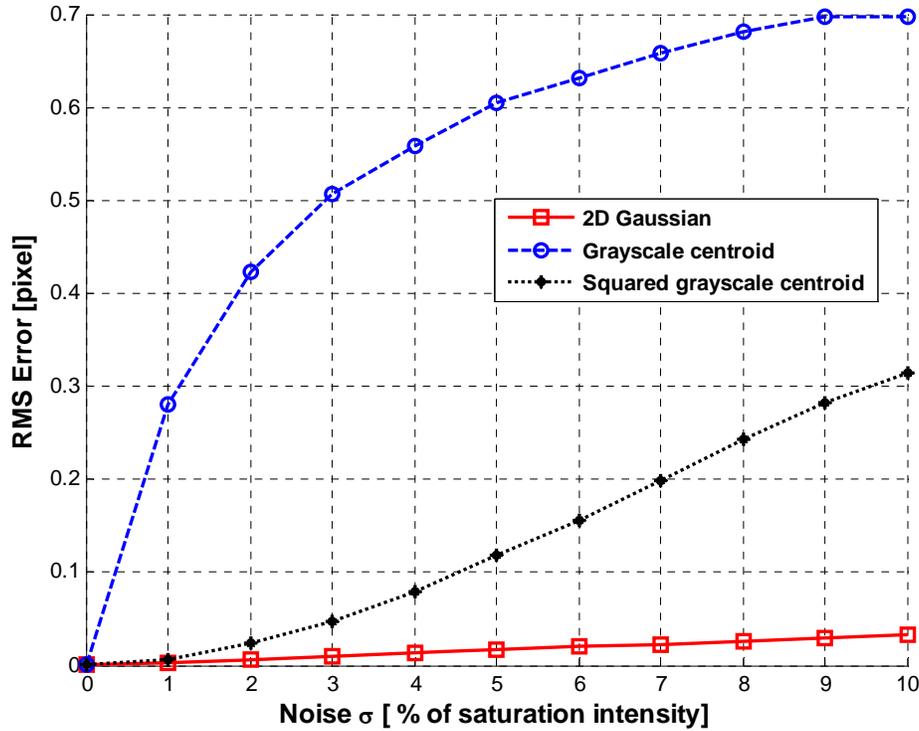
Figure 3.5: Random noise effect on target location accuracy.

b). DC offset effect on target location accuracy

It has been shown by Clarke [33] that a DC offset to the AD converter on a frame grabber will adversely affect the accuracy of target location. In this simulation test, the DC offset effect on the tracking accuracy for three target location algorithms is investigated. The simulated target image is ideal Gaussian shaped with sigma of 2 in x and y axis and peak intensity of 4095 with 12 bits data output. The ROC is 50 x 50 pixels and the RMS error is obtained by moving the target to 400 different locations. The results for tracking accuracy with three methods are shown in Figure 3.6.
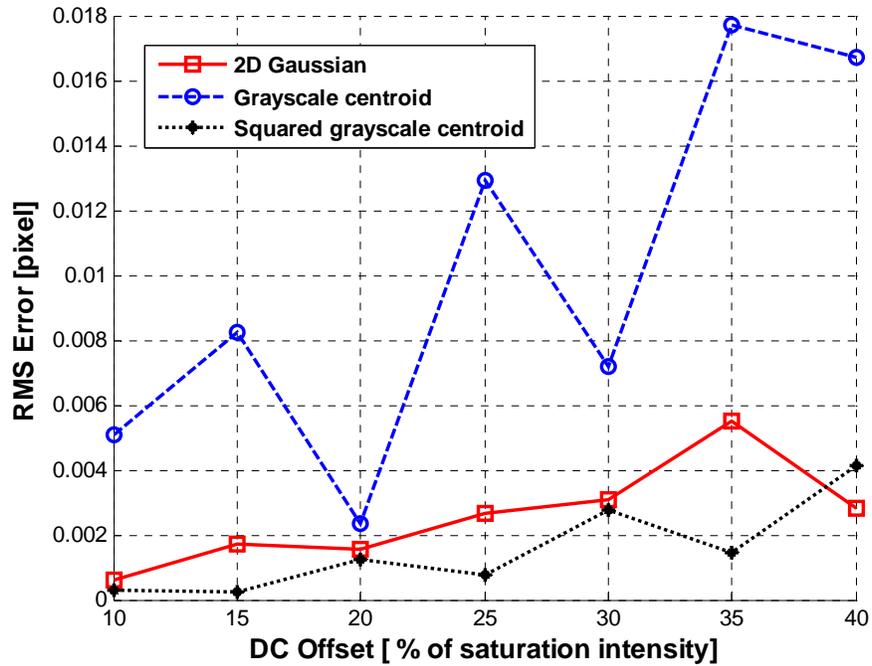
Figure 3.6: DC offset effect on target location accuracy.

The results indicate that the tracking accuracy of grayscale centroid method is affected by the DC offset level a lot and the RMS error of grayscale centroid is about 5~6 times larger than those of 2D Gaussian fitting and squared grayscale centroid when the DC offset intensity level is about 40% of saturation. 2D Gaussian fitting and squared grayscale centroid have close performance with DC offset varying.

c). Fill factor effect on target location accuracy

Typically, the fill factor of CMOS image sensors is not as good as CCD image sensors due to different pixel architectures. Most CCD image sensors can achieve 100% fill factor for every pixel, while CMOS image sensors only have less than 50% fill factor and the fill factor for our selected CMOS image sensor is only 36%. The simulation test for fill factor effect here is to verify whether the fill factor has effect on tracking accuracy for each algorithm and how much effect it has. The simulated target image is a Gaussian distribution of sigma of 2 in *x* and *y* axis

and peak intensity of 4095 with 12 bits data output. For different cases of fill factors, the target image is generated with fill factors of 9%, 16%, 25%, 36%, 49%, 64%, 81% and 100%. The light sensing parts on each pixel varying fill factor are shown in Figure 3.7. Then the tracking accuracy of each algorithm is evaluated based on the resulting target images with varying fill factors.
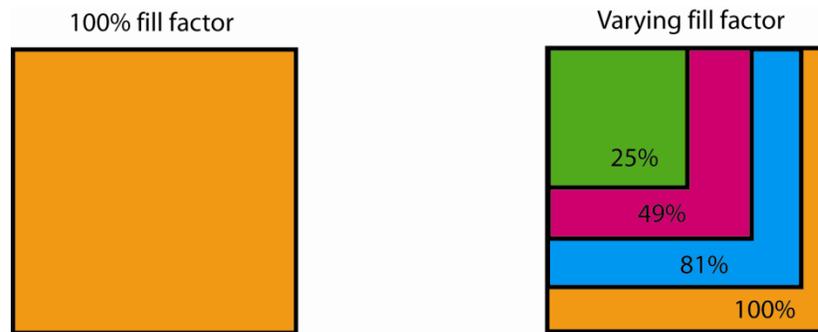
Figure 3.7: Fill factor variation illustration.

The RMS error for each algorithm is shown in Figure 3.8. From the results, it is clear that the fill factor has very small effect on the tracking accuracy of these three algorithms. Grayscale centroid method has the largest RMS error compared to the other two methods, while this RMS error is less than 0.02% pixel which is negligible. So it is concluded that the fill factor effect on tracking accuracy is very small and can be ignored.
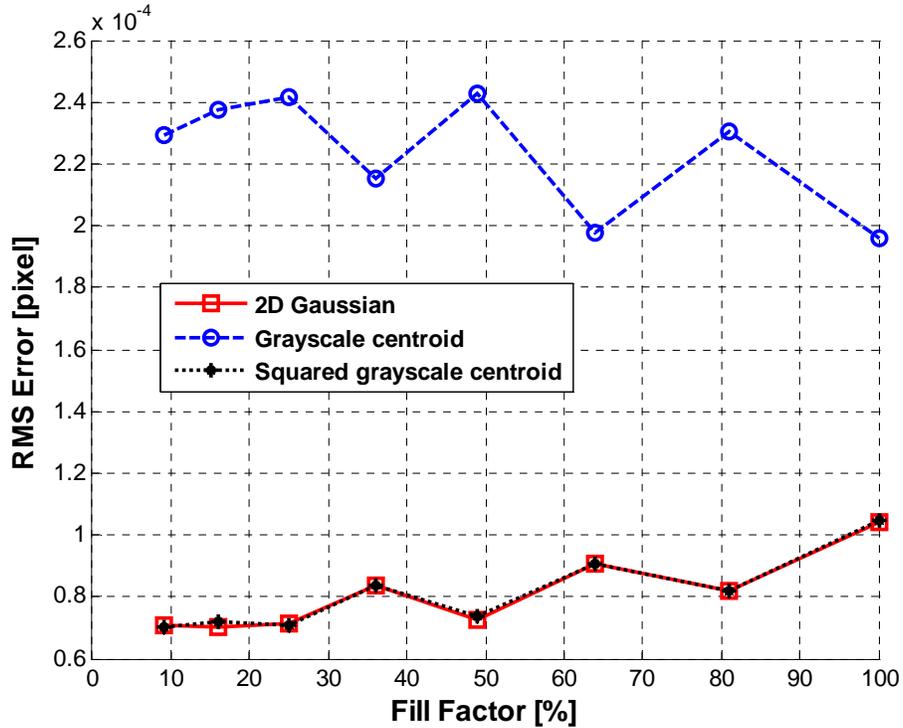
Figure 3.8: Fill factor effect on target location accuracy.

d). Saturation effect on target location accuracy

The effect of saturation is analyzed by varying the peak intensity from 100% to 150% of the saturation intensity. The target image is Gaussian shaped with sigma of 2 in *x* and *y* axis and the peak intensity varies from 4095 to 6143 with 12 bits. The ROC is 50 x 50 pixels and the RMS error is obtained by moving the target to 400 different locations. The results for tracking accuracy with three methods are shown in Figure 3.9.
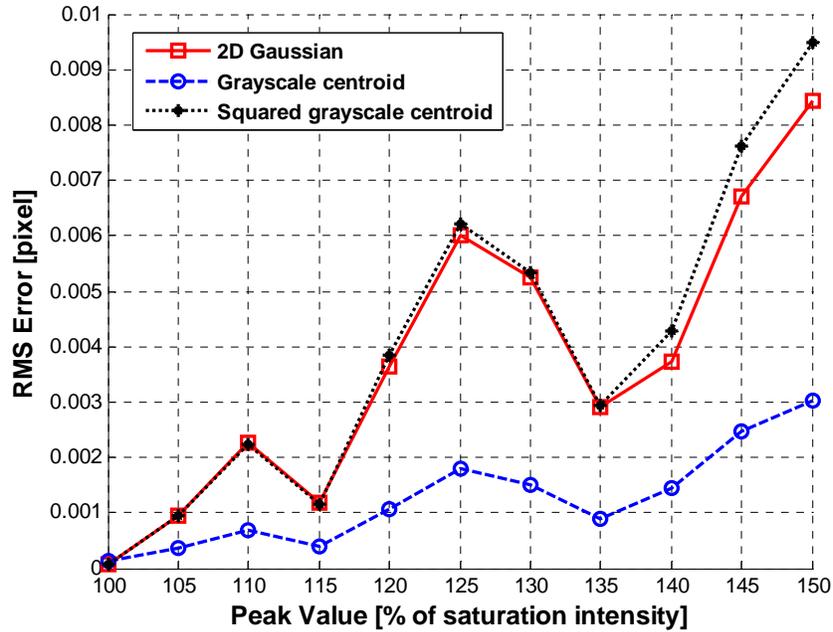
Figure 3.9: Saturation effect on target location accuracy.

The results indicate that 2D Gaussian fitting and squared grayscale centroid method are much more sensitive to saturation than grayscale centroid method. Grayscale centroid method achieves better tracking accuracy than the other two methods and it does not increase much with saturation level increasing. Based on this simulation results, it is concluded that for high saturation level of the tracking target with other conditions fixed, grayscale centroid method provides the best accuracy.

e). Conclusion based on simulation investigation

Based on the simulation investigation on these target location algorithms, it is observed that squared grayscale centroid method and 2D Gaussian fitting are not sensitive to the noise, DC offset and fill factor, while grayscale centroid is very sensitive to these factors and has much worse estimation accuracy. Squared grayscale centroid method and 2D Gaussian fitting are more sensitive to saturation level than grayscale centroid method.

If the target image is not saturated, squared grayscale centroid method and 2D Gaussian fitting can achieve better accuracy for target location. However, 2D Gaussian fitting algorithm is more complicated compared to squared grayscale centroid. It is concluded that squared centroid method is the best target location algorithm which combines high accuracy and simplicity if unsaturated target images are applied.

3.3   Summary

This chapter presented noise correction algorithms, target location algorithms, and comparison for three target location algorithms. Section 3.1 discussed the noise correction algorithms for processing raw images. In Section 3.2, three target location algorithms were described and simulation investigation was implemented for these methods under different conditions. For instance, the effect of random noise on tracking accuracy for each algorithm was analyzed and compared using these algorithms. The effect of DC offset, fill factor and saturation were also analyzed and the tracking accuracy from different algorithms was compared. It is concluded that squared centroid method is the best target location algorithm which combines high accuracy and simplicity if the unsaturated target images are applied.

Chapter 4

Camera Calibration and 3D Reconstruction Algorithms

Camera calibration is the process of estimating a set of camera parameters required to reconstruct 3D positions of the tracking target. These camera parameters includes intrinsic parameters that describe the combination of lens and image sensor, and extrinsic parameters that relate the position and orientation of the camera to a world coordinate system [38]. For 3D optical tracking systems and many other applications such as robotic vision, metrology, manufacturing inspection, and automated assembly [39], the overall performance of the system strongly depends on the accuracy of camera calibration.

Figure 4.1 shows the flow chart for a single camera calibration. The tracking target is mounted on a coordinate measuring machine (CMM) and can move together with the CMM to different positions. A camera is used to capture images for this tracking target. Then noise correction and target location algorithms are implemented to obtain accurate target locations for the tracking target. Based on target locations and 3D positions of the tracking target measured by the CMM, camera calibration algorithms are designed to obtain precise intrinsic and extrinsic parameters for this camera. To obtain camera parameters for other cameras, the same calibration procedure can be applied.

With two sets of camera parameters and target locations from two cameras, 3D reconstruction algorithms can be employed to estimate 3D positions of the tracking target.
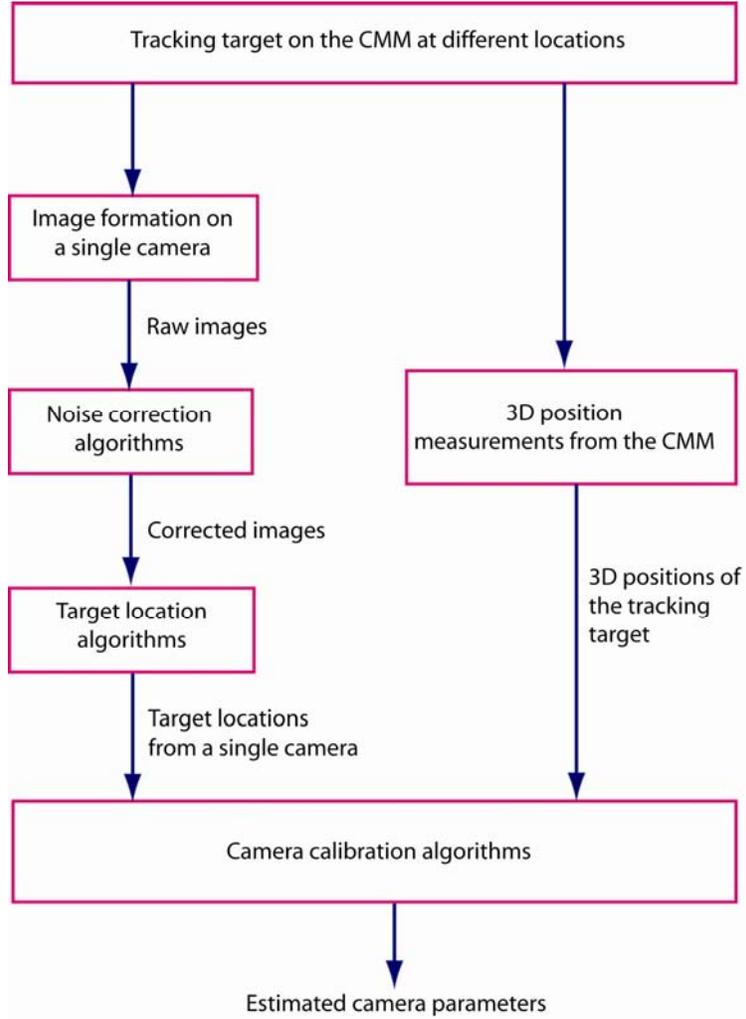
Figure 4.1: Flow chart of camera calibration for a single camera.

In this chapter, camera calibration and 3D reconstruction algorithms are presented. Section 4.3 introduces the reasons for camera calibration and what camera calibration is. Section 4.2 describes the camera model which is also the fundamental for camera calibration, including the definition of coordinate frames, mathematical modeling for 4 step image formation and summarized this section with a full camera model. Section 4.3 discusses the existing camera calibration algorithms with their advantages and disadvantages and presents the proposed

solution. Section 4.4 explains the 3D reconstruction algorithms. Section 4.5 presents simulation investigation on the DLT and proposed solution, and compares the performance of these two methods. Finally, this chapter is summarized in Section 4.6.

4.1 Introduction to camera calibration

For the 3D optical tracking system, camera calibration is very important as the overall performance of the system strongly depends on the accuracy of camera calibration. The main reasons are as follows:

a) The geometric parameters for the cameras are always unknown and the camera assembly sustains considerable internal misalignments. For instance, the distance from the lens to the image sensor is often not available. The image sensor may not be orthogonal to the optical axis, and the center of the image sensor may not coincide with the intersection of the optical axis and the image plane.

b) The lens and optics used for the cameras are not perfect and always sustain a certain amount of distortion.

Camera calibration is to calculate these camera parameters, such as the position of the optical center of the lens on the image sensor u_0, v_0, the distance from the optical center of the lens to the image sensor d, scaling factors for column pixels S_x and row pixels S_y, rotation matrix R and translation matrix T from 3D world coordinates to 3D camera coordinates, and lens distortion coefficients.

According to various criteria, the existing methods and techniques for camera calibration can be classified as follows [40]:

- **Implicit calibration and explicit calibration**

  Implicit calibration is calibrating a camera without explicitly computing its physical parameters [41] [42]. This method can be used for 3D measurement when there is no need to obtain the physical parameters for the camera, e.g. the distance of the lens optical center to the image sensor, and the optical center location on the image sensor. Explicit calibration is a process of estimating all the physical parameters of the cameras. Most of the existing research work belongs to this category [43] [44] [45] [46] [47] [48] [49] [50]. It is easy to check whether the calibration is effective or not using the explicit calibration method. However, choosing which method to use depends on the application requirements.

- **Photogrammetric calibration and self-calibration**

  The main difference between photogrammetric calibration and self-calibration is the use of calibration objects. Photogrammetric calibration is performed using 3D calibration objects with high precision [48]. The calibration objects used can be two planes of patterns with precise 3D positions for edge points of the pattern, e.g. Tsai [47] and Weng [44] employed a planar pattern with squares undergoing a precisely known translation for photogrammetric calibration, and Heikkila [50] [51] [52] used two planes orthogonal to each other with grids of circles as the calibration objects.

  Self-calibration does not use any calibration object for estimating intrinsic parameters. Just by moving a camera around within a static scene, the rigidity of the scene provides some constraints on the camera's intrinsic parameters from different views of the camera. With 3 or more images of the same scene, the intrinsic parameters can be recovered by this method. However, the self-calibration process is not as accurate as photogrammetric calibration even though it is very flexible.

Later in this chapter, only explicit photogrammetric calibration methods, which can easily check all the physical parameters of the cameras and has the potential of achieving high accuracy, are discussed.

4.2   Camera model

Before camera calibration algorithms are described, a camera model which is the fundamental for camera calibration is presented here. A camera model is a mathematical model for the mapping of the tracking target in a 3D world coordinate system to a 2D image coordinate system using the corresponding camera parameters.

In this section, coordinate frames for the camera model are defined and an overview of the four-step image formation process is given, followed by detailed description of modeling for each step in the image formation process. This includes rigid body coordinate transformation, perspective projection with a pinhole model, transformation of ideal image coordinates to distorted image coordinates, and transformation of distorted image coordinates to digitized pixel coordinates. Lastly, the full camera model is presented as a summary of this section.

4.2.1   Definition for coordinate frames

Figure 4.2 shows all the coordinate frames for the camera model. As shown in Figure 4.2, $O_wX_wY_wZ_w$ is a 3D world coordinate system, $O_cX_cY_cZ_c$ is a 3D camera coordinate system, xO_1y is a 2D image coordinate system, and uO_2v is a digitized pixel coordinate system.

- **3D world coordinates**

The coordinates of the tracking target with respect to a 3D world coordinate frame are denoted by (X_w, Y_w, Z_w).
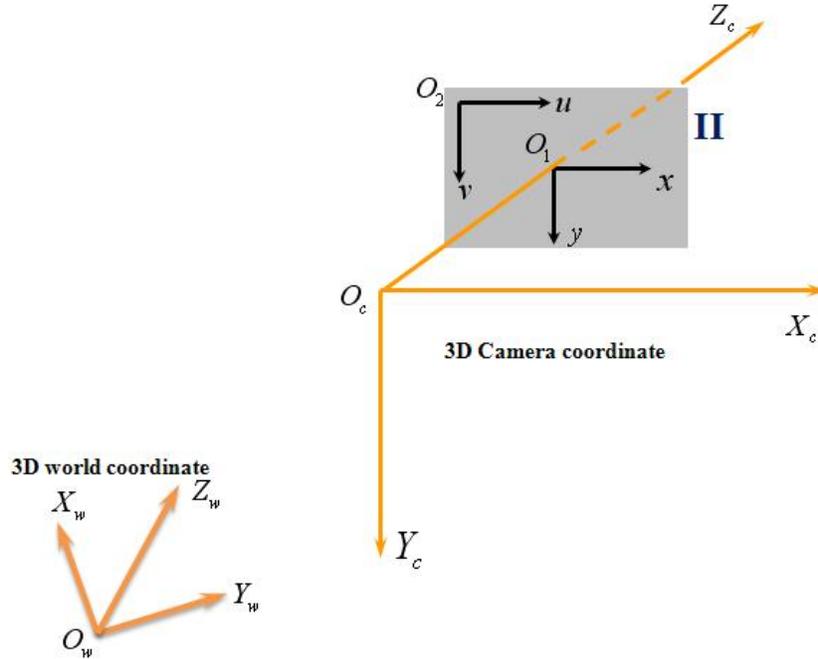
Figure 4.2: Coordinate frames for camera model: $O_w X_w Y_w Z_w$ is a 3D world coordinate system, $O_c X_c Y_c Z_c$ is a 3D camera coordinate system, $x O_1 y$ is a 2D image coordinate system, $u O_2 v$ is a digitized pixel coordinate system.

- **3D camera coordinates**

The coordinates of the tracking target with respect to the 3D camera coordinate frame are denoted by (X_c, Y_c, Z_c). The origin of the camera coordinate frame O_c is located at the optical center of the lens. The image plane II is parallel to the $X_c O_c Y_c$ plane and it is displaced a distance d from O_c along the Z_c axis [38].

- **2D image coordinates**

The 2D image coordinates of the tracking target on the image plane are denoted by (x, y) in SI units. The origin of the image coordinate frame O_1 lies at the intersection of the optical axis of the lens and image plane. This point is known as the principal point or the image center of the

67

image plane. Due to the lens distortion effect, the image points are displaced. The ideal image coordinates are denoted by (x_i, y_i), while the distorted image coordinates are denoted by (x_d, y_d). Once the image correction is applied to the distorted image, these points are denoted by (x_c, y_c).

- **2D pixel coordinates**

Using scaling factors for unit conversion from SI (meters) to pixels, 2D image coordinates (x, y) in meters can be converted to digitized 2D pixel coordinates (u, v) in pixels. The origin of the pixel coordinate frame is defined as O_2.

4.2.2 Four steps for image formation process

Figure 4.3 represents four steps for image formation process from 3D world coordinates to 2D pixel coordinates. The flow chart is described simply as follows: Step 1 represents the rigid transformation from 3D world coordinates (X_w, Y_w, Z_w) to 3D camera coordinates (X_c, Y_c, Z_c). Step 2 represents the transformation from the 3D camera coordinates (X_c, Y_c, Z_c) to ideal 2D image coordinates (x_i, y_i) using a pinhole model. Step 3 represents the transformation from ideal 2D image coordinates (x_i, y_i) to distorted 2D image coordinates (x_d, y_d). Step 4 represents the transformation from distorted 2D image coordinates (x_d, y_d) in SI units to digitized pixel coordinates (u, v) in pixels using scaling factors.

In this section, each step in the image formation process is described in detail and the parameters requiring calibration are introduced. The mathematical model of the image formation process is given as well.
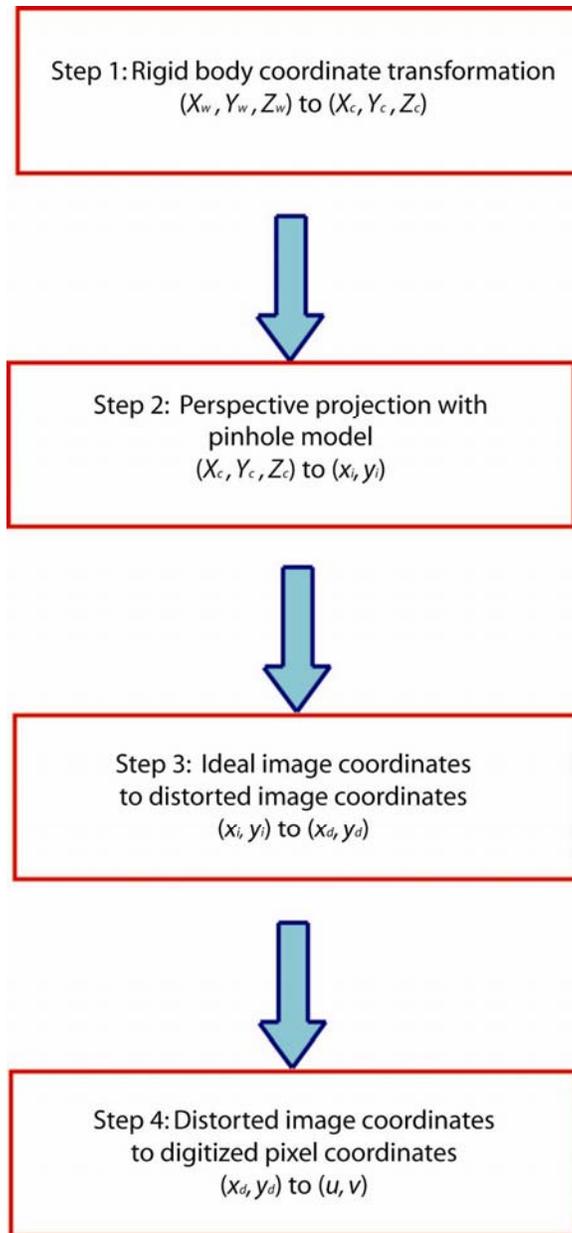
Figure 4.3: Four steps of the image formation process.

- **Step 1: Rigid body coordinate transformation**

This step represents the rigid body transformation for a 3D point in a world coordinate system (X_w, Y_w, Z_w) to a camera coordinate system (X_c, Y_c, Z_c) using a translation vector $\boldsymbol{T}$ and a rotation matrix $\boldsymbol{R}$. The mathematical relation for this step is shown in the following equation:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T , \tag{4.1}$$

where R is a 3 by 3 matrix with 3 independent Euler angles α (rotation angle around the X axis), β (rotation angle around the Y axis), and γ (rotation angle around the Z axis). The rotation matrix R is expressed as a function of α, β, and γ using:

$$R = R_z(\gamma)R_y(\beta)R_x(\alpha) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} . \tag{4.2}$$

It is important to note that the rotation matrix R performs the rotation around the X axis first, then the Y axis, and finally the Z axis. If the order of this operation is changed, the rotation matrix will be different. So in this thesis, the rotation matrix is defined to be fixed in this order. To simplify the above equation, the rotation matrix R can be rewritten as:

$$R = \begin{bmatrix} \cos\gamma\cos\beta & \sin\gamma\cos\alpha + \cos\gamma\sin\beta\sin\alpha & \sin\gamma\sin\alpha - \cos\gamma\sin\beta\cos\alpha \\ -\sin\gamma & \cos\gamma\cos\alpha - \sin\gamma\sin\beta\sin\alpha & \cos\gamma\sin\alpha + \sin\gamma\sin\beta\cos\alpha \\ \sin\beta & -\sin\alpha\cos\beta & \cos\beta\cos\alpha \end{bmatrix} . \tag{4.3}$$

If 9 parameters are used to represent the R matrix, the expression is as follows:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} , \tag{4.4}$$

where $r_{11} = \cos\gamma\cos\beta$,

$r_{12} = \sin\gamma\cos\alpha + \cos\gamma\sin\beta\sin\alpha$,

$r_{13} = \sin\gamma\sin\alpha - \cos\gamma\sin\beta\cos\alpha$,

$r_{21} = -\sin\gamma$,

$$r_{22} = \cos\gamma\cos\alpha - \sin\gamma\sin\beta\sin\alpha,$$

$$r_{23} = \cos\gamma\sin\alpha + \sin\gamma\sin\beta\cos\alpha,$$

$$r_{31} = \sin\beta,$$

$$r_{32} = -\sin\alpha\cos\beta,$$

$$r_{33} = \cos\beta\cos\alpha.$$

In Eqn. (4.1), the translation vector $\boldsymbol{T}$ can be represented as:

$$\boldsymbol{T} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = \begin{bmatrix} -r_{11}t_1 - r_{12}t_2 - r_{13}t_3 \\ -r_{21}t_1 - r_{22}t_2 - r_{23}t_3 \\ -r_{31}t_1 - r_{32}t_2 - r_{33}t_3 \end{bmatrix}. \tag{4.5}$$

The 6 independent parameters α, β, γ, t_1, t_2, and t_3, which are used to represent the rotation (α, β, γ) and translation (t_1, t_2, t_3) are also called extrinsic parameters in the camera calibration process and need to be calibrated. The rigid body coordinate transformation process is shown in Figure 4.4.
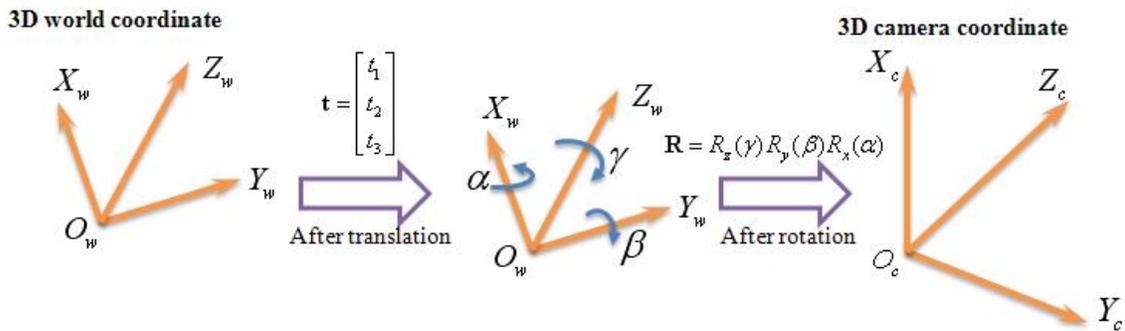
Figure 4.4: Rigid body coordinate transformation.

The rigid body transformation can also be written in a homogeneous form for matrix transformation as follows:

$$
\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}.
\tag{4.6}
$$

- **Step 2: Perspective projection using a pinhole model**

Step 2 represents the transformation from 3D camera coordinates to ideal 2D image coordinates using perspective projection using a pinhole camera model. The image formation of a tracking target M based on the pinhole model is shown in Figure 4.5.
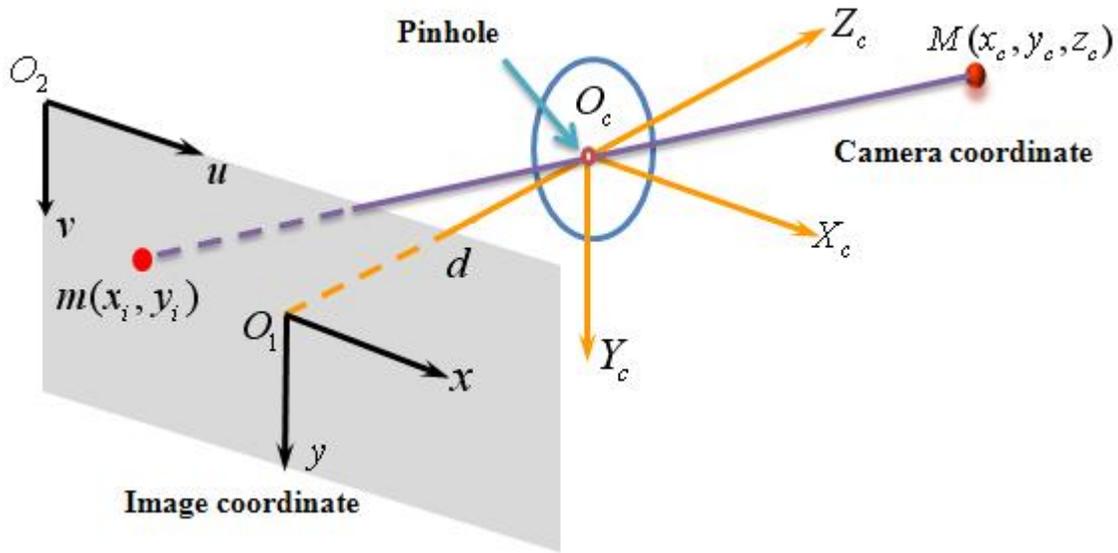
Figure 4.5: Perspective projection using a pinhole model.

In Figure 4.5, M(X_c, Y_c, Z_c) is the tracking target in the 3D camera coordinate system. The light rays from the tracking target M go through the optical center (pinhole) of the lens and interact with the 2D image sensor to form a sharp image of the point. The image coordinate of M

72

is denoted by m(x_i, y_i) on the image sensor. The mathematical equations which relate M(X_c, Y_c, Z_c) and m(x_i, y_i) are as follows:

$$
\begin{aligned}
x_i &= -d\,\frac{X_c}{Z_c} \\
y_i &= -d\,\frac{Y_c}{Z_c}
\end{aligned}
\quad,
\tag{4.7}
$$

where d is the distance from the optical center O_c to the image sensor. The parameter to be calibrated in this step is d. Eqn. (4.7) can be expressed using a homogeneous form:

$$
z_c \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}.
\tag{4.8}
$$

- **Step 3: Ideal image coordinates to distorted image coordinates**

The pinhole model in Step 2 assumes that the camera has perfect lens without any distortion. However, there is no perfect lens in reality. All the off-the-shelf lenses have distortions due to imperfections in the lens design, manufacturing and lens assembly [44]. In this step, the transformation from ideal image coordinates to distorted image coordinates is presented and the lens distortion model is discussed. Figure 4.6 shows the distorted image coordinate $m'(x_d, y_d)$ after lens distortion transformation.
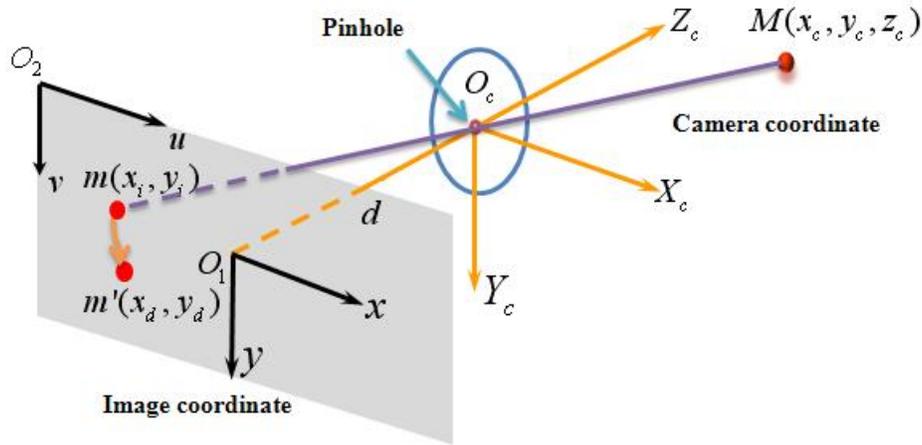
Figure 4.6: Distorted image coordinates after lens distortion.

The expressions for the distorted image coordinates from ideal image coordinates are given below:

$$
\begin{aligned}
x_d &= x_i + \delta_x(x_i, y_i) \\
y_d &= y_i + \delta_y(x_i, y_i)
\end{aligned}
\tag{4.9}
$$

where $\delta_x(x_i, y_i)$ and $\delta_y(x_i, y_i)$ are displacements due to lens distortion effects on the x axis and the y axis of the image coordinate system.

Modeling of lens distortion can be found in [44] [49] [50] [53] [54]. Typically, lens distortion can be classified into three types: radial distortion, decentering distortion and thin prism distortion [44]. Radial distortion only results in radial error in ideal image coordinates, while decentering distortion and thin prism distortion result in both radial and tangential error, see Figure 4.7.

74
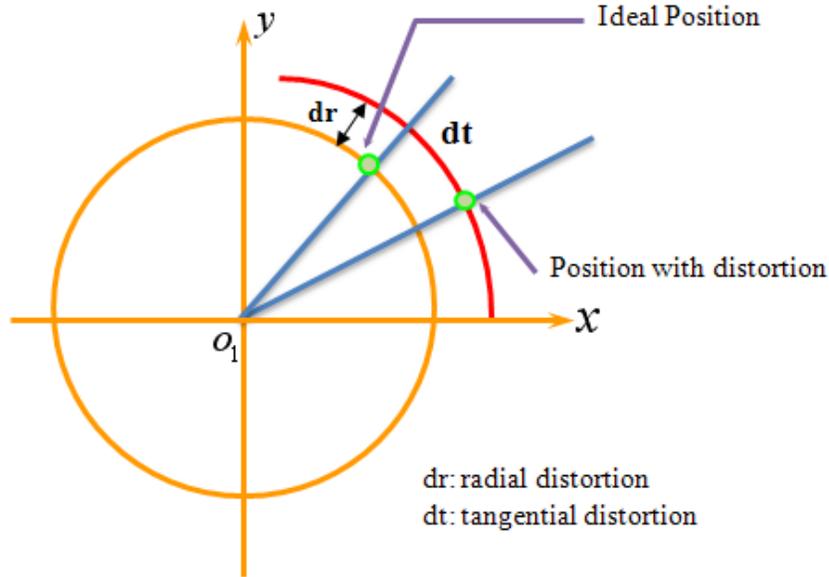
Figure 4.7: Radial and tangential distortions [44].

The causes and effects of each type of distortion are discussed as follows:

a) Radial distortion

Radial distortion is caused by a flaw in the radial curvature of the lens elements [44]. The approximated expressions for radial distortion are as follows [54]:

$$\delta x_i(x_i, y_i)^{(r)} = K_1 x_i(x_i^2 + y_i^2) + K_2 x_i(x_i^2 + y_i^2)^2 + K_3 x_i(x_i^2 + y_i^2)^3 + ...$$
$$\delta y_i(x_i, y_i)^{(r)} = K_1 y_i(x_i^2 + y_i^2) + K_2 y_i(x_i^2 + y_i^2)^2 + K_3 y_i(x_i^2 + y_i^2)^3 + ...' \qquad (4.10)$$

where K_1, K_2, K_3 are the radial distortion coefficients. Typically, the first and second terms are dominant while the other terms are negligible. The radial distortion effect on the ideal image is shown in Figure 4.8. There are two main radial distortions. One is called barrel distortion with a negative radial displacement of the image points, and the other is called pincushion distortion with a positive radial displacement of the image points. Radial distortion is a strictly symmetric distortion about the optical axis.

a : barrel distortion
b : pincushion distortion
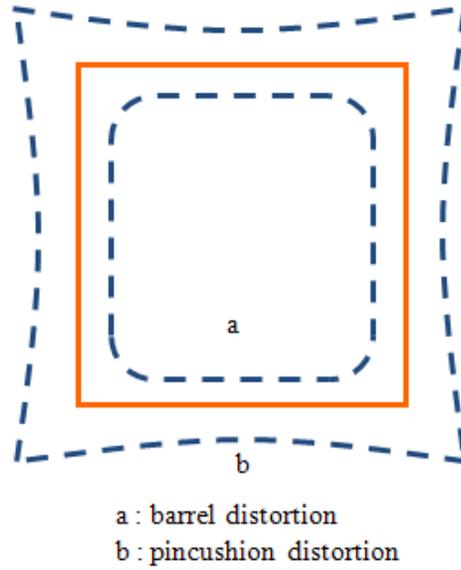
Figure 4.8: Effect of radial distortion: solid lines show no distortion and dashed lines show radial distortion (a: negative, b: positive) [44].

b) Decentering distortion

Decentering distortion is caused by the non colinearity of all the optical centers of the lens elements [44]. This distortion has both radial and tangential components, which can be expressed as follows [49] [54]:

$$
\begin{aligned}
\delta x_i(x_i, y_i)^{(d)} &= P_1(3x_i^2 + y_i^2) + P_2 x_i y_i \\
\delta y_i(x_i, y_i)^{(d)} &= P_2(3y_i^2 + x_i^2) + P_1 x_i y_i
\end{aligned}
\tag{4.11}
$$

where P_1, P_2 are the decentering distortion coefficients. Figure 4.9 illustrates the tangential distortion component effect on the ideal image positions.
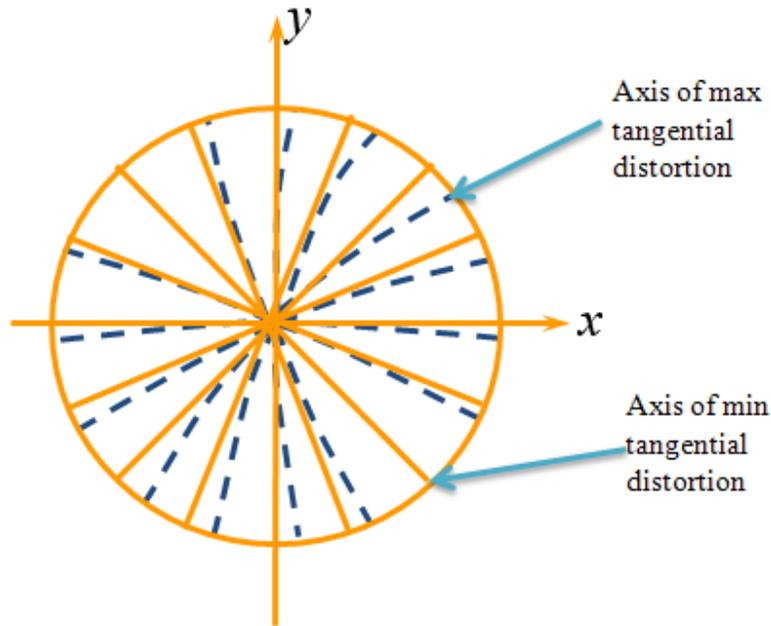
Figure 4.9: Effect of tangential distortion: solid lines show no distortion and dashed lines show tangential distortion [44].

c) Thin prism distortion

Thin prism distortion is caused by imperfections in lens design and manufacturing as well as camera assembly, such as slight tilt angle of the lens relative to the image sensor [44]. This type of distortion results in additional amount of radial and tangential distortion which can be represented by:

$$
\begin{aligned}
\delta x_i(x_i, y_i)^{(p)} &= S_1(x_i^2 + y_i^2) \\
\delta y_i(x_i, y_i)^{(p)} &= S_2(y_i^2 + x_i^2)
\end{aligned}
\tag{4.12}
$$

where S_1, S_2 are the thin prism distortion coefficients.

As the lenses vary from manufacturer to manufacturer, different lens may have different distortion due to different manufacturing errors. For instance, some lenses may only suffer from radial distortion and some may have radial distortion as well as decentering distortion. So the total lens distortion is not necessary the combination of all these distortion types. Typically,

radial distortion is common for all the off-the-shelf lenses. The distortion amount of the other two distortion types depend on the selected lenses and need to be verified by the experiments. So coefficient identification for lens distortion model based on experimental results is necessary. This is discussed in Chapter 5.

- **Step 4: Distorted image coordinates to digitized pixel coordinates**

  The last step in the image formation process is the transformation from distorted image coordinates to digitized pixel coordinates. It is well known that for all digital cameras, the outputs of the image sensors are all digitized images using pixels to represent the target location information. As the pixel shape on the image sensors is not strictly square or rectangle due to manufacturing errors, skew angle θ is introduced for modeling the transformation process, see Figure 4.10.
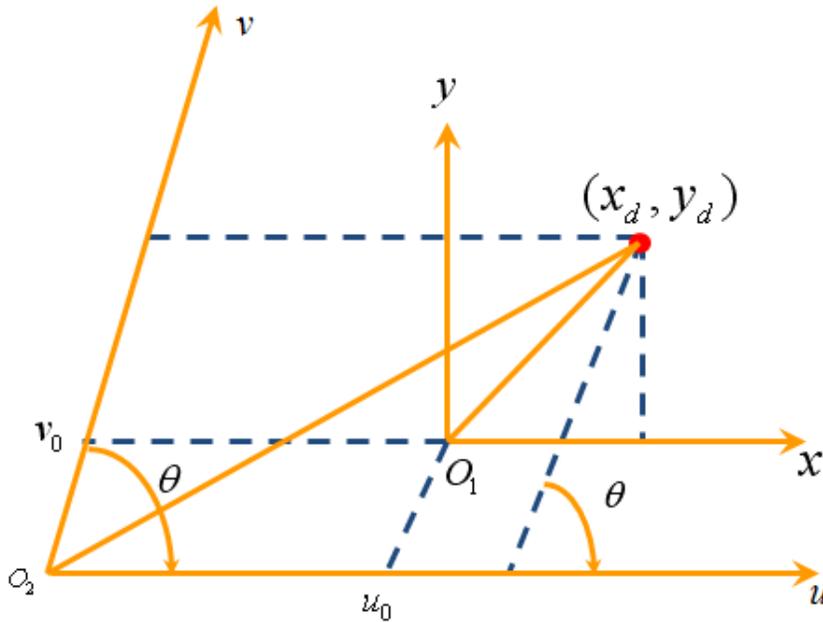
Figure 4.10: Transformation from image coordinates to pixel coordinates.

In Figure 4.10, the image coordinate system is denoted by xO_1y, and the digitized pixel coordinate system is denoted by uO_2v. If a distorted image point is located at (x_d, y_d) in the image coordinate system, then the location of this point in the digitized pixel coordinate system can be expressed as:

$$
\begin{aligned}
u &= u_0 + S_x x_d - S_x y_d \cot\theta \\
v &= v_0 + S_y \frac{y_d}{\sin\theta}
\end{aligned}
\qquad ,
\tag{4.13}
$$

where θ is the skew angle of the u and v axes, and (u_0, v_0) is the distance of the optical center O_1 to the origin of pixel coordinate O_2 in the pixel coordinate system. S_x and S_y are the scaling factors that convert the distance into pixels for both the x and y directions. The homogeneous form for Eqn. (4.13) is given as:

$$
\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} =
\begin{bmatrix}
S_x & -S_x \cot\theta & u_0 \\
0 & S_y / \sin\theta & v_0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}.
\tag{4.14}
$$

The parameters to be calibrated are u_0, v_0, S_x, S_y and θ.

4.2.3 Full camera model

To summarize the above steps, a full camera model is shown as follows:

1) 3D world coordinate system (WCS) to 3D camera coordinate system (CCS):

$$
\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} =
\begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix}
\begin{bmatrix} x_w \\ y_w \\ x_w \\ 1 \end{bmatrix}.
\tag{4.15}
$$

2) 3D camera coordinate system (CCS) to 2D ideal image coordinates:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} -d & 0 & 0 & 0 \\ 0 & -d & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \tag{4.16}$$

3) 2D ideal image coordinates to 2D distorted image coordinates:

$$\begin{aligned} x_d &= x_i + \delta_x(x_i, y_i) \\ y_d &= y_i + \delta_y(x_i, y_i) \end{aligned}. \tag{4.17}$$

As the ideal image coordinates cannot be obtained with an analytical solution to the inverse mapping [50], the above distortion model is normally replaced by the following [44]:

$$\begin{aligned} x_d &= x_i + \delta_x(x_d, y_d) \\ y_d &= y_i + \delta_y(x_d, y_d) \end{aligned}. \tag{4.18}$$

4) 2D distorted image coordinates to digitized pixel coordinates:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & -S_x \cot\theta & u_0 \\ 0 & S_y/\sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix}. \tag{4.19}$$

For all the image sensors, the difference between the real skew angle θ and its ideal value of 90° is negligible [47, 48, 52]. So during the camera calibration process, the skew angle θ is set to 90° for the simplicity of calculation.

4.3 Camera calibration methods

In the literature, there exists much research work on camera calibration. In this subsection, only the most widely used methods, such as direct linear transformation (DLT) method, Tsai's method, Weng's method and Heikkila's method are illustrated. Additionally, the advantages and

disadvantages of these methods are discussed. Later, a new camera calibration method for higher accuracy is proposed.

4.3.1   DLT method

The DLT method was first developed by Abdel-Aziz and Karara [43] and later refined in many publications. This method solves the problem of camera calibration based on a pinhole camera model and ignores nonlinear lens distortion.

Neglecting the lens distortion in the full camera model, the transformation from (X_w, Y_w, Z_w) to (u,v) can be expressed by the following equation:

$$
\begin{aligned}
u &= -d\,\frac{r_{11}(X_w-t_1)+r_{12}(Y_w-t_2)+r_{13}(Z_w-t_3)}{r_{31}(X_w-t_1)+r_{32}(Y_w-t_2)+r_{33}(Z_w-t_3)}\,S_x + u_0 \\
v &= -d\,\frac{r_{21}(X_w-t_1)+r_{22}(Y_w-t_2)+r_{23}(Z_w-t_3)}{r_{31}(X_w-t_1)+r_{32}(Y_w-t_2)+r_{33}(Z_w-t_3)}\,S_y + v_0
\end{aligned}
\tag{4.20}
$$

As the distance from the lens optical center to image sensor d is always coupled with scaling factors S_x and S_y, these three parameters are reduced to two parameters f_x, f_y which can be calculated as

$$
\begin{aligned}
f_x &= dS_x \\
f_y &= dS_y
\end{aligned}
\tag{4.21}
$$

Then Eqn. (4.20) can be rewritten as:

$$
\begin{aligned}
u &= -f_x\,\frac{r_{11}(X_w-t_1)+r_{12}(Y_w-t_2)+r_{13}(Z_w-t_3)}{r_{31}(X_w-t_1)+r_{32}(Y_w-t_2)+r_{33}(Z_w-t_3)} + u_0 \\
v &= -f_y\,\frac{r_{21}(X_w-t_1)+r_{22}(Y_w-t_2)+r_{23}(Z_w-t_3)}{r_{31}(X_w-t_1)+r_{32}(Y_w-t_2)+r_{33}(Z_w-t_3)} + v_0
\end{aligned}
\tag{4.22}
$$

The DLT method uses 11 parameters L_1 to L_{11} to represent the mathematical relationship between (X_w, Y_w, Z_w) and (u,v) without considering the physical constraints within parameters. The equation is expressed as follows [55]:

$$u = \frac{L_1 X_w + L_2 Y_w + L_3 Z_w + L_4}{L_9 X_w + L_{10} Y_w + L_{11} Z_w + 1}$$
$$v = \frac{L_5 X_w + L_6 Y_w + L_7 Z_w + L_8}{L_9 X_w + L_{10} Y_w + L_{11} Z_w + 1} \qquad (4.23)$$

where $D = -(r_{31} t_1 + r_{32} t_2 + r_{33} t_3)$,

$$L_1 = (u_0 r_{31} - f_x r_{11}) / D,$$

$$L_2 = (u_0 r_{32} - f_x r_{12}) / D,$$

$$L_3 = (u_0 r_{33} - f_x r_{13}) / D,$$

$$L_4 = \frac{(f_x r_{11} - u_0 r_{31}) t_1 + (f_x r_{12} - u_0 r_{32}) t_2 + (f_x r_{13} - u_0 r_{33}) t_3}{D},$$

$$L_5 = (v_0 r_{31} - f_y r_{21}) / D,$$

$$L_6 = (v_0 r_{32} - f_y r_{22}) / D,$$

$$L_7 = (v_0 r_{33} - f_y r_{23}) / D,$$

$$L_8 = \frac{(f_y r_{21} - v_0 r_{31}) t_1 + (f_y r_{22} - v_0 r_{32}) t_2 + (f_y r_{23} - v_0 r_{33}) t_3}{D},$$

$$L_9 = r_{31} / D,$$

$$L_{10} = r_{32} / D,$$

$$L_{11} = r_{33} / D. \qquad (4.24)$$

If there are N points for camera calibration, the solution is calculated for these 11 parameters by using the following matrix form:

$$
\begin{bmatrix}
X_{w1} & Y_{w1} & Z_{w1} & 1 & 0 & 0 & 0 & 0 & -X_{w1}u_1 & -Y_{w1}u_1 & -Z_{w1}u_1 \\
0 & 0 & 0 & 0 & X_{w1} & Y_{w1} & Z_{w1} & 1 & -X_{w1}v_1 & -Y_{w1}v_1 & -Z_{w1}v_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_{wi} & Y_{wi} & Z_{wi} & 1 & 0 & 0 & 0 & 0 & -X_{wi}u_i & -Y_{wi}u_i & -Z_{wi}u_i \\
0 & 0 & 0 & 0 & X_{wi} & Y_{wi} & Y_{wi} & 1 & -X_{wi}v_i & -Y_{wi}v_i & -Z_{wi}v_i \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_{wN} & Y_{wN} & Z_{wN} & 1 & 0 & 0 & 0 & 0 & -X_{wN}u_N & -Y_{wN}u_N & -Z_{wN}u_N \\
0 & 0 & 0 & 0 & X_{wN} & Y_{wN} & Z_{wN} & 1 & -X_{wN}v_N & -Y_{wN}v_N & -Z_{wN}v_N
\end{bmatrix}_{2N\times11}
\begin{bmatrix}
L_1 \\ L_2 \\ L_3 \\ L_4 \\ L_5 \\ L_6 \\ L_7 \\ L_8 \\ L_9 \\ L_{10} \\ L_{11}
\end{bmatrix}_{11\times1}
=
\begin{bmatrix}
u_1 \\ v_1 \\ \vdots \\ u_i \\ v_i \\ \vdots \\ u_N \\ v_N
\end{bmatrix}_{2N\times1} .
$$

$$(4.25)$$

Here, a simple form is used to represent the above equation:

$$ \mathbf{X} \cdot \mathbf{L} = \mathbf{Y} , \tag{4.26}$$

where $\mathbf{X} =
\begin{bmatrix}
X_{w1} & Y_{w1} & Z_{w1} & 1 & 0 & 0 & 0 & 0 & -X_{w1}u_1 & -Y_{w1}u_1 & -Z_{w1}u_1 \\
0 & 0 & 0 & 0 & X_{w1} & Y_{w1} & Z_{w1} & 1 & -X_{w1}v_1 & -Y_{w1}v_1 & -Z_{w1}v_1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_{wi} & Y_{wi} & Z_{wi} & 1 & 0 & 0 & 0 & 0 & -X_{wi}u_i & -Y_{wi}u_i & -Z_{wi}u_i \\
0 & 0 & 0 & 0 & X_{wi} & Y_{wi} & Y_{wi} & 1 & -X_{wi}v_i & -Y_{wi}v_i & -Z_{wi}v_i \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
X_{wN} & Y_{wN} & Z_{wN} & 1 & 0 & 0 & 0 & 0 & -X_{wN}u_N & -Y_{wN}u_N & -Z_{wN}u_N \\
0 & 0 & 0 & 0 & X_{wN} & Y_{wN} & Z_{wN} & 1 & -X_{wN}v_N & -Y_{wN}v_N & -Z_{wN}v_N
\end{bmatrix}$, $\mathbf{Y} =
\begin{bmatrix}
u_1 \\ v_1 \\ \vdots \\ u_i \\ v_i \\ \vdots \\ u_N \\ v_N
\end{bmatrix}$

and $\mathbf{L} = \begin{bmatrix} L_1 & L_2 & L_3 & L_4 & L_5 & L_6 & L_7 & L_8 & L_9 & L_{10} & L_{11} \end{bmatrix}^T$. The matrix $\mathbf{L}$ can be obtained by using linear least square methods:

$$ \mathbf{X}^T \mathbf{X} \mathbf{L} = \mathbf{X}^T \mathbf{Y} $$

$$ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \mathbf{L} = \mathbf{X}^T \mathbf{Y} $$

$$ \mathbf{L} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} , \tag{4.27}$$

where $(X^T X)^{-1} X^T$ is the pseudo-inverse of matrix X ($2N \times 11$). The singular value decomposition (SVD) which is the most reliable matrix decomposition method is employed for computing the pseudo-inverse. Suppose the SVD of matrix X is [56]:

$$X = U \sum V^T, \tag{4.28}$$

where U is a 2N by 2N orthogonal matrix over X, matrix $\sum$ is a $2N \times 11$ diagonal matrix with nonnegative real numbers on the diagonal and V^T denotes the transpose of matrix V which is a 11 by 11 orthogonal matrix over X. So the solution for matrix L can be represented by:

$$L = V \sum{}^{-1} U^T Y. \tag{4.29}$$

Note that in order to obtain all the 11 parameters of matrix L, at least 6 points with 12 equations are needed. All the points should not be on the same plane to avoid the singularity of matrix X.

After obtaining matrix L, extrinsic parameters and intrinsic parameters without lens distortion can be calculated. Based on Eqn. (4.24), the following expression is obtained as:

$$\begin{aligned} L_1 t_1 + L_2 t_2 + L_3 t_3 &= -L_4 \\ L_5 t_1 + L_6 t_2 + L_7 t_3 &= -L_8 \\ L_9 t_1 + L_{10} t_2 + L_{11} t_3 &= -1 \end{aligned} \tag{4.30}$$

Then the translation matrix T can be calculated as:

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = \begin{bmatrix} L_1 & L_2 & L_3 \\ L_5 & L_6 & L_7 \\ L_9 & L_{10} & L_{11} \end{bmatrix}^{-1} \begin{bmatrix} -L_4 \\ -L_8 \\ -1 \end{bmatrix}. \tag{4.31}$$

Similarly, D, u_0, v_0, f_x and f_y can be calculated as:

$$D^2 = \frac{1}{L_9{}^2 + L_{10}{}^2 + L_{11}{}^2}, \tag{4.32}$$

$$u_0 = \frac{L_1 L_9 + L_2 L_{10} + L_3 L_{11}}{L_9{}^2 + L_{10}{}^2 + L_{11}{}^2},$$
$$v_0 = \frac{L_5 L_9 + L_6 L_{10} + L_7 L_{11}}{L_9{}^2 + L_{10}{}^2 + L_{11}{}^2}, \tag{4.33}$$

$$f_x{}^2 = \frac{(u_0 L_9 - L_1)^2 + (u_0 L_{10} - L_2)^2 + (u_0 L_{11} - L_3)^2}{L_9{}^2 + L_{10}{}^2 + L_{11}{}^2},$$
$$f_y{}^2 = \frac{(v_0 L_9 - L_5)^2 + (v_0 L_{10} - L_6)^2 + (v_0 L_{11} - L_7)^2}{L_9{}^2 + L_{10}{}^2 + L_{11}{}^2}. \tag{4.34}$$

Then the rotation matrix R can be expressed as:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = D \begin{bmatrix} \dfrac{u_0 L_9 - L_1}{f_x} & \dfrac{u_0 L_{10} - L_2}{f_x} & \dfrac{u_0 L_{11} - L_3}{f_x} \\ \dfrac{v_0 L_9 - L_5}{f_y} & \dfrac{v_0 L_{10} - L_6}{f_y} & \dfrac{v_0 L_{11} - L_7}{f_y} \\ L_9 & L_{10} & L_{11} \end{bmatrix}. \tag{4.35}$$

For the signs on D, f_x and f_y, it is obvious that f_x and f_y are positive. However, D can be positive or negative. Use the positive value first and compute the determinant of the transformation matrix R obtained [55]. If the determinant is positive, D must be positive and the sign for R is correct. Otherwise, D is negative. Multiply -1 for matrix R to get the correct sign for it. Three rotation angles can be computed from matrix R with the specific rotation order defined previously.

As no iteration is required, the DLT method is simple and computationally fast. It is also the best solution when there is no lens distortion. However, the assumption for no lens distortion is not realistic. This method does not consider lens distortion effects and cannot incorporate the lens distortion into it, which makes this method less accurate since all the off-the-shelf lenses are not perfect and have a certain amount of distortion. Especially for 3D motion measurement,

ignoring lens distortion is not acceptable and will result in poor accuracy (The reported accuracy obtained by the DLT method is one in 2000 parts [44]).

4.3.2 Tsai's method

Tsai's method is a two-step method [47] that solves most of the parameters using a direct linear transformation based on the radial alignment constraint (RAC) in the first step, while the remaining parameters are determined by nonlinear iteration algorithm in the second step. The RAC is based on an assumption that the lens of the camera only has radial distortion and this assumption helps to establish an additional radial alignment constraint $\overline{O_2 P_i} \,/\!/\, \overline{O_2 P_d} \,/\!/\, \overline{P_{O_1Z} P}$ for the target location on the image, as shown in Figure 4.11.
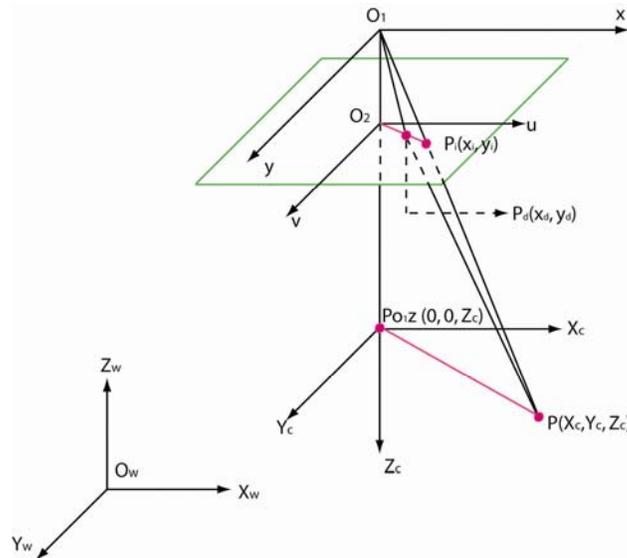
Figure 4.11: Illustration of radial alignment constraint. Radial distortion does not alter direction of vector from origin to image point, which leads to $\overline{O_2 P_i} \,/\!/\, \overline{O_2 P_d} \,/\!/\, \overline{P_{O_1Z} P}$.

Based on the RAC, the direct linear transformation for solving extrinsic parameters of α, β, γ, t_1, and t_2 is derived. Then, an iterative scheme is used to estimate the translation along Z axis (t_3), distance from the optical center of the lens to the image sensor (d), and radial distortion coefficients (K_1, K_2).

There are several advantages of Tsai's method. First, it has fast computation speed with less nonlinear iteration for only 4 parameters. Additionally, radial lens distortion is incorporated to achieve better accuracy for camera calibration (e.g. one part in 4000). With these considerable advantages, Tsai's method has been widely used by many applications in computer vision and machine vision with very good accuracy.

However, there are some disadvantages of Tsai's method. One is that this method does not work when the lens distortion is very small because the closed-form solution is based on RAC which assumes large radial distortion (>1% distortion over a full frame). Also this method cannot be extended to other types of distortion. If there is any tangential distortion error for the lenses due to misalignment of lenses, this method will become less accurate. Besides, this method assumes the principal point to be known in advance which is not always feasible. Lastly, not all the calibrated parameters are optimized which means the final solution is not a fully optimized solution.

4.3.3 Weng's method

Weng's method is also a two-step method [44] which firstly uses the DLT method to calculate all the extrinsic and intrinsic parameters without lens distortion coefficients. Then it applies nonlinear optimization method for all the camera parameters including lens distortion coefficients based on the initial guess from the DLT method. Compared with Tsai's method,

Weng's method does not apply any additional radial alignment constraints and considers the lens distortion model with 2nd and 3rd order as

$$\delta x_i(x_i, y_i) = K_1 x_i(x_i^2 + y_i^2) + P_1(3x_i^2 + y_i^2) + P_2 x_i y_i + S_1(x_i^2 + y_i^2)$$
$$\delta y_i(x_i, y_i) = K_1 y_i(x_i^2 + y_i^2) + P_2(3y_i^2 + x_i^2) + P_1 x_i y_i + S_2(y_i^2 + x_i^2)$$

(4.36)

There are several advantages for Weng's method. Firstly, Weng's algorithm fully optimizes all the parameters instead of partial parameters in Tsai's method. In addition, Weng's algorithm involves not only radial lens distortion but also other types of lens distortion in the lens distortion model.

Despite these advantages of Weng's method, it also has some disadvantages. For instance, only 2nd and 3rd order components of lens distortion are considered but 5th order components or higher may also be important for the lens distortion model. Besides, the convergence rate of his method is too slow and it is difficult to converge within a few iterations.

4.3.4   Heikkila's method

The method developed by Heikkila [50] [51] [52] also uses the DLT method as an initial guess for all the intrinsic and extrinsic parameters without lens distortion and then applies nonlinear optimization algorithms for all the parameters including lens distortion coefficients. Compared with Weng's method, Heikkila's method includes 5th order radial distortion and decentering distortion components in the lens distortion model developed by Brown [49]:

$$\delta x_i(x_i, y_i) = K_1 x_i(x_i^2 + y_i^2) + K_2 x_i(x_i^2 + y_i^2)^2 + P_1(3x_i^2 + y_i^2) + P_2 x_i y_i$$
$$\delta y_i(x_i, y_i) = K_1 y_i(x_i^2 + y_i^2) + K_2 y_i(x_i^2 + y_i^2)^2 + P_2(3y_i^2 + x_i^2) + P_1 x_i y_i$$

(4.37)

This method is more accurate than Weng's method by considering the 5th order component in the lens distortion model. In addition, this method uses Levenberg-Marquardt optimization method to solve the nonlinear least square problem which has been shown to provide the fastest convergence [50].

However, Heikkila's lens distortion model includes the decentering distortion which is not important for the wide angle lenses in the optical tracking systems. Additionally, more distortion coefficients make the parameter estimation less accurate in the nonlinear optimization process. So it is necessary to identify the accurate lens distortion model for the selected lenses to achieve high accuracy.

4.3.5 Proposed method

The proposed method for camera calibration is a combination of Heikkila's method and lens distortion identification. The novel part of the proposed method is the lens distortion identification. Instead of fixing the lens distortion model to a certain type which may give relative low accuracy, the lens distortion identification is to identify the distortion coefficients for the selected lens based on the experimental results. The distortion coefficients of the selected lens can be determined based on the accuracy of the 3D estimation. This method is more flexible than Heikkila's method and can achieve higher accuracy with less distortion coefficients based on the identification.

The proposed method contains the DLT method and nonlinear least square optimization to compute the intrinsic parameters (d, u_0, v_0, S_x, S_y) and extrinsic parameters (α, β, γ, t_1, t_2, t_3) as well as lens distortion coefficients (K_1, K_2). The lens distortion model used is as:

$$\delta x_i(x_i, y_i) = K_1 x_i(x_i^2 + y_i^2) + K_2 x_i(x_i^2 + y_i^2)^2$$
$$\delta y_i(x_i, y_i) = K_1 y_i(x_i^2 + y_i^2) + K_2 y_i(x_i^2 + y_i^2)^2 \cdot \tag{4.38}$$

The flow chart for the proposed method is shown in Figure 4.12. The initial values for

intrinsic and extrinsic parameters are calculated from the DLT method. However, lens distortion

parameters cannot be estimated from the DLT method. In order to obtain an accurate solution for

all the camera parameters, a nonlinear iteration method is needed for the full camera model.
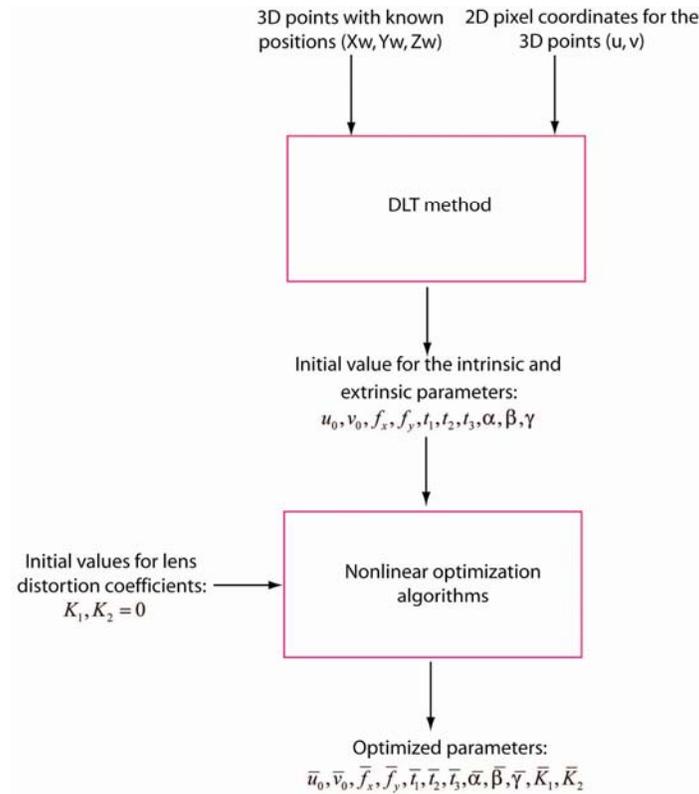
Figure 4.12: Flow chart of the proposed solution.

The objective of the nonlinear optimization is to obtain optimized intrinsic and extrinsic

parameters for the camera model as well as lens distortion coefficients. Meanwhile, the estimated

2D pixel coordinates (u_i, v_i) should be as close as the observed coordinates ($\hat{u}_i$, $\hat{v}_i$), using the

estimated camera parameters and the 3D positions. In this case, the cost function for the nonlinear optimization method can be expressed as

$$f = \sum_{i=1}^{N} [(\hat{u}_i - u_i(R,T,f_x,f_y,u_0,v_0,P))^2 + (\hat{v}_i - v_i(R,T,f_x,f_y,u_0,v_0,P))^2],$$ (4.39)

where $\hat{u}_i$ and $\hat{v}_i$ are the observed pixel coordinates of the 3D points, $u_i(R,T,f_x,f_y,u_0,v_0,P)$ and $v_i(R,T,f_x,f_y,u_0,v_0,P)$ are the optimized pixel coordinates, and P represents the lens distortion parameters (K_1, K_2). Now the estimation problem becomes a nonlinear least square optimization problem. To solve this nonlinear least square problem, the Levenberg-Marquadt optimization method is known to give the fastest convergence with high accuracy [50]. For the sake of a global minimum solution for the nonlinear optimization, a good initial guess is needed to start the iterations. The calculation results from the DLT method and the lens distortion coefficients set to zero are used as the initial condition.

4.4   3D reconstruction algorithms

After all the camera parameters are estimated from the camera calibration method, 3D reconstruction is the process of estimating 3D positions (X_w, Y_w, Z_w) of the tracking target based on these camera parameters and two image coordinates of the target (u_l, v_l), (u_r, v_r), as shown in Figure 4.13.
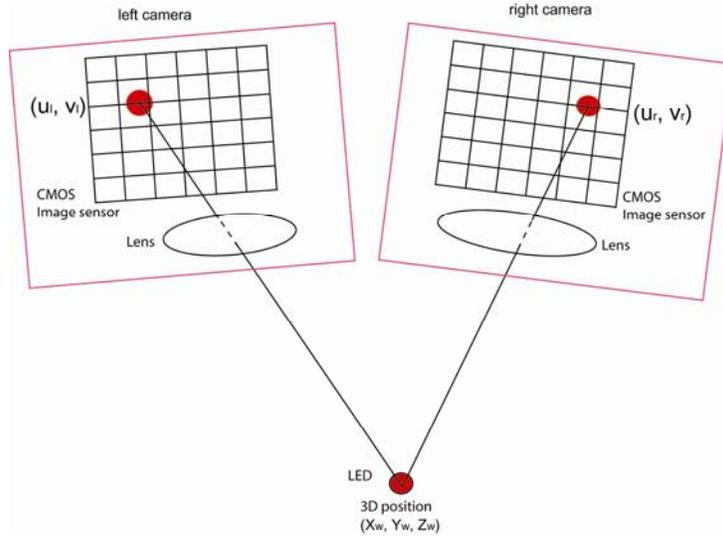
Figure 4.13: 3D reconstruction algorithms for 3D optical tracking.

Once all the intrinsic and extrinsic parameters of the camera as well as lens distortion coefficients are obtained from camera calibration, the distorted pixel coordinates of the tracking target can be corrected and the ideal pixel coordinates (u_l, v_l) from left camera and (u_r, v_r) from right camera are obtained. According to the DLT method, the image formation process from 3D world coordinate to 2D pixel coordinates for the left camera can be expressed as:

$$
\begin{aligned}
u_l &= \frac{L_{1l}X_w + L_{2l}Y_w + L_{3l}Z_w + L_{4l}}{L_{9l}X_w + L_{10l}Y_w + L_{11l}Z_w + 1} \\
v_l &= \frac{L_{5l}X_w + L_{6l}Y_w + L_{7l}Z_w + L_{8l}}{L_{9l}X_w + L_{10l}Y_w + L_{11l}Z_w + 1}
\end{aligned}
\tag{4.40}
$$

where L_{1l}, L_{2l}, L_{3l}, L_{4l}, L_{5l}, L_{6l}, L_{7l}, L_{8l}, L_{9l}, L_{10l}, L_{11l} represents all the information for intrinsic and extrinsic parameters of the left camera. Similarly, image formation process for the right camera can also be represented by

$$
\begin{aligned}
u_r &= \frac{L_{1r}X_w + L_{2r}Y_w + L_{3r}Z_w + L_{4r}}{L_{9r}X_w + L_{10r}Y_w + L_{11r}Z_w + 1} \\
v_r &= \frac{L_{5r}X_w + L_{6r}Y_w + L_{7r}Z_w + L_{8r}}{L_{9r}X_w + L_{10r}Y_w + L_{11r}Z_w + 1}
\end{aligned}
\tag{4.41}
$$

where L_{1r}, L_{2r}, L_{3r}, L_{4r}, L_{5r}, L_{6r}, L_{7r}, L_{8r}, L_{9r}, L_{10r}, L_{11r} represents all the information for intrinsic and extrinsic parameters of the right camera.

Rearranging Eqn. (4.40), the following expression is obtained:

$$(u_l L_{9l} - L_{1l})X_w + (u_l L_{10l} - L_{2l})Y_w + (u_l L_{11l} - L_{3l})Z_w = L_{4l} - u_l$$
$$(v_l L_{9l} - L_{5l})X_w + (v_l L_{10l} - L_{6l})Y_w + (v_l L_{11l} - L_{7l})Z_w = L_{8l} - v_l \qquad (4.42)$$

For 3D position estimation of the tracking target, pixel coordinates from both left camera and right camera are needed for triangulation calculation. So 3D positions of the tracking target can be calculated using the matrix as

$$\begin{bmatrix} u_l L_{9l} - L_{1l} & u_l L_{10l} - L_{2l} & u_l L_{11l} - L_{3l} \\ v_l L_{9l} - L_{5l} & v_l L_{10l} - L_{6l} & v_l L_{11l} - L_{7l} \\ u_r L_{9r} - L_{1r} & u_r L_{10r} - L_{2r} & u_r L_{11r} - L_{3r} \\ v_r L_{9r} - L_{5r} & v_r L_{10r} - L_{6r} & v_r L_{11r} - L_{7r} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \begin{bmatrix} L_{4l} - u_l \\ L_{8l} - v_l \\ L_{4r} - u_r \\ L_{8r} - v_r \end{bmatrix}, \qquad (4.43)$$

where the first two rows are parameters from the left camera, and the last two rows are parameters from the right camera.

The solution for this matrix is based on linear least square method. So the accuracy of 3D position estimation depends on the accuracy of parameter estimation from camera calibration and the accuracy of target location methods. If all these estimations are error free, the 3D position estimation from 3D reconstruction algorithms will also be error free.

4.5 Simulation investigation

In this section, the simulation investigation for evaluating the performance of the DLT method and the proposed method (denoted as NDLT in the following sections) using synthetic image data is discussed. The generation of synthetic image data includes using the full camera

model and defined camera parameters, which are later referred as the ground truth for all the simulations.

In this subsection, three cases are discussed in order to verify which method (DLT or NDLT method) is better under different conditions. Case I is a simulation comparison with these two methods using synthetic data without noise or distortion. Case II is a simulation comparison using distortion-free synthetic data with noise. Case III is a simulation comparison using synthetic data with noise and distortion. Also, the effect of the constraints for matrix R is studied.

4.5.1   Case I: comparison using synthetic data without noise or distortion

Based on the full camera model, synthetic pixel coordinates without noise or lens distortion can be obtained using a set of defined camera parameters and 3D positions. The performance of the DLT method and NDLT method are evaluated and compared based on the estimation errors of all the camera parameters. The procedure for this simulation is shown in Figure 4.14.

As shown in Figure 4.14, the tracking target moves to 192 different positions in the 3D world coordinate and 192 target locations are generated based on the full camera model without lens distortion. Note that the number of 192 is used for all the simulations due to the fact that the number of data points affects the estimation accuracy. Based on the simulation results, a number smaller than this value degrades the estimation accuracy and a number larger than this one does not help to improve the accuracy. Therefore, in all the simulation and experimental tests, the value of 192 is applied.
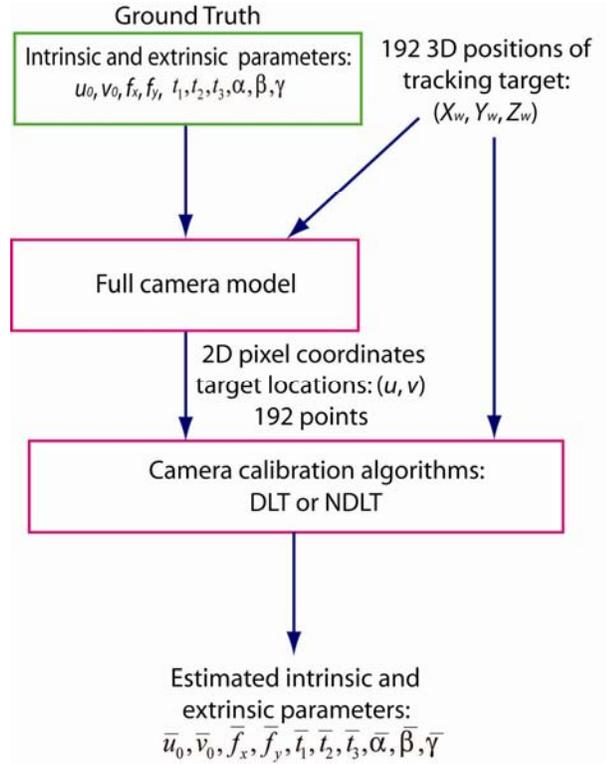
Figure 4.14: Flow chart of simulation using synthetic data without noise or distortion.

The estimated results for the intrinsic and extrinsic parameters are shown in Table 4.1. The error is calculated using the following expression:

$$\text{Error} = \left| (\text{Estimation - Ground truth}) / \text{Ground truth} \right| \times 100 \quad (\%). \qquad (4.44)$$

Table 4.1: Estimation using synthetic data without noise or lens distortion.

Parameters	Ground Truth	DLT_esti	NDLT_esti	DLT_error(%)	NDLT_error(%)
t_1(mm)	100	100.000	100.000	0.000	0.000
t_2(mm)	200	200.000	200.000	0.000	0.000
t_3(mm)	-900	-900.000	-900.000	0.000	0.000
α (deg)	2	2.000	2.000	0.000	0.000
β (deg)	5	5.000	5.000	0.000	0.000
γ (deg)	-179	-179.000	-179.000	0.000	0.000
f_x(scaling)	2141.6667	2141.667	2141.667	0.000	0.000
f_y(scaling)	2141.6667	2141.667	2141.667	0.000	0.000
u_0(pixel)	1024	1024.000	1024.000	0.000	0.000
v_0(pixel)	1024	1024.000	1024.000	0.000	0.000

The results indicate that both the DLT method and the NDLT method obtain zero error for parameter estimation when there is no noise or lens distortion added to the 2D pixel coordinates. This means that both DLT method and NDLT method achieves the best accuracy for the data without noise or lens distortion. In this case, the DLT method is better than NDLT method as it requires no iteration and less computation cost.

4.5.2   Case II: comparison using distortion-free synthetic data with noise

In this case, the procedure is the same as previous case except that Gaussian noise with a mean of zero is added to the generated target locations, as shown in Figure 4.15.
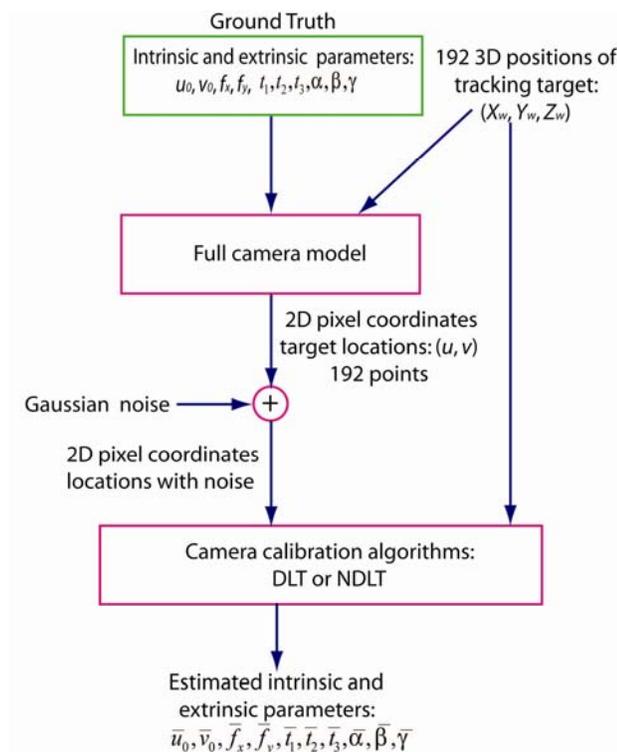
Figure 4.15: Flow chart of simulation using distortion-free synthetic data with noise.

Given the standard deviation σ of the Gaussian noise as 0.01 pixels, the estimation results for the DLT method and NDLT method are shown in Table 4.2. The error calculation is the same as Eqn. (4.44). The results demonstrate that the estimation errors from the DLT method and NDLT method are very close. Nonlinear optimization results have a marginal improvement from the DLT initial value. This also indicates that the DLT method is as accurate as nonlinear optimization if there is no lens distortion and only Gaussian noise is considered.

Table 4.2: Estimation using distortion-free synthetic data with noise (σ=0.01 pixels).

Parameters	Ground Truth	DLT_esti	NDLT_esti	DLT_error(%)	NDLT_error(%)
t_1(mm)	100	99.951	99.960	0.049	0.040
t_2(mm)	200	200.090	200.055	0.045	0.027
t_3(mm)	-900	-900.666	-900.638	0.074	0.071
α (deg)	2	2.001	2.001	0.055	0.052
β (deg)	5	5.005	5.004	0.090	0.087
γ (deg)	-179	-179.000	-179.000	0.000	0.000
f_x(scaling)	2141.6667	2143.263	2143.191	0.075	0.071
f_y(scaling)	2141.6667	2143.265	2143.192	0.075	0.071
u_0(pixel)	1024	1024.193	1024.202	0.019	0.020
v_0(pixel)	1024	1024.113	1024.035	0.011	0.003

However, the above results only show one particular case for considering the noise (σ = 0.01 pixels). As noise is unavoidable in all the images, the sensitivity of these two methods to noise should also be investigated. Figure 4.16 shows the parameter estimation results using the DLT method and the nonlinear optimization method with the noise effect (σ varies from 0 to 3 pixels).

(a) t1

(b) t2

(c) t3

(d) α

(e) β

(f) γ
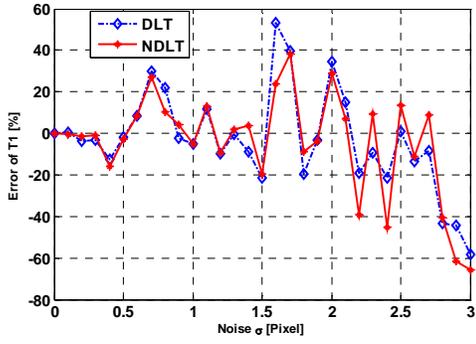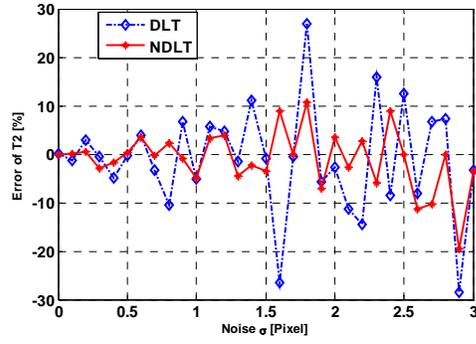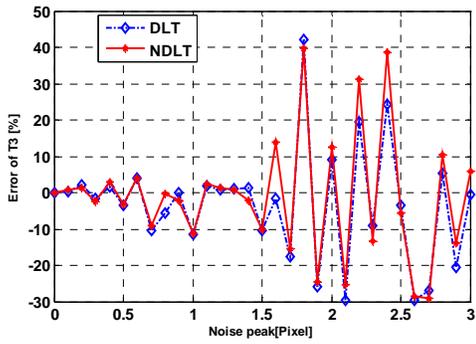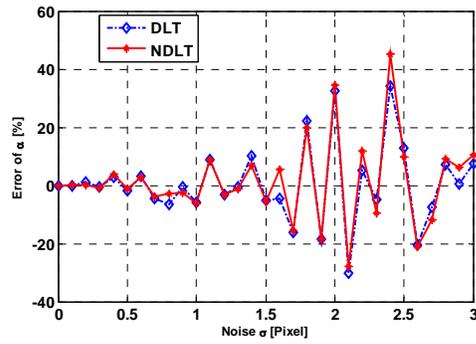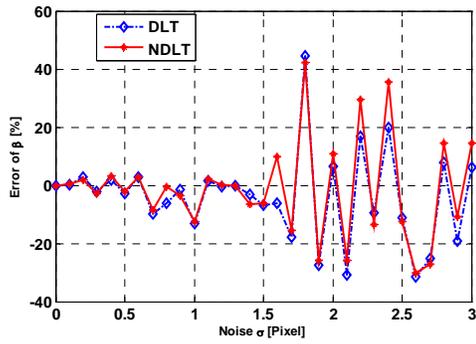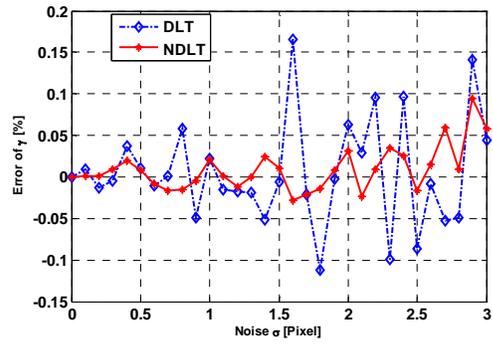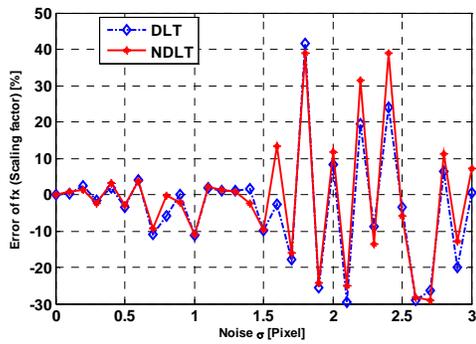
(g) fx
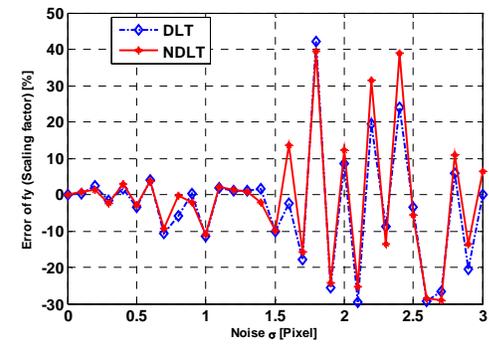
(h) fy
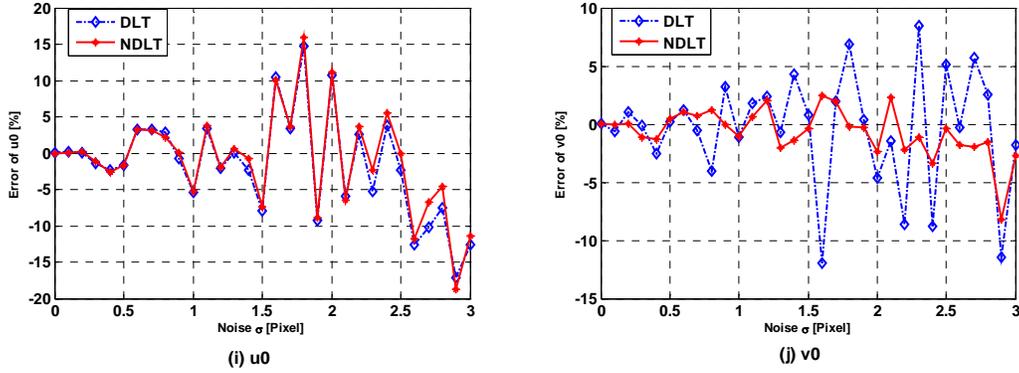
Figure 4.16: Noise effect on the estimation errors of all the camera parameters using the DLT method and NDLT method.

From the above results, it is difficult to conclude which method is better in case of noise. For instance, the estimation of some parameters such as t_1 and β is slightly better using the DLT method than those using the NDLT method, while the estimation of some other parameters such as t_2, γ and v_0 using the NDLT method is slightly better than those using the DLT method. Meanwhile, the estimation errors for t_3, α, β, f_x, f_y and u_0 are almost the same using both methods.

Another way to evaluate the calibration methods is to compare 3D position estimation errors, which completely depend on the accuracy of camera calibration algorithms. As shown in Figure 4.17, this evaluation procedure contains two sets of camera parameter estimation using the same calibration algorithms, and two cameras are used for the 3D reconstruction. As shown in Figure 4.17, 100 2D pixel coordinates are used for calibration algorithms and 92 points are used for the accuracy evaluation of the 3D position estimation.
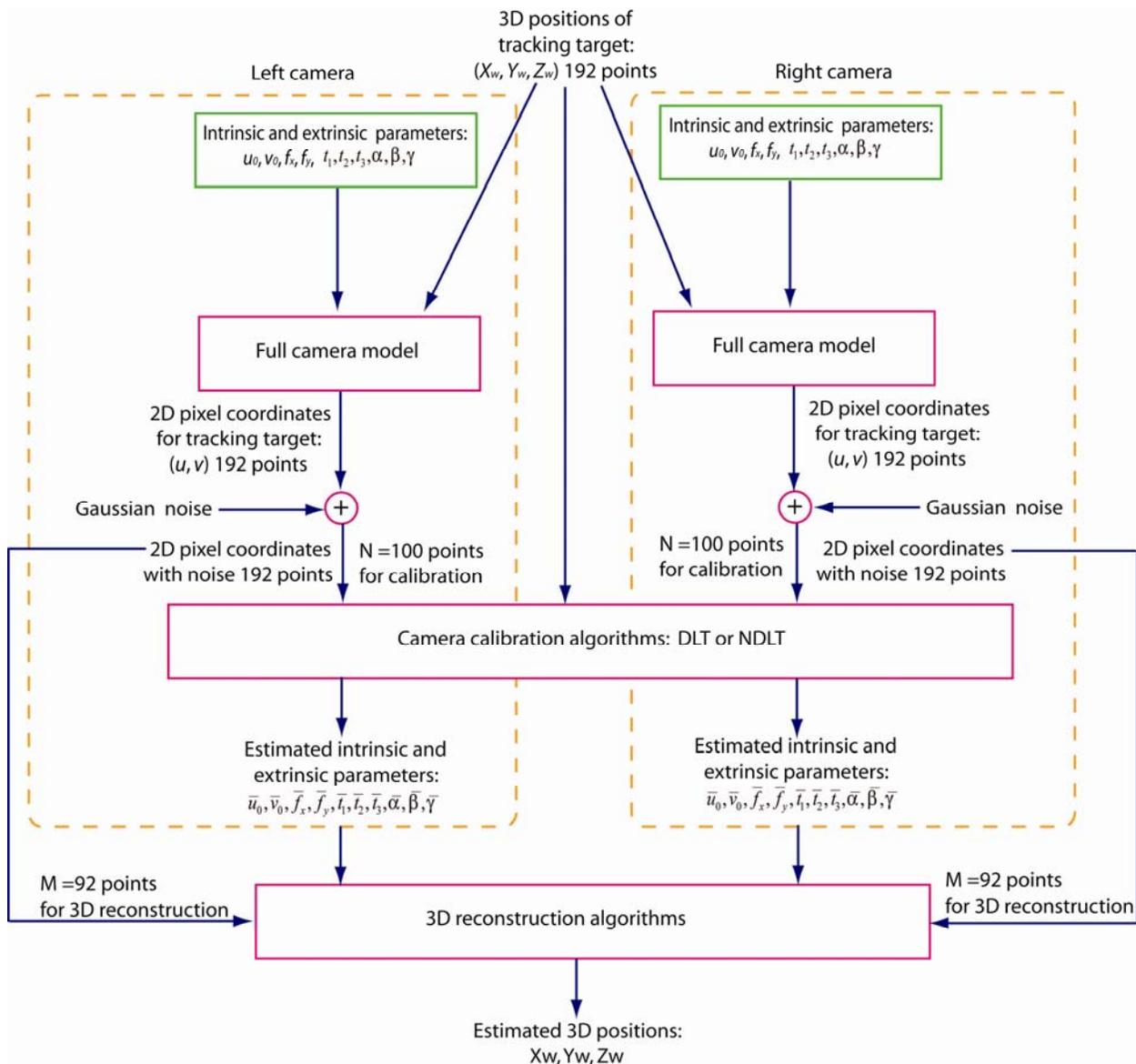
Figure 4.17: Flow chart of the 3D position estimation using distortion-free noisy synthetic data.

After the 3D positions are estimated from the 3D reconstruction algorithms, the RMS error, which is the root mean square of the difference between the estimated 3D positions and true 3D positions, can be calculated and applied to evaluate the performance of camera calibration algorithms. Here, the RMS error in 3D is also used for the performance evaluation and defined as:

$$RMS\_3D = \sqrt{(RMS\_X_w)^2 + (RMS\_Y_w)^2 + (RMS\_Z_w)^2}$$ (4.45)

With the noise σ varying from 0 to 3 pixels, the RMS errors in X_w, Y_w, Z_w and 3D using DLT method and NDLT method are shown in Figure 4.18.
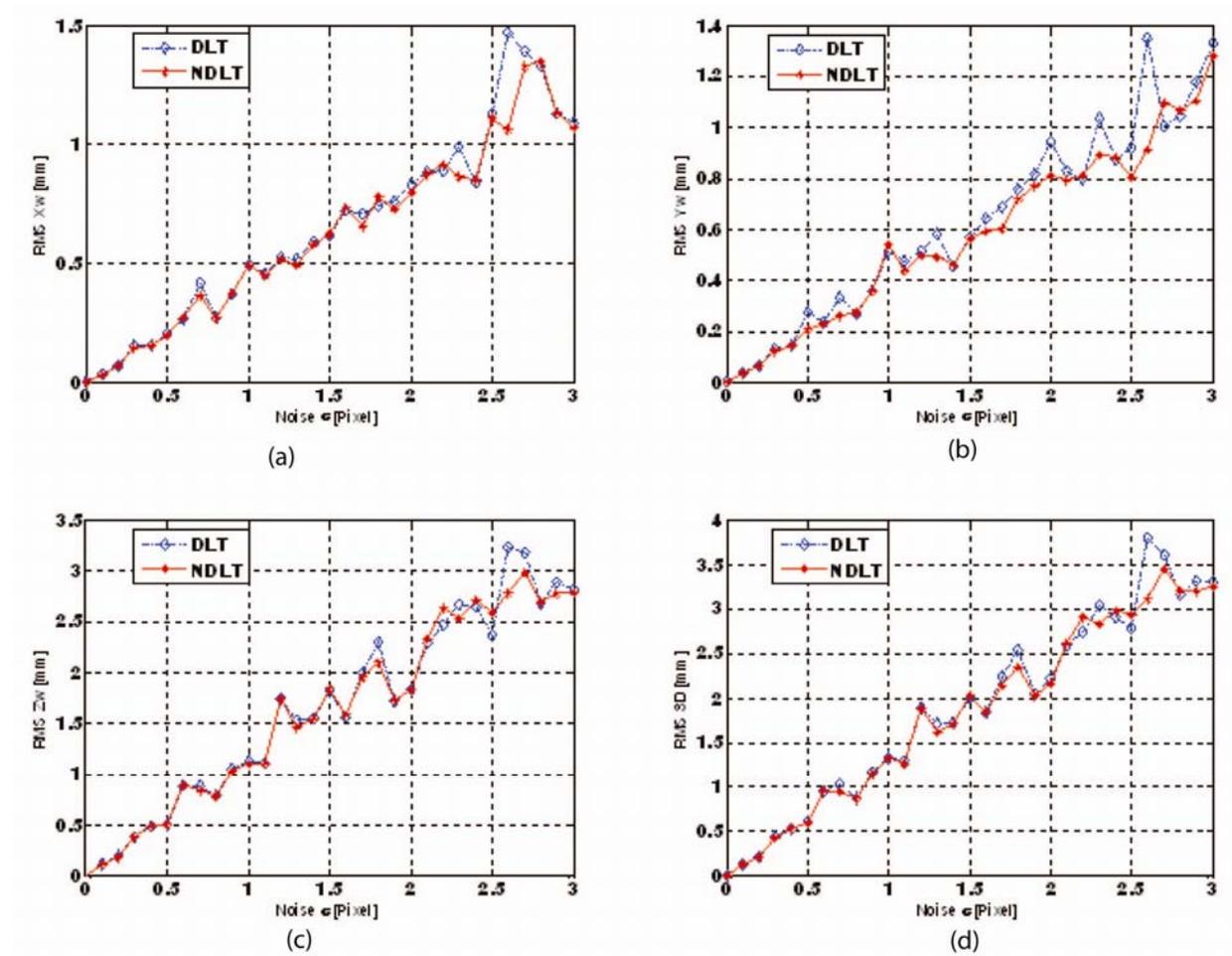
Figure 4.18: (a) RMS errors in X_w; (b) RMS errors in Y_w; (c) RMS errors in Z_w; (d) RMS errors in 3D using the DLT method and NDLT method with noise σ varying from 0 to 3 pixels.

From the results, it is obvious that the DLT method achieves almost the same accuracy as the NDLT method does. This means that NDLT (DLT-based nonlinear optimization using constraints on the matrix **R**) does not improve the final results with distortion-free data. These

101

results confirm that more constraints on the matrix $\boldsymbol{R}$ do not help to improve the accuracy for noisy data at all [44]. So it is concluded that the DLT method is the best solution which combines accuracy and simplicity when the lenses are ideal.

4.5.3 Case III: calibration using synthetic data with noise and distortion

In Case III, synthetic data with noise and lens distortion is used for evaluation of the DLT method and NDLT method. The lens distortion model with 3rd, 5th order radial distortion coefficients (K_1, K_2) and decentering distortion coefficients (P_1, P_2) are employed for this simulation. These coefficients are set to experienced values [50] for wide angle lenses, as shown in Table 4.3. Both left and right cameras are given the same values for these coefficients. The distortion errors caused by this lens distortion on left and right cameras are also shown in this table. It indicates that the maximum distortion error in X (denoted as Max ((Δu))) is 4.775 pixels and in Y (denoted as Max ((Δv))) is 5.027 pixels for the left camera. Similarly, for the right camera, the maximum distortion error in X (denoted as Max ((Δu))) is 7.241 pixels and in Y (denoted as Max ((Δv))) is 5.250 pixels. The mean and minimum distortion errors are also given in this table.

Table 4.3: 2D pixel coordinates difference with certain distortion coefficients.

View side	K_1 [mm^{-2}]	K_2 [mm^{-4}]	P_1 [mm^{-1}]	P_2 [mm^{-1}]	Mean (Δu) [pixel]	Mean (Δv) [pixel]	Max (Δu) [pixel]	Max (Δv) [pixel]	Min (Δu) [pixel]	Min (Δv) [pixel]
Left	-1×10^{-4}	1.2×10^{-7}	2×10^{-5}	-1×10^{-5}	1.0772	1.173	4.775	5.027	-1×10^{-4}	5×10^{-3}
Right	-1×10^{-4}	1.2×10^{-7}	2×10^{-5}	-1×10^{-5}	1.7913	1.1545	7.241	5.250	-1.3×10^{-3}	1.1×10^{-3}

The evaluation procedure for this simulation is similar to the previous evaluation method shown in Figure 4.17, except that the lens distortion model is included in the camera calibration, as shown in Figure 4.19.
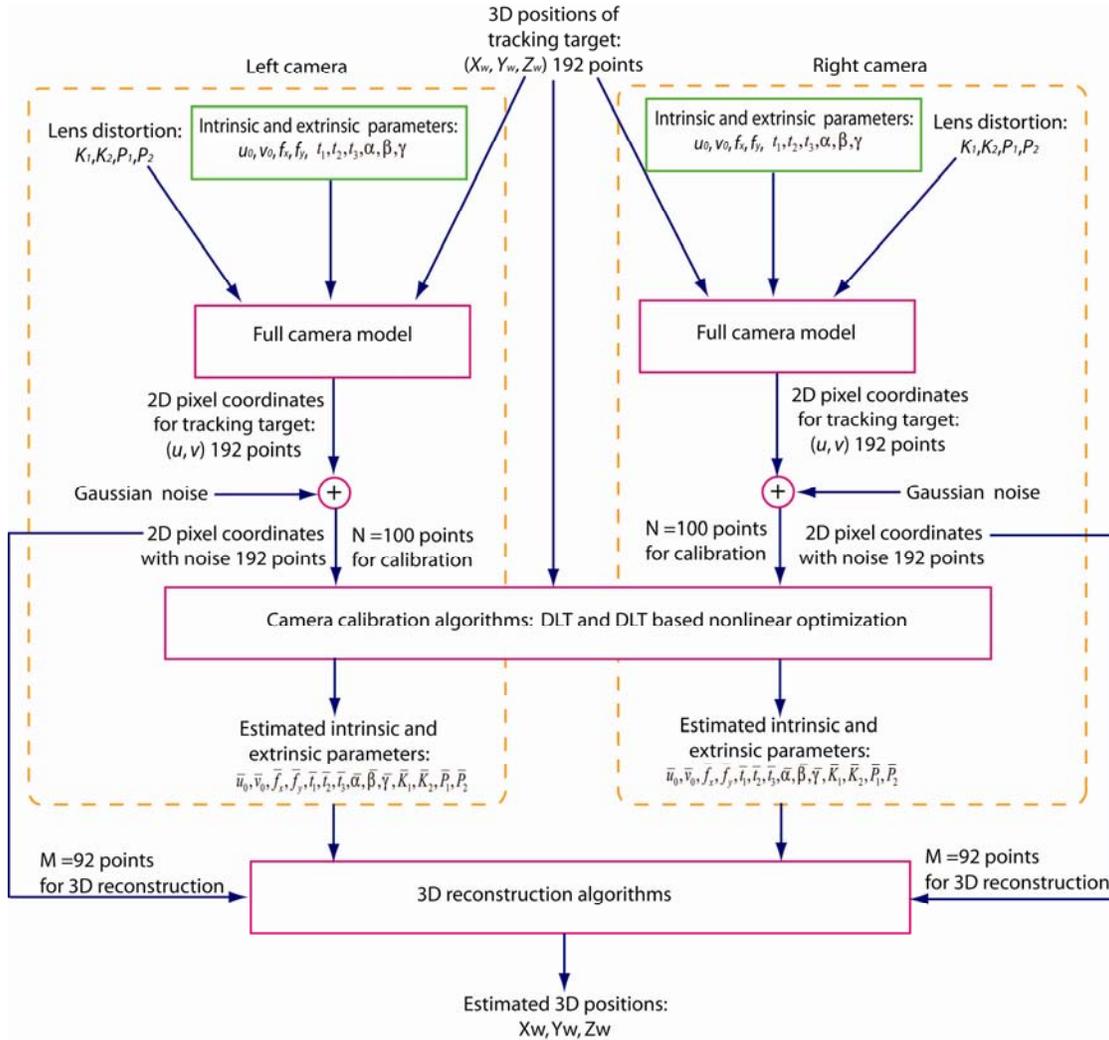
Figure 4.19: Flow chart of 3D position estimation using synthetic data with noise and lens distortion.

The estimation results for the camera parameters using the DLT method and NDLT method for both cameras are shown in Table 4.4 (left camera) and Table 4.5 (right camera).

Table 4.4: Parameter estimation results for the left camera with noise (σ=0.01 pixels).

Parameters	Left Ground_truth	DLT left_esti	NDLT left_esti	DLT_error (%)	NDLT_error (%)
t_1(mm)	100	99.479	99.836	0.521	0.164
t_2(mm)	200	200.074	200.009	0.037	0.005
t_3(mm)	-900	-906.041	-901.429	0.671	0.159
α (deg)	2	1.585	1.999	20.745	0.038
β (deg)	5	5.194	5.011	3.871	0.216
γ (deg)	-179	-178.963	-179.000	0.021	0.000
f_x(scaling)	2141.6667	2148.128	2145.081	0.302	0.159
f_y(scaling)	2141.6667	2148.039	2145.076	0.298	0.159
u_0(pixel)	1024	1031.684	1024.308	0.750	0.030
v_0(pixel)	1024	1039.169	1023.926	1.481	0.007
$K_1[\text{mm}^{-2}]$	-1×10^{-4}	0	-1.01×10^{-4}	100	0.587
$K_2[\text{mm}^{-4}]$	1.2×10^{-7}	0	1.01×10^{-7}	100	15.735
$P_1[\text{mm}^{-1}]$	2×10^{-5}	0	2.09×10^{-5}	100	4.440
$P_2[\text{mm}^{-1}]$	-1×10^{-5}	0	-0.95×10^{-5}	100	5.049

Table 4.5: Parameter estimation results for the right camera with noise (σ=0.01 pixels).

Parameters	Right Ground_truth	DLT right_esti	NDLT right_esti	DLT_Error (%)	NDLT_error (%)
t_1(mm)	700	701.697	699.753	0.242	0.035
t_2(mm)	200	200.000	199.982	0.000	0.009
t_3(mm)	-900	-905.104	-899.889	0.567	0.012
α (deg)	2	1.600	2.001	20.012	0.039
β (deg)	-20	-21.244	-19.996	6.222	0.018
γ (deg)	-179	-179.139	-179.000	0.078	0.000
f_x(scaling)	2141.6667	2150.328	2141.144	0.404	0.024
f_y(scaling)	2141.6667	2150.192	2141.268	0.398	0.019
u_0(pixel)	1024	977.020	1023.698	4.588	0.029
v_0(pixel)	1024	1038.503	1023.946	1.416	0.005
$K_1[\text{mm}^{-2}]$	-0.0001	0	-1.01×10^{-4}	100	0.579
$K_2[\text{mm}^{-4}]$	1.2×10^{-7}	0	1.03×10^{-7}	100	14.142
$P_1[\text{mm}^{-1}]$	2×10^{-5}	0	2.17×10^{-5}	100	8.457
$P_2[\text{mm}^{-1}]$	-1×10^{-5}	0	-0.96×10^{-5}	100	4.073

From the above results, it is obvious that the NDLT method achieves much higher accuracy than the DLT method using synthetic data with lens distortion. Meanwhile, the DLT method cannot handle distortion coefficient estimation for the lenses as discussed before. The 3D estimation results (RMS_X_w, RMS_Y_w, RMS_Z_w, RMS_3D) using these two methods are also shown in Table 4.6.

Table 4.6: 3D position estimation errors using the DLT method and NDLT method.

3D estimation RMS errors	DLT	NDLT
RMS_X_w(mm)	0.158	0.00372
RMS_Y_w(mm)	0.175	0.00333
RMS_Z_w(mm)	0.3556	0.0104
RMS_3D(mm)	0.4267	0.0115

The results indicate that 3D position estimation from the NDLT method is much more accurate than the DLT method when lens distortion is involved. With larger lens distortion, the estimation accuracy using the DLT method is worse, while the estimation accuracy using NDLT method remains the best. For instance, the estimation error using the NDLT method is always close to the noise level. So it is concluded that the proposed NDLT method achieves the best accuracy for camera calibration with lens distortion and noise taken into account.

4.5.4   Effect of the constraints on the matrix R

In this section, the effect of constraints on the matrix R is studied. Section 4.5.2 has already indicated that adding constraints on the matrix R for calibration algorithms does not improve the accuracy with distortion-free data. In reality, it is well known that distortion-free

lenses do not exist. So it is very meaningful to understand whether considering constraints on the matrix R will help improve the accuracy using data that includes lens distortion. The comparison in this section is between the DLT-based nonlinear optimization without considering constraints on the matrix R (denoted as M1) and the proposed solution which is DLT-based nonlinear optimization with considering constraints on the matrix R (NDLT).

The synthetic data used for the performance evaluation is the same as in Section 4.5.3, except that the lens distortion is increased from 1 to 10 times of the distortion applied in 4.5.3 (e.g. $K_1 = -1 \times 10^{-4}$, $K_2 = 1.2 \times 10^{-7}$, $P_1 = 2 \times 10^{-5}$ and $P_2 = -1 \times 10^{-5}$). The lens distortion effect on M1 and NDLT is shown in Figure 4.20 with noise σ=0.01 pixels.
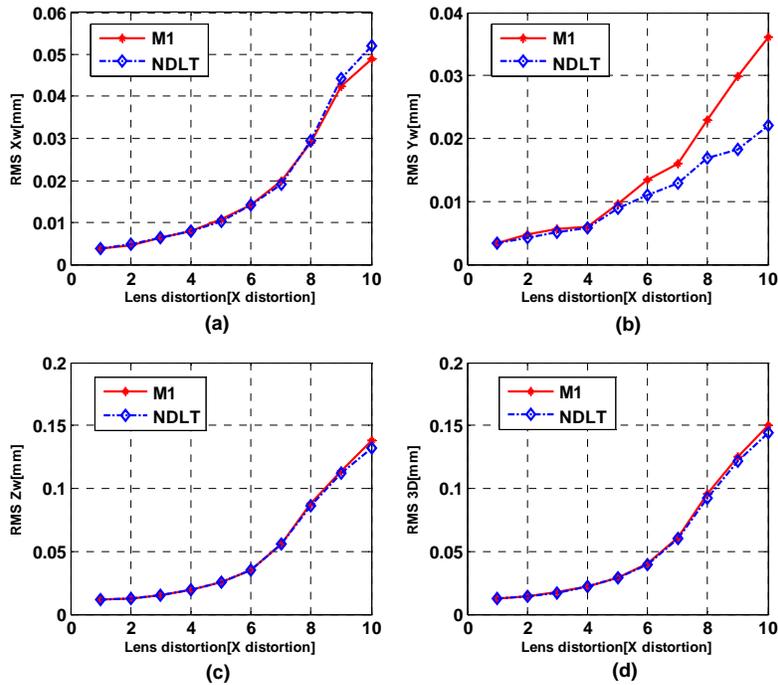
Figure 4.20: (a) RMS errors in X_w; (b) RMS errors in Y_w; (c) RMS errors in Z_w; (d) RMS errors in 3D using the M1 method and NDLT method with noise σ=0.01 pixels.

Based on the results, it is clear that the M1 and NDLT methods have similar performance for 3D estimation when lens distortion is relatively small (distortion error less than 2% of the full image). However, when lens distortion is larger than that, the estimation results using the NDLT method is always better than those using the M1 method as shown above. It is concluded that for large lens distortion applications (distortion error > 2% of the full image), constraints on the matrix R should be considered. In addition, the NDLT method which involves the constraints on the matrix R performs the best among all the camera calibration algorithms.

4.6   Summary

This chapter presented camera calibration algorithms and techniques for the 3D optical tracking system as well as 3D reconstruction algorithms. In Section 4.1, a brief introduction to camera calibration was given. Section 4.2 presented the camera model with definitions of coordinate frames first, and then explained the four steps from 3D world coordinates to digitized pixel coordinates in detail and a full camera model was also provided. Section 4.3 discussed the existing camera calibration methods with advantages and disadvantages of each. Also, the proposed solution for camera calibration was presented, involving DLT algorithms as an initial guess and a nonlinear least square optimization for a full scale optimization. In Section 4.4, the 3D reconstruction algorithms for estimating 3D positions were described. Section 4.5 discussed the simulation investigation for the DLT method and the NDLT method under different data conditions, such as data without noise or distortion, distortion-free data with noise, and data with distortion and noise. The simulation results indicated that the DLT method achieved as good accuracy as the NDLT method for distortion-free data, while the NDLT method achieved the

107

best accuracy for data with distortion and noise. As no lens is distortion-free in reality, it is concluded that the proposed NDLT provides the best accuracy for camera calibration. This section also studied the effect of constraints on the matrix R in camera calibration. It is concluded that constraints on the matrix R should be considered for data with large lens distortion.

Chapter 5

Experimental Results and Discussion

When the 3D optical tracking system has been designed, the implementation and tests for the designed system are presented in this chapter. The IR LED HE8812SG from Opnext, which is the tracking target of the system, is tested and compared with the other two IR LEDs: SE3470 from Honeywell and LN152 from Panasonic. Noise correction algorithms and target location algorithms are implemented to obtain 2D pixel coordinates. The repeatability of the tracking system is tested and improved by reducing possible error sources.

Camera calibration algorithms are carried out and evaluated using real image data in this chapter. With all the intrinsic and extrinsic parameters as well as lens distortion coefficients obtained from camera calibration, 3D reconstruction algorithms are implemented to acquire 3D positions for the tracking target. The performance of the 3D optical tracking system is evaluated based on the results of 3D RMS errors.

This chapter is organized as follows: Section 5.1 describes the system setup for the 3D optical tracking system. Section 5.2 presents the tests on the selected IR LEDs, which involve the performance comparison for three types of IR LEDs. The FPN correction is implemented and the properties of the FPN are studied in Section 5.3. In Section 5.4, the PRNU correction is described. The repeatability tests for target locations are shown in Section 5.5. Section 5.6 introduces evaluation of the 3D optical tracking system. Section 5.7 gives the summary of this chapter.

5.1   System setup

The real setup for the 3D optical tracking system is shown in Figure 5.1. Here, the IR LED acts as the tracking target. A power supply is used to provide required current for the IR LED which is mounted on a CMM. The CMM which has a measuring range of 705 mm in X axis, 605 mm in Y axis, and 1005 mm in Z axis, can achieve several micrometers accuracy for 3D measurements. The IR LED then moves together with the CMM and can locate at known 3D positions using the CMM. The selected CMOS camera is placed on a tripod and is used to form images for the IR LED. The frame grabber is employed to obtain the image data from the camera and send them to a computer. Then the computer provides all the raw images for further image processing and all the algorithms are implemented using MATLAB on the computer.
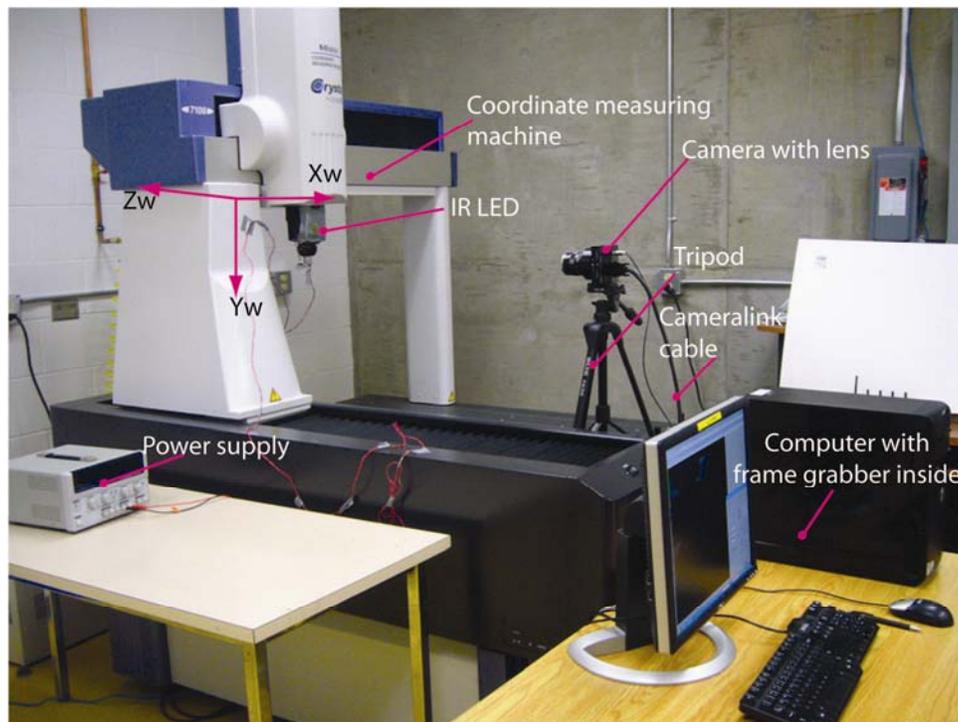
Figure 5.1: Setup for the 3D optical tracking System.

5.2   Tests on the selected IR LEDs

Theoretically, the best IR LED which can satisfy system requirements for better uniformity of light distribution and high power efficiency is the HE8812SG from Opnext. However, the IR LEDs SE3470 from Honeywell and LN152 from Panasonic also have good performance for the uniformity of light distribution with lower cost. From the application perspective, it is necessary to compare the performance of these three IR LEDs and verify the theoretical analysis above. In the following subsections, the names of the companies Opnext, Honeywell and Panasonic are used to represent the name of each IR LED for convenience.

In order to have a direct comparison for these IR LEDs, they are all mounted on one test board and are energized by a power supply. As the maximum current for each IR LED is different (I_{max}= 100 mA for Panasonic and Honeywell, I_{max}= 250 mA for Opnext), the maximum current on the circuit should not exceed 100 mA to prevent damaging of any IR LED. The CMOS camera with LUPA 4000 image sensor and selected lens from Zeiss is used to capture the images for these IR LEDs. Based on the images obtained from the CMOS camera, the performance of the IR LEDs can be compared.

In the following subsection, performance comparison for the IR LEDs based on varying F-number and varying supply current is investigated. However, for varying F-number conditions, all the other parameters need to be fixed. For instance, the current supply is fixed to be 100 mA, and the image format is a 12 bit TIFF file. The exposure time for the image capturing is set to the minimum value of 69 μs at the pixel rate of 30 MHz. Other parameters on the image capturing software are set as default values adjusted by the camera manufacturer to provide the best quality for all the images. A gray image of these IR LEDs and their intensity distribution are shown in Figure 5.2.
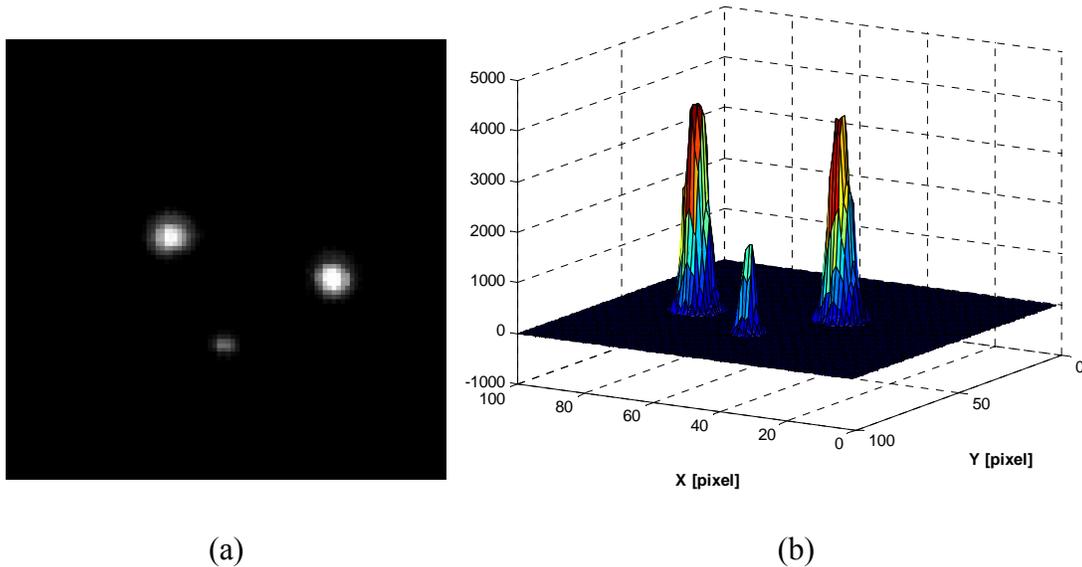
Figure 5.2: (a) Gray image for three IR LEDs; (b) Intensity distribution for these IR LEDs.

- **F-number effect**

The F-number of the lens affects the depth of field for the 3D optical tracking system as larger F-number allows larger depth of field. However, when the F-number increases, the received light on the image sensor from the tracking target will decrease and the SNR of the system will degrade. The F-number effect is to test the intensity variation for these IR LEDs with varying F-number.

The test conditions are as follows: the supply current for the test board I=100 mA, the exposure time t=69 μs, the pixel rate is 30 MHz, the bit depth is 12 bits with TIFF file image format, the other parameters are set to default values provided by the manufacturer and the F-number F of the lens varies from 2.8, 4, 5.6, 8, 11, 16 to 22.

To have a better comparison with the performance of these IR LEDs, the intensity distribution for the IR LEDs with F=2.8 and F=22 are shown in Figure 5.3 and the comparison

results for the peak intensity varying with *F* from 2.8, 4, 5.6, 8, 11, 16 to 22.can be seen in Figure

5.3.
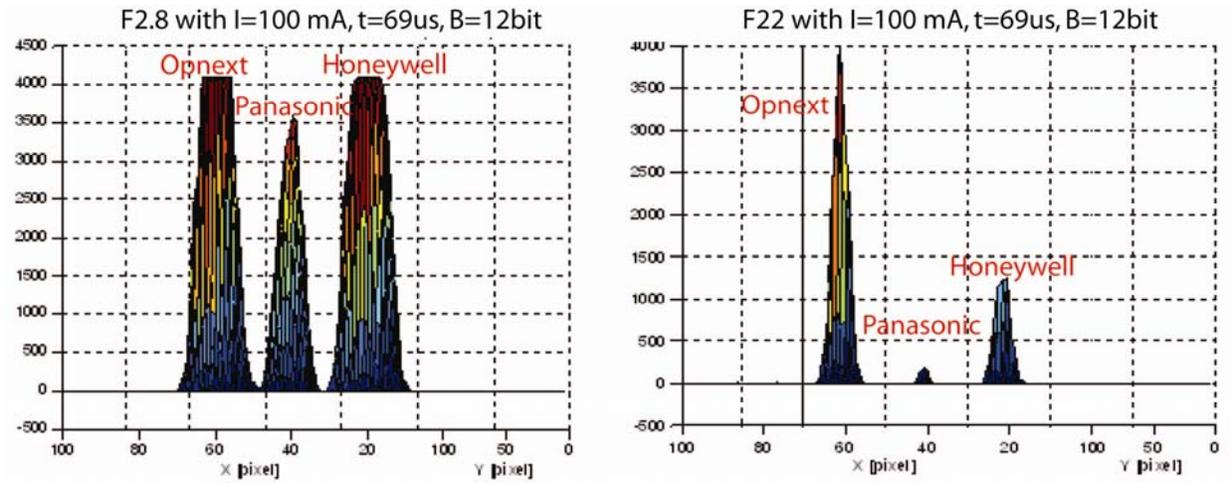
Figure 5.3: Intensity distribution for the IR LEDs with varying F-number.

Table 5.1: Peak intensity of the IR LEDs with F-number varying.

F-number	Peak Intensity (12 bits)		
	Opnext	Honeywell	Panasonic
2.8	4095	4095	3591
4	4095	4095	3481
5.6	4095	4095	2658
8	4095	4054	1760
11	4095	3880	1262
16	4077	3043	731
22	3971	1238	174

Based on the results, it is clear that when F-number increases from 2.8 to 22, the light

intensity of each IR LED reduces accordingly. However, the peak intensity of Opnext does not

change much with F-number and is the highest among the three IR LEDs. Instead, the light

intensity of Panasonic is the lowest and is less than 5% of the saturation intensity when the

aperture size is smallest. There are two reasons for this phenomenon: one is due to the low

sensitivity of the CMOS image sensor to 950 nm wavelength of light and the quantum efficiency (QE) for Panasonic is less than 5%; the other reason is that the optical power output from this type of IR LED is relative low compared to other two IR LEDs based on the theoretical analysis. The light intensity of Honeywell is also very high when the aperture size is large. But its peak intensity decreases dramatically with the aperture size decreasing, only 30% of the saturation intensity with $F=22$. It is concluded that the IR LED from Opnext has the highest peak intensity and is capable of achieving high SNR with large depth of field.

- **Supply current effect**

The supply current effect is to compare the power efficiency of these IR LEDs with varying supply current.

The test conditions are as follows: the F-number $F = 11$, the exposure time $t = 69 \ \mu s$, the pixel rate is 30 MHz, the bit depth $= 12$ bits with image format of TIFF file, the other parameters are set to default values and supply current varies from 100 mA to 10 mA with 10 mA decrement for each test.

The intensity distribution for these IR LEDs with $I = 100$ mA and $I = 10$ mA are shown in Figure 5.4 and the comparison results for the peak intensity varying with supply current I from 100 mA to 10 mA with 10 mA decrement are given in Table 5.2.
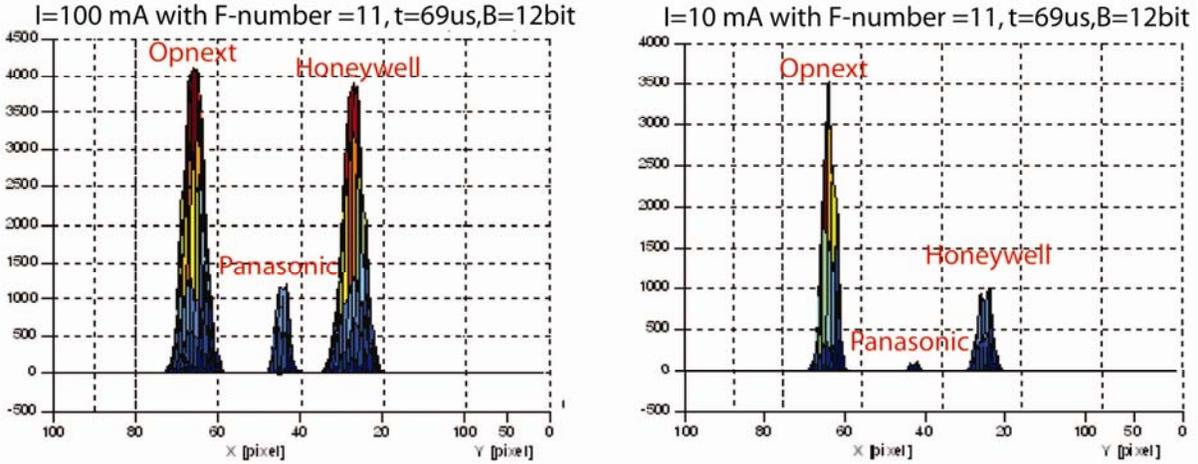
Figure 5.4: Intensity distribution for the IR LEDs with varying supply current.

Table 5.2: Peak intensity for three IR LEDs with varying supply current.

Supply Current	Peak Intensity (12 bits)		
(mA)	Opnext	Honeywell	Panasonic
100	4095	3885	1189
90	4095	3878	1128
80	4095	3832	1059
70	4095	3832	1044
60	4095	3705	945
50	4081	3365	831
40	4069	3141	751
30	4038	2468	440
20	4026	2376	427
10	3522	990	118

Figure 5.4 shows that the peak intensity of Opnext does not change much with supply current and the peak is still high even when I is 10 mA. The intensity of Panasonic is always the lowest and this causes low SNR for the tracking system. So the comparison is mainly focused on Opnext and Honeywell here. It is observed that both Opnext and Honeywell have very high intensity value when the supply current I varies from 100 mA to 50 mA. The slight difference of these two IR LEDs is that Opnext has marginally higher intensity. However, the peak intensity of Honeywell begins to drop after I=50 mA and reduces to 50% of the highest intensity at I=20 mA,

while the peak intensity of Opnext only decreases by 4%. This means that the IR LED from Opnext has high power output on image sensor even with very low current supply, e.g. I=20 mA. If batteries are used to power the LEDs, the IR LED from Opnext will be the best choice for the highest power efficiency. However, the supplied current should not be less than 20 mA as lower current supply (I<20 mA) decreases the performance of the IR LED.

After the comparison tests on these IR LEDs described above, conclusions on the performance of these IR LEDs are given as follows:

a) The peak intensity of the IR LED from Opnext does not change much with the F-number and supply current, while the other two IR LEDs are affected a lot. In addition, the IR LED from Opnext always maintains the highest peak intensity among the three IR LEDs.

b) The IR LED from Opnext has the highest power efficiency compared with the other two.

c) The IR LED from Opnext sustains the largest depth of field with the largest F-number, while the other two have very low intensity with small aperture size.

It is concluded that the IR LED HE8812SG from Opnext performs the best for the optical tracking system with largest depth of field and highest power efficiency. This is also coincident with the previous theoretical analysis.

From the above results, the test conditions for IR LED HE8812SG are also obtained to avoid saturation: the F-number F =11, the supply current I=20 mA, the exposure time t_{exp} = 69 μs , the pixel rate is 30 MHz, the bit depth = 12 bits with image format of TIFF file and the other parameters are set to default values.

5.3   FPN correction

The CMOS image sensors suffer from fixed pattern noise (FPN) and pixel response non-uniformity (PRNU) as well as temporal noise. These noise terms degrade the image quality obtained from the CMOS image sensor and affect the accuracy of target location estimation. In this subsection, the properties of the FPN for the selected CMOS image sensor are investigated, and the FPN correction is implemented for accurate target location estimation.

The black images of the FPN are obtained by covering the lens with no light reaching the CMOS image sensor. In order to reduce the temporal noise effect on the FPN, a mean image of the FPN is acquired by averaging 10 black images and applied for the FPN correction. A typical black image of the FPN with exposure time of 69 μs at 30 MHz pixel rate is shown in Figure 5.5.
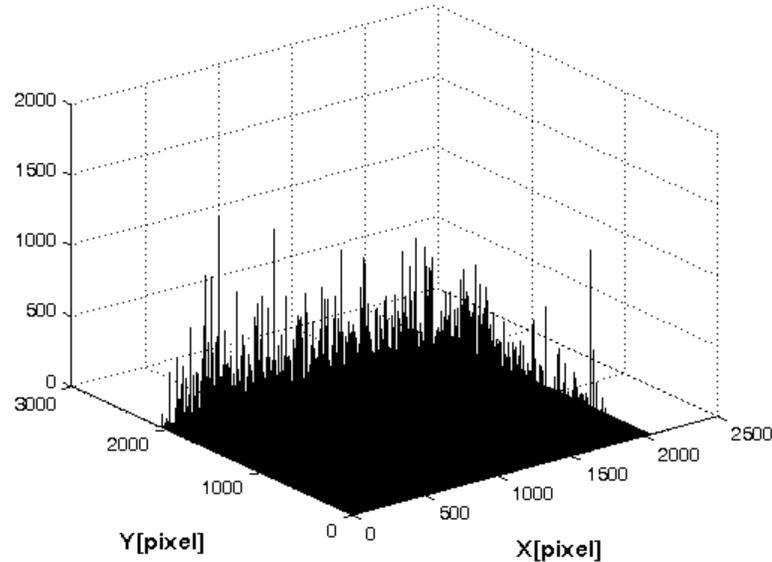
Figure 5.5: A black image of FPN with exposure time of $69\,\mu s$: mean intensity is 0.14 for 12 bits output, and the RMS intensity is 6.4. There are 9 pixels with intensity over 25% of the saturation intensity and 248 pixels with intensity over 10% of the saturation intensity.

The result shows that for the selected CMOS camera, the FPN of some pixels is very large while the FPN of most pixels are zero. In the above figure, there are 9 pixels with intensity over 1000 out of 4095 saturation intensity (25%) on this image. For intensity over 400 which is 10% of the saturation intensity, there are 248 pixels on the image. This is why this black image looks so noisy with many peaks. This large FPN peaks are resulting from the large dark current on these pixels due to sensor manufacturing errors. As the image has 4 mega pixels, those peaks with high intensity do not affect much on the mean intensity of this image, about 0.14 and the RMS of this image is also not large, about 6.4.

However, as the tracking target on the image is a small dot with image area approximately 5x5 to 10x10 pixels, these huge peaks will have great impact on the accuracy of the target locations if these noise peaks are within the ROC. This also emphasizes on the importance of the FPN correction.

- **FPN varying with sensor temperature**

During the investigation process for the FPN, it is observed that the FPN is not fixed but varies with sensor temperature and exposure time. Since the sensor temperature is difficult to be measured, the sensor temperature effect on the FPN can be represented by the FPN variation during sensor warm-up period. The results of mean intensity, RMS intensity and worst case intensity variation of the FPN during the sensor warm-up period are illustrated in Figure 5.6.

As shown in Figure 5.6, the mean and RMS intensity of the FPN are increasing during the sensor warm-up period and then remain to be stable after about 2 hours. This also indicates that it takes about 2 hours for the FPN to be stable after turning on the camera. It is noted that the values of the mean and RMS intensity for the FPN are very small over the full frame, less than 0.0036% for the mean intensity and less than 0.16% for the RMS intensity. However, in the

worst case, the single pixel intensity of the FPN varies a lot during the sensor warm-up period, about 2.5% of the saturation intensity.
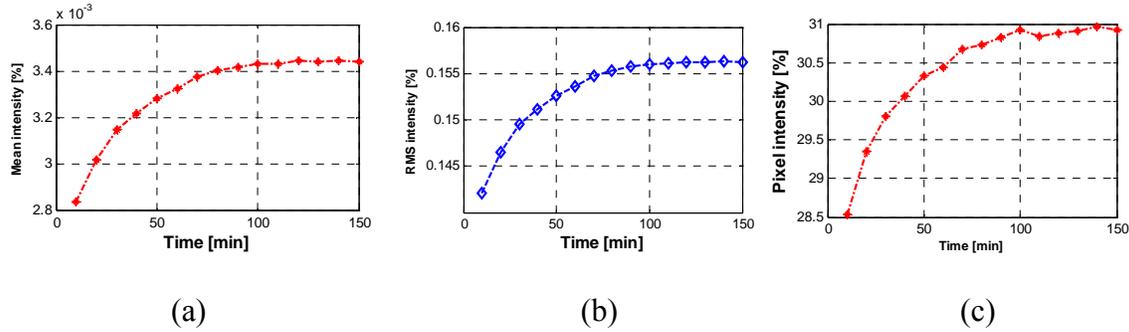
Figure 5.6: (a) Mean intensity for the FPN varying during sensor warm-up; (b) RMS intensity of the FPN during sensor warm-up; (c) The worst case for the FPN varying during sensor warm-up.

This large variation will degrade the accuracy of target location estimation dramatically if it is not considered for the FPN correction. So the best way to correct the FPN with this CMOS camera is to capture 10 black images of the FPN after 2 hours and subtract the averaged black image from the raw image of the tracking target. In this case, the sensor temperature effect on the FPN can be avoided and the accuracy of target location could also be improved. This large variation of the FPN during sensor warm-up period is due to the camera design problem.

- **FPN varying with exposure time t_{exp}**

Another observation for the FPN is that the FPN varies with the exposure time. Black images of the FPN with different exposure time are obtained to verify this observation. Exposure time of 69 μs, 1 ms, 10 ms and 100 ms at 30 MHz pixel rate are used for the verification. The results of mean and RMS intensity of the FPN varying with the exposure time can be seen in Figure 5.7. From the results, it is obvious that the FPN varies a lot with exposure time. When the exposure time increases from 69 μs to 1 ms, the variation of the FPN is relatively small. Once

the exposure time is larger than 10 ms, both mean and RMS intensity of the FPN increases dramatically with the exposure time.

According to the effect of exposure time on the FPN variation, short exposure time should be applied to avoid large FPN. Additionally, attention should be paid to the FPN correction. The black images of the FPN should be taken at the same exposure time as the target images are captured. Otherwise, if the FPN correction uses black images with a different exposure time, the target location estimation of the tracking target will be less accurate.

(a)   (b)
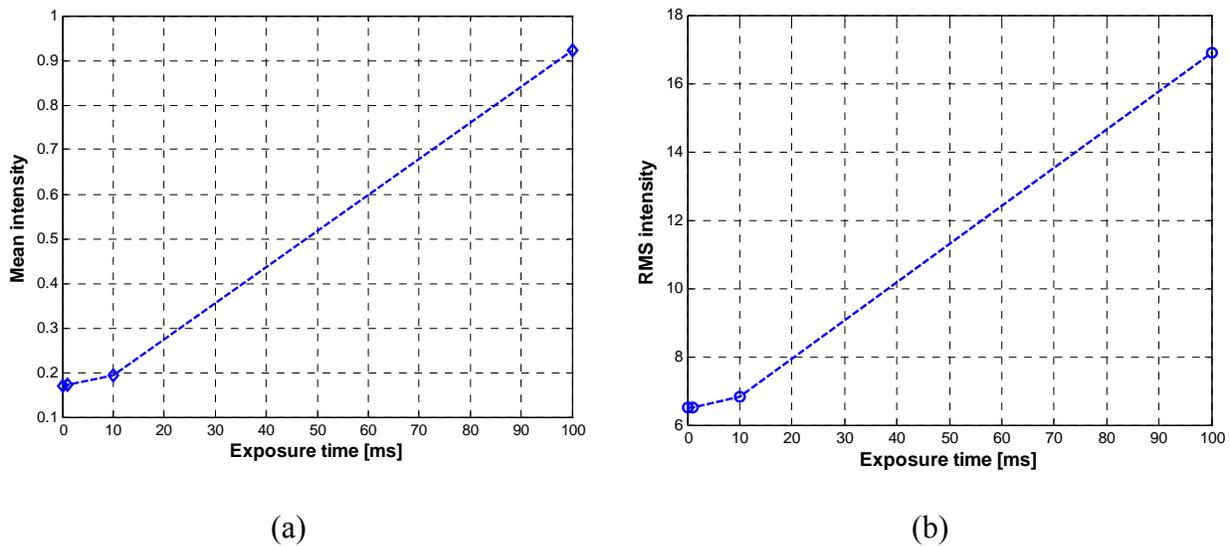
Figure 5.7: (a) Mean intensity of the FPN on the full frame varying with the exposure time; (b) RMS intensity of the FPN on the full frame varying with the exposure time.

Based on all the tests on the FPN for the CMOS image sensor, it is concluded that the minimum exposure time should be used (69 μs for the selected CMOS camera and this exposure time is limited by the camera design from Epix). Additionally, black images with the same exposure time as the target images should be employed for the FPN correction. Moreover, due to the variation of the FPN with the sensor temperature caused by the current camera design, all the

experiments should be done after the sensor warm-up period (about 2 hours after the camera is on). This can be avoided by a better camera design with sensor temperature control.

5.4 PRNU correction

In this subsection, the PRNU correction is discussed. Before the implementation of the PRNU correction, grey images with intensity between 70% and 100% of saturation intensity using uniform illumination should be obtained. The FPN correction should be implemented first on these gray images. Then the pixel gain for each pixel on the full image can be calculated using corrected gray images. To get relatively uniform illumination, the room light passing through white papers (acting as frosted glass) is used. The real setup for obtaining uniform illumination using white paper is illustrates in Figure 5.8.
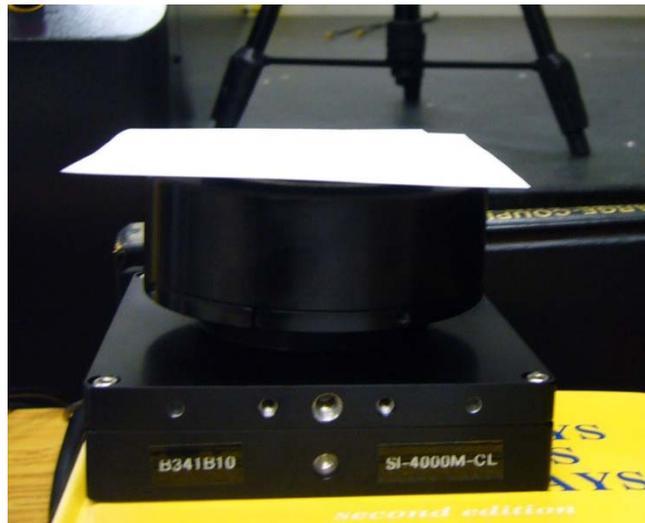
Figure 5.8: Setup for uniform illumination using white paper.

It is known that light passing through the white paper is not absolutely uniform and more pieces of white paper overlapping may improve the uniformity. In order to understand how much impact the number of the white paper has on the uniformity of illumination, gray images with 1 piece of white paper and 2 pieces of white paper are used for comparison. The exposure time for the gray images is set to 10 ms for 1 piece of paper and 20 ms for 2 pieces of paper as the illumination from 2 pieces of paper is much darker than that from 1 piece of paper. Before the calculation of the pixel gains, 10 black images of the FPN with exposure time of 10 ms and 20 ms are obtained respectively for the FPN correction. The calculation for normalized pixel gains obtained using 1 piece of paper and 2 pieces of paper can be expressed as

$$
\begin{aligned}
G_{1P} &= (\overline{I}_{1p} - \overline{I}_{blk}) ./ mean(mean(\overline{I}_{1p} - \overline{I}_{blk})) \\
G_{2P} &= (\overline{I}_{2p} - \overline{I}_{blk}) ./ mean(mean(\overline{I}_{2p} - \overline{I}_{blk}))
\end{aligned}
\tag{5.1}
$$

Based on Eqn. (5.1), the pixel gain of each pixel using 1 piece of paper and 2 pieces of paper are calculated and illustrated in Figure 5.9. The results show that the maximum value of the normalized pixel gain for 1 piece of paper is 1.376 and the minimum value is 0.541, and the maximum value for 2 pieces of paper is 1.335 and the minimum value is 0.552. Both plots for the pixel gain of the full image indicate that the pixel gains of the image sensor are quite different from pixel to pixel and increase a lot from the first row to the last row along X axis.
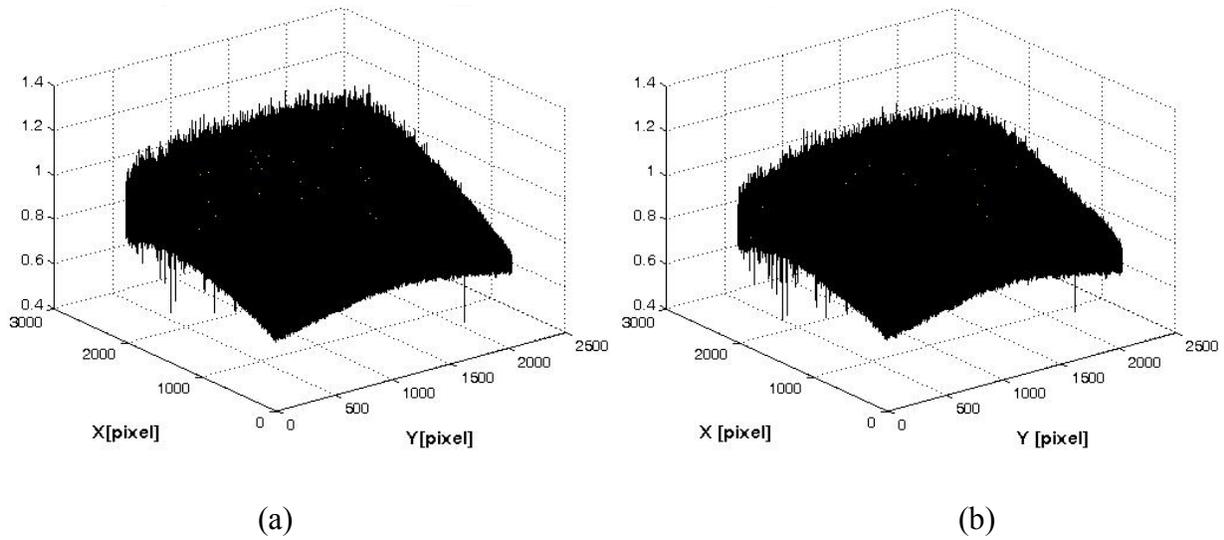
Figure 5.9: (a) Normalized pixel gains using illumination of 1 piece of paper: the maximum value is 1.376 and minimum value is 0.541; b) Normalized pixel gain using illumination of 2 pieces of paper: the maximum value is 1.335 and minimum value is 0.551.

The difference between pixel gains using 1 piece of paper and 2 pieces of paper is shown in Figure 5.10. The maximum value of G1p/ G2p is 1.11 and the minimum is 0.80. This figure demonstrates the illumination uniformity of 1 piece of paper and 2 pieces of paper are slightly different. This means, the illumination of 1 piece of paper is already very uniform and more pieces do not have much improvement on the uniformity. So it is concluded that the illumination of 1 piece of paper is good enough for the PRNU correction. This will also be verified in Chapter 5 using the 3D position estimation results.
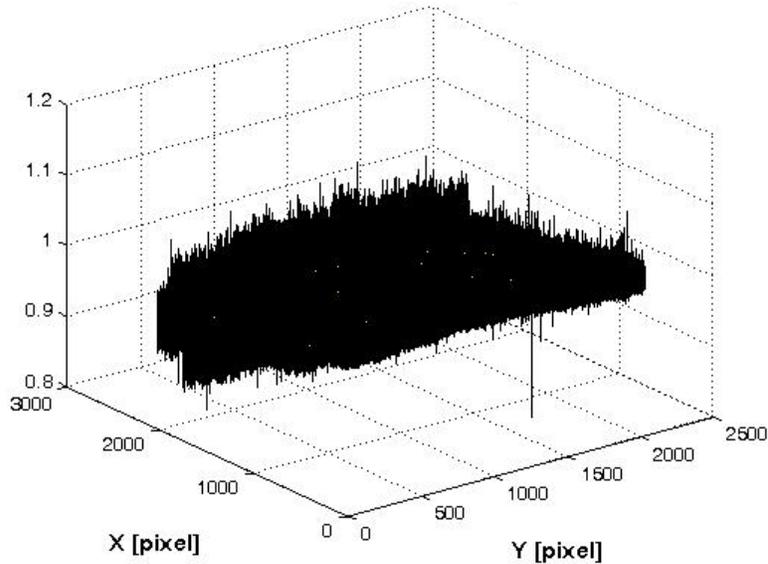
Figure 5.10: The difference between pixel gains using 1 piece of paper and 2 pieces of paper: the maximum value is 1.11 and the minimum value is 0.80.

5.5 Repeatability tests for target locations

The repeatability tests for target locations are to evaluate the repeatability of the target locations based on the system setup, noise correction algorithms and target location algorithms. These tests are done with both camera and the selected IR LED stationary and taking 10 images of the IR LED continuously within 1 minute. Noise correction algorithms are first implemented on these images and then target location algorithms are applied to estimate the 2D target locations. The repeatability of the 2D target locations illustrates that the limitation for the achievable accuracy by the current system.

The test conditions for the repeatability are as follows: supply current for the IR LED I is 20 mA, the exposure time t is $69\,\mu s$, the pixel rate is 30 MHz, F-number F=11, ADC and column

offsets are set to default values with 12 bit TIFF file output and the ROC for target locations is 50x50 pixels. Here, grayscale centroid method is used.

The repeatability results within 1 minute are shown in Figure 5.11. It shows that the repeatability in X axis is about 0.02 pixels and the repeatability in Y axis about 0.01 pixels.
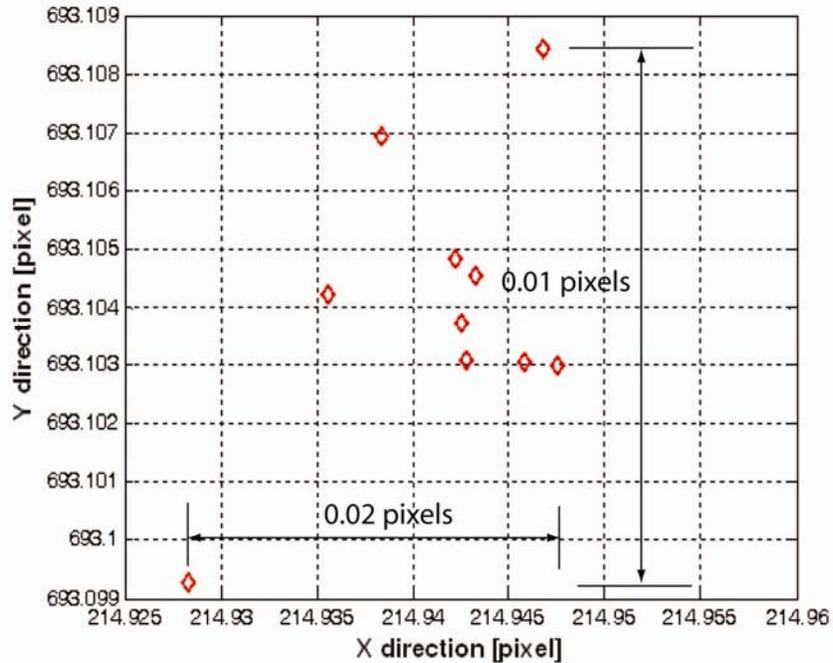
Figure 5.11: Repeatability results using test conditions of I=20 mA, F =11, 12bit data, t_{exp}=69 μs and the ROC is 50x50 pixels.

More tests have been done by varying F-number, exposure time, pixel clock, connected cable length, ADC phase, column offset, and data format. The results show that these factors have no impact on the repeatability results at all. Other factors such as non-rigidity of the camera mounting on the tripod, non-rigidity of the IR LED mounting on the CMM, vibrations from the ground, and temperature variation of the image sensor can affect the repeatability results. Due to limited time, the effects of these factors are not studied in this thesis.

The repeatability results shown in Figure 5.11 do not involve the effect of temperature drift. In the real case, it normally takes several hours to obtain images for the tracking target at 192 different positions for the experiments. So it is necessary to check the repeatability of target locations within a longer time period. Here, the repeatability tests within 2 hours and 35 minutes are done and the results are shown in Figure 5.12. The total drift over this period is 0.15 pixels in X axis and 0.16 pixels in Y axis within 2 hours and 35 minutes.
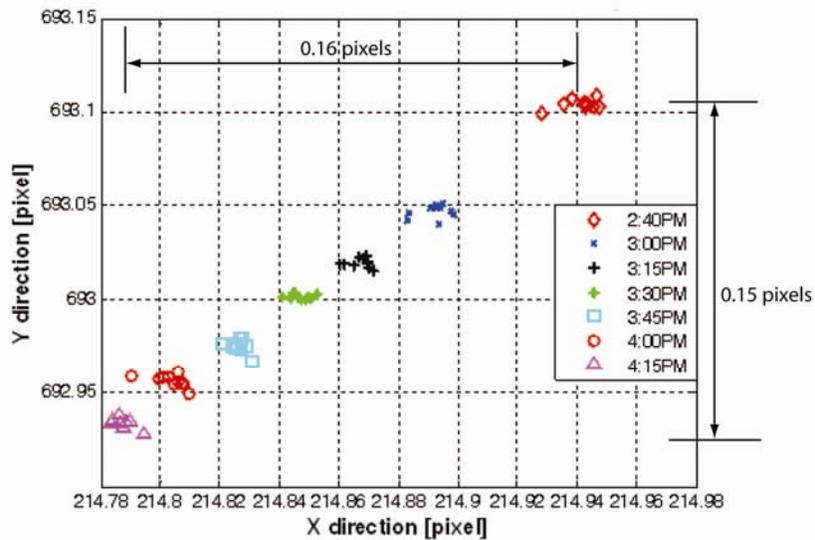
Figure 5.12: Repeatability tests within 2 hours and 35 minutes using the test conditions of I=20 mA, F =11, 12bit data, t_{exp}=69 μs and the ROC is 50x50 pixels.

The results indicates that the 2D location of the IR LED is drifting all the time due to the temperature drift of the CMOS image sensor or some other error sources, which need to be investigated more in the future work.

5.6   Evaluation of the 3D optical tracking system

In this section, the evaluation of the 3D optical tracking system based on all the algorithms is presented. 3D positions of the tracking target can be estimated using two images from two different cameras or from one camera viewing at two different locations. As this thesis mainly focuses on achieving high accuracy for the 3D optical tracking system, one camera viewing from two different locations is applied to capture two images for the tracking target, as shown in Figure 5.13.

(a)   (b)

Figure 5.13: (a) Camera setup for the left view; (b) Camera setup for the right view.

For the implementation of camera calibration and 3D reconstruction algorithms, 192 different 3D positions of the tracking target are required as discussed in Chapter 4. These 192 positions are located on 3 parallel planes with 5 mm distance between each plane. On each plane, there are 64 points with 8 points along X axis and 8 points along Y axis and the distance between

each point is 60 mm on both X and Y axis, as shown in Figure 5.14. Here, 100 points are used

for camera calibration and the remaining 92 points are used for the 3D reconstruction.
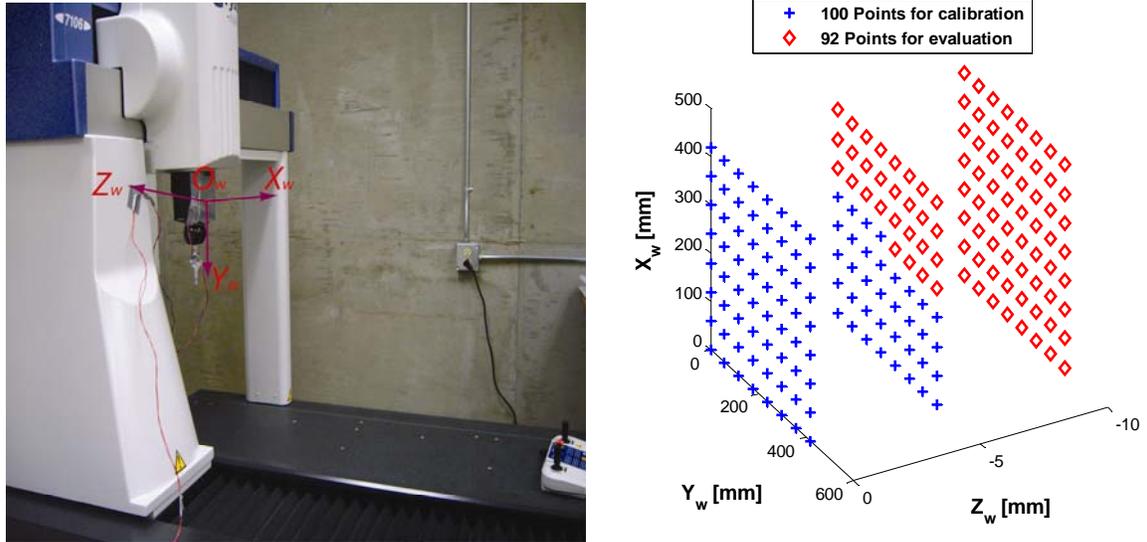
Figure 5.14: 3D positions of the tracking target in a 3D world coordinate.

Based on the experimental setup as well as designed algorithms, 3D estimation errors are

calculated to evaluation the performance of the designed 3D optical tracking system, as shown in

Figure 5.15.
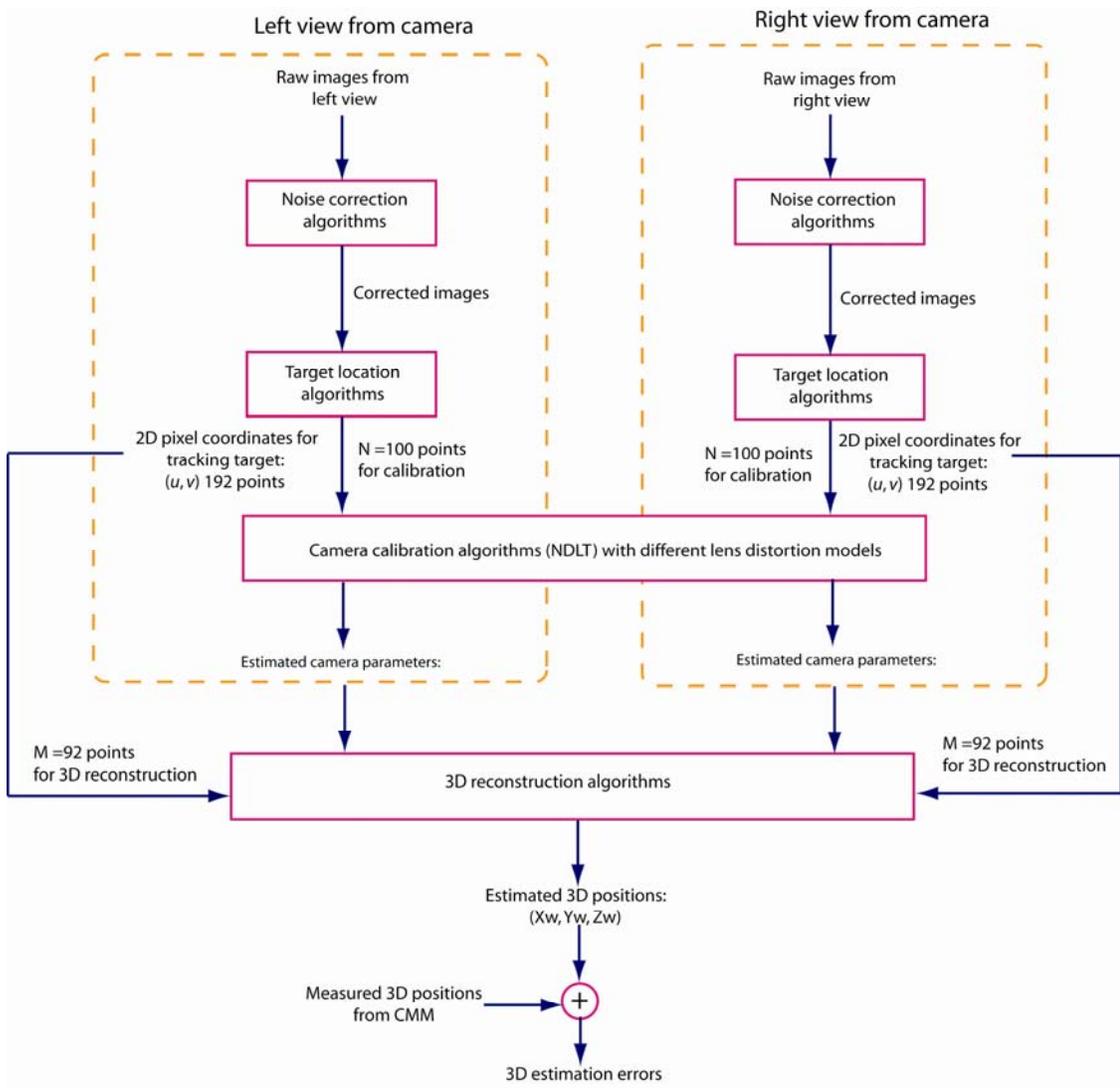
Figure 5.15: Flow chart for evaluation of the 3D optical tracking system.

In this section, three target location algorithms (2D Gaussian fitting, grayscale centroid and squared grayscale centroid) are evaluated using real image data. Additionally, in the camera calibration simulation, the 5th order radial distortion and decentering distortion are considered. But in reality, the lens distortion model for the selected lens is unknown. So the lens distortion coefficients need to be identified and which coefficient weights more in the lens distortion model should be determined. Furthermore, the effect of the PRNU correction with different illumination

sources (e.g. 1 piece of paper and 2 pieces of paper) on the final accuracy of the 3D position estimation is also studied in this subsection. For better comparison, the 3D estimation errors without PRNU correction are also given.

To compare the accuracy of using different lens distortion models for the tracking system, DLT method without lens distortion models and NDLT method with three different lens distortion models are considered:

a). No distortion (denoted as DLT)

b). NDLT with 3rd and 5th order radial distortion (denoted as YL's lens model)

c). NDLT with 3rd, 5th order radial + decentering distortion (denoted as Heikkila's lens model)

d). NDLT with 3rd order radial + decentering + thin prism distortion (denoted as Weng's model).

Using the combination of specific target location algorithms and NDLT calibration method with specific lens distortion model above, the 3D RMS errors of the 3D position estimation are calculated and shown in Table 5.3. These results are obtained without PRNU correction. Note that the FPN correction is applied to all the experimental data.

Table 5.3: 3D RMS errors using different target location methods and lens distortion models (no PRNU correction).

Lens distortion models / Target location methods	DLT (mm)	NDLT with YL's model (mm)	NDLT with Heikkila's model (mm)	NDLT with Weng's model (mm)
2D Gaussian fitting	0.7530	0.0471	0.0596	0.1152
Grayscale centroid	0.7530	**0.0468**	0.0576	0.1189
Squared grayscale centroid	0.9238	0.4804	0.5225	0.5206

The results show that 2D Gaussian fitting method and grayscale centroid method achieve much higher accuracy than squared grayscale centroid method, about 10 times better with lens distortion model of (K_1K_2). Additionally, the grayscale centroid method is as accurate as the 2D Gaussian fitting method and its calculation is much faster than 2D Gaussian fitting method with low computation cost. Therefore, the grayscale centroid method is the best for the 3D optical tracking system which combines both simplicity and high accuracy. However, this conclusion based on the experimental results is quite different from the conclusion based on the simulation results. It may be caused by the simplification of the simulation model and the real case is more complicated than the simulation case. There are still some unknown factors affecting the accuracy of target locations that are not included in the simulation and this investigation is included in the future work.

Another comparison is the lens distortion models. Here, only the row of grayscale centroid method with different lens distortion models is observed. It is very clear that lens distortion model of (K_1K_2) with 5th order radial distortion provides the best accuracy for the 3D position estimation, about 47 μm for the 3D RMS error. When no lens distortion is applied, the 3D RMS error is the worst for all the cases no matter which target location algorithms is used. Compared with the 3D RMS error resulting from Weng's lens distortion model, the accuracy using YL's lens distortion model is 2 times better (3D RMS error reduces from 119 μm to 47 μm). This means the fifth order radial distortion component is very important for the selected lens. The comparison between the results using YL's lens model and Heikkila's lens model show that decentering distortion is not important for the selected lens and more parameters degrade the estimation accuracy (3D RMS error increases from 47 μm to 58 μm). It is concluded that the

best lens distortion model for the selected lens is the YL's model with 3rd order and 5th order radial lens coefficients. This conclusion can also be applied to other wide angle lenses.

Lastly, the PRNU correction effect describes the importance of PRNU correction by using image data without PRNU correction, with PRNU correction using 1 piece of paper and with PRNU correction using 2 pieces of paper. The results are shown in Table 5.3 (without PRNU correction), Table 5.4 (with PRNU correction using 1 piece of paper) and Table 5.5 (with PRNU correction using 2 pieces of paper) as follows.

Table 5.4: 3D RMS errors using different target location methods and lens distortion models (with PRNU correction using 1 piece of paper for illumination).

Lens distortion models / Target location methods	DLT (mm)	NDLT with YL's model (mm)	NDLT with Heikkila's model (mm)	NDLT with Weng's model (mm)
2D Gaussian fitting	0.7532	0.0417	0.0530	0.1134
Grayscale centroid	0.7530	**0.0414**	0.0523	0.1167
Squared grayscale centroid	1.0086	0.5540	0.6559	0.6644

Table 5.5: 3D RMS errors using different target location methods and lens distortion models (with PRNU correction using 2 pieces of paper for illumination).

Lens distortion models / Target location methods	DLT (mm)	NDLT with YL's model (mm)	NDLT with Heikkila's model (mm)	NDLT with Weng's model (mm)
2D Gaussian fitting	0.7516	0.0416	0.0523	0.1121
Grayscale centroid	0.7521	**0.0415**	0.0519	0.1159
Squared grayscale centroid	1.0299	0.5478	0.6483	0.6707

The results show that the PRNU correction is important for improving the accuracy of the 3D position estimation. Without the PRNU correction, the 3D RMS error is 47 μm using YL's lens model and grayscale centroid method. And with the PRNU correction using 1 piece of paper, the 3D RMS error is 41 μm, which is about 13% improvement on the accuracy. However, there is no difference for the results using 1 piece of paper and 2 pieces of paper. This confirms the previous prediction that the illumination using 1 piece of paper is very uniform and 2 pieces do not improve the results.

Based on all these experimental results, it is concluded that a high accuracy 3D optical tracking system is achieved, which has less than 50 μm 3D RMS error. The lens distortion model with radial distortion coefficients K_1 and K_2 is the best fit for the selected lens and more parameters degrade the accuracy for the 3D tracking. Also, by considering accuracy and simplicity, the grayscale centroid method is the best method for target locations. Furthermore, the PRNU correction is very necessary for high accuracy tracking and the illumination of 1 piece of paper can be used.

5.7   Summary

This chapter presented the experimental results for the 3D optical tracking system, including IR LEDs comparison, FPN correction, PRNU correction, repeatability for target location, and 3D position estimation results.

Section 5.1 described the system setup for the 3D optical tracking system. Section 5.2 presented the tests on the selected IR LEDs, including three types of infrared LEDs comparison and identification of the test conditions for the selected IR LED. In Section 5.3, the properties of

133

the FPN and the FPN correction were investigated. The setup for the PRNU correction and the pixel gain calculation were introduced in Section 5.4. Section 5.5 presented the repeatability tests for target locations. In Section 5.6, the performance evaluation for the 3D optical tracking system was described and different target locations algorithms as well as lens distortion models were implemented on the 3D tracking accuracy.

It was concluded that a high accuracy 3D optical tracking system (less than 50 μm 3D RMS error) was achieved based on the combination of FPN correction, PRNU correction with white paper, grayscale centroid method for target location, and lens distortion model using radial distortion coefficients K_1 and K_2.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Based on the research work during these two years, several conclusions are drawn as follows:

1. Designed and implemented a high accuracy CMOS camera based 3D optical tracking system. The 3D RMS error achieved for the 3D position estimation is 41 μm for a measuring range of 500 m by 500 m on XY plane and 10 mm in Z.

2. Investigated into three target location algorithms and compared their performance under different conditions. It is concluded that grayscale centroid method is the best target location algorithm for the 3D optical tracking system.

3. Investigated into different camera calibration algorithms and lens distortion models. It was concluded that the proposed NDLT method provides the best accuracy for the 3D optical tracking system with lens distortion. The constraints for the matrix R should be considered in camera calibration for large lens distortion cases but have no influence on the calibration accuracy with distortion free data. The lens distortion model using radial distortion coefficients K_1 and K_2 only best fits the selected lens distortion. This model can also be applied to other similar wide angle lenses.

6.2 Future work

Here are some suggestions for the future work:

a). High speed real time tracking

As discussed in Chapter 2, the selected CMOS image sensor LUPA 4000 has the ability of random windowing and can achieve several kHz frame rate with small ROI. However, there is no camera available in the world that has achieved random windowing for the LUPA 4000. In order to fully exert the ability of high frame rate for this CMOS sensor, special electronics should be designed. The designed CMOS camera can be used for high speed real time tracking and the sampling rate of 10 kHz with ROI of 25 x 25 pixels is expected.

b). Large volume evaluation

In this thesis, the achieved accuracy for the 3D tracking (3D RMS error of 41 μm) only covers 500 mm by 500 mm in XY and 10 mm in depth. So the tracking accuracy within a larger volume should be investigated to evaluate the performance of the designed system.

c). Different types of lenses investigation

The conclusion for the lens distortion model is based on the selected lens with focal length f =25 mm. Other types of lenses should be studied to verify this conclusion and extend the model to all types of lenses.

d). Better design for system setup and camera electronics

For the current setup of the tracking system, the camera was placed on the tripod. A more rigid support with less vibration should be designed for the camera. Besides, a better design is

also needed for locating the IR LED. Moreover, a better camera design with less FPN and PRNU will also improve the accuracy of the system a lot.

e). Further investigation on target location algorithms

As discussed in Chapter 5, the conclusions based on the simulation and the experimental results are quite different. This means, there are still some unknown factors not considered in the simulation model. More research work should be done to figure out these unknown factors.

References

[1]   J.C. Compter, "Electro-dynamic planar motor," *Precision Engineering,* vol. 28, pp. 171-180, 2004.

[2]   CD Lovell-Smith, "A Prototype Optical Tracking System: Investigation and Development," Master Thesis, University of Canterbury, pp. 1-8, 2009.

[3]   H Mathieu, "The Cyclope: A 6 DOF Optical Tracker Based on a Single Camera," 2005, "http://perception.inrialpes.fr/Publications/2005/Mat05/TheCyclopeTracker.pdf".

[4]   DK Bhatnagar, "Position Trackers for Head Mounted Display Systems: A survey," *University of North Carolina, Chapel Hill TR93-010,* 1993.

[5]   RJ Valkenburg, DW Penman, JA Schoonees, NS Alwesh, and GT Palmer, "Interactive Hand-held 3D Scanning," *Proceedings of Image & Vision Computing New Zealand*, pp. 245–250, 2006.

[6]   R Khadem, CC Yeh, M Sadeghi-Tehrani, MR Bax, JA Johnson, JN Welch, EP Wilkinson, and R Shahidi, "Comparative tracking error analysis of five different optical tracking systems," *Computer Aided Surgery,* vol. 5, pp. 98-107, 2000.

[7]   AD Wiles, DG Thompson, and DD Frantz, "Accuracy assessment and interpretation for optical tracking systems," *Proceedings of SPIE*, pp. 421–432, 2004.

[8]   M Ribo, "State of the art report on optical tracking," *Vienna Univ. Technol., Vienna, Austria, Tech. Rep,* vol. 25, pp. 2-8, 2001.

[9]   DW Denning, "State-of-the-art clinical article," *Clinical infectious diseases,* vol. 26, pp. 781-803, 1998.

[10]   E Foxlin, "Motion tracking requirements and technologies," *Handbook of virtual environment technology,* pp. 163–210, 2002.

[11]   Optical Metrology Center, "OMC Technical Brief - LaserTracker," 2001, "http://www.optical-metrology-centre.com/Downloads/Tech_Briefs/TechBrief_LaserTracker.pdf".

[12]   Northern Digital Inc., "Laser Tracker Technology," 2010, "http://www.ndigital.com/industrial/lasertracker.php".

[13]   Leica, "Leica Laser Tracker System," 1999, "http://www.a-solution.com.au/pages/downloads/LTD500_Brochure_EN.pdf".

[14]   G Welch, G Bishop, L Vicci, S Brumback, K Keller, and D Colucci, "High-performance wide-area optical tracking: The HiBall tracking system," *Presence: Teleoperators & Virtual Environments,* vol. 10, pp. 1-21, 2001.

[15]   3rd Tech, "HiBall-3100™ Wide-Area, High-Precision Tracker and 3D Digitizer," 2006, "http://www.3rdtech.com/HiBall.htm".

[16]   A.R.T. GmbH, "Optical Tracking System," "http://www.ar-tracking.de/System-overview.20.0.html".

[17]   V Macellari, "CoSTEL: a computer peripheral remote sensing device for 3-dimensional monitoring of human motion," *Medical and Biological Engineering and Computing,* vol. 21, pp. 311-318, 1983.

[18]   Rafel C. Gonzalez and Richard E. Woods, "Digital image processing*"*: Prentice-Hall, Inc., pp. 42-44, 2002.

[19]   Gerald C Holst, "Radiometry and Photometry," in *CCD Arrays Cameras and Displays*: SPIE Optical Engineering Press, pp. 33-35,117, 1998.

[20]   D. Litwiller, "CMOS vs. CCD: Maturing technologies, maturing markets," *Photonics Spectra,* vol. 39, p. 54, Aug 2005.

[21]   Dave Litwiller, "CCD vs. CMOS: facts and fiction," *Photonics Spectra,* 2001.

[22]   HowStuffWorks Inc, "What is the difference between CCD and CMOS image sensors in a digital camera," 2008, "http://electronics.howstuffworks.com/cameras-photography/digital/question362.htm".

[23]   Epix Inc, "PIXCI® EL1DB," 2010, "http://www.epixinc.com/products/pixci_el1db.htm".

[24]   Wikipedia, "Depth of field," 2010, "http://en.wikipedia.org/wiki/Depth_of_field#DOF_limits".

[25]   Lloyd L. Chambers, "Zeiss ZF Lenses for Infrared," 2010, "http://diglloyd.com/articles/Infrared/ZeissZF-prototypes-infrared.html".

[26]   Wikipedia, "Infrared," 2010, "http://en.wikipedia.org/wiki/Infrared".

[27]   Cypress Semiconductor Corporation, "LUPA 4000: 4 MegaPixel CMOS Image Sensor," 2009.

[28]   Opnext Japan Inc., "HE8812SG GaAIAs Infrared Emitting Diode," 2008, "http://www.opnext.com/products/pdf/ode2_063_he8812sg.pdf".

[29]   Honeywell Sensing and Control, "SE3470 AIGaAs Infrared Emitting Diode," "http://sensing.honeywell.com/index.cfm/ci_id/131470/la_id/1/document/1/re_id/0".

[30]   Panasonic Corporation, "LN152 GaAs Infrared Light Emitting Diode," 2002, "http://www.semicon.panasonic.co.jp/ds4/LN152_EEK.pdf".

[31]   DH Sliney, "ICNIRP Statement on Light Emitting Diodes (LEDs) and Laser Diodes: Implications for Hazard Assessment," *Health Physics,* vol. 78, pp. 744-752, 2000.

[32] D. Sliney, D. Aron-Rosa, F. DeLori, F. Fankhauser, R. Landry, M. Mainster, J. Marshall, B. Rassow, B. Stuck, S. Trokel, T. M. West, and M. Wolffe, "Adjustment of guidelines for exposure of the eye to optical radiation from ocular instruments: statement from a task group of the International Commission on Non-Ionizing Radiation Protection (ICNIRP)," *Applied Optics,* vol. 44, pp. 2162-2176, Apr 10 2005.

[33] MR Shortis, TA Clarke, and T Short, "A comparison of some techniques for the subpixel location of discrete target images," *Videometrics III,* vol. 2350, 1994.

[34] Silicon Imaging Inc., "SI-4000F MegaCamera 4.2 Million Pixel Progressive Scan Digital Camera," 2005.

[35] MK Cheezum, WF Walker, and WH Guilford, "Quantitative comparison of algorithms for tracking single fluorescent particles," *Biophysical Journal,* vol. 81, pp. 2378-2388, 2001.

[36] TA Clarke, MA Cooper, and JG Fryer, "Estimator for the random error in subpixel target location and its use in the bundle adjustment," *Proceedings of SPIE,* pp. 161-168, 1994.

[37] MR Shortis, TA Clarke, and S Robson, "Practical testing of the precision and accuracy of target image centring algorithms," *Proceedings of SPIE,* pp. 65–76, 1995.

[38] B. Tech. Swapna Puthukkudichalil, "The Question of Accuracy in Camera Clibration and a Novel Technique for Estimating Camera Geometry," MEng, Memorial University of Newfoundland, pp. 1-20, 2009.

[39] RK Lenz and RY Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 713-720, 1988.

[40] J Salvi, X Armangué, and J Batlle, "A comparative review of camera calibrating methods with accuracy evaluation* 1," *Pattern Recognition,* vol. 35, pp. 1617-1635, 2002.

[41] GQ Wei and SD Ma, "A complete two-plane camera calibration method and experimentalcomparisons," *IEEE 4th International Conference on Computer Vision*, pp. 439-446, 1993.

[42] GQ Wei and Ma Song De, "Implicit and explicit camera calibration: theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 16, pp. 469-480, 1994.

[43] Y.I. Abdel-Aziz and H.M. Karara, "Mathematical formulation in close-range photogrammetry*"*, Fourth ed.: American Society of Photogrammetry, pp. 801-803, 1971.

[44] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 965-980, 1992.

[45] J Weng, TS Huang, and N Ahuja, "Motion and structure from two perspective views: algorithms, erroranalysis, and error estimation," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 11, pp. 451-476, 1989.

[46] OD Faugeras and G Toscani, "Camera calibration for 3D computer vision," *Proceedings of International Workshop on Machine Vision and Machine Intelligence*, Tokyo, Japan, 1987.

[47] R 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,* vol. 3, pp. 323-344, 1987.

[48]   Zhenyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, pp. 1330-1334, 2000.

[49]   DC Brown, "Close-range camera calibration," *Photogrammetric engineering,* vol. 37, pp. 855-866, 1971.

[50]   J Heikkila and O Silven, "A four-step camera calibration procedure with implicit imagecorrection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106-1112, 1997.

[51]   J Heikkila and O Silven, "Calibration procedure for short focal length off-the-shelf CCD-cameras," *Proceedings of the 13th International Conference on Pattern Recognition*, pp. 166-170, 1996.

[52]   J Heikkila, "Geometric camera calibration using circular control points," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, pp. 1066-1077, 2000.

[53]   W Faig, "Calibration of close-range photogrammetry systems: Mathematical formulation," *Photogrammetric Engineering and Remote Sensing,* vol. 41, pp. 1479-1486, 1975.

[54]   CC Slama, C Theurer, and SW Henriksen, "Manual of photogrammetry*"*, American Society of Photogrammetry, 1980.

[55]   Young-Hoo Kwon, "DLT method," 1998, "http://www.kwon3d.com/theory/dlt/dlt.html".

[56]   Wikipedia, "Singular value decomposition," 2010, "http://en.wikipedia.org/wiki/Singular_value_decomposition".