2010

3D Reconstruction Using a Stereo Vision System with Simplified Inter-Camera Geometry

Zhuo Wang
University of Windsor

3D RECONSTRUCTION USING A STEREO VISION SYSTEM WITH SIMPLIFIED INTER-CAMERA GEOMETRY

by
Zhuo Wang

A Thesis
Submitted to the Faculty of Graduate Studies
through School of Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Master of Master of Science at the
University of Windsor

Windsor, Ontario, Canada
2010

3D RECONSTRUCTION USING A STEREO VISION SYSTEM WITH SIMPLIFIED INTER-CAMERA GEOMETRY

by

Zhuo Wang

APPROVED BY:

Dr. Majid Ahmadi
Department of Electrical and Computer Engineering

Dr. Imran Ahmad
School of Computer Science

Dr. Boubakeur Boufama, Advisor
School of Computer Science

Dr. Yung H. Tsin, Chair of Defense
School of Computer Science

January 18, 2010

Declaration of Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research in collaboration with Dr. Boubakeur Boufama. The collaboration is covered in Chapter 4 of the thesis. In all cases, the key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primarily through the provision of guidance and criticism.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from the co-author to include the above material in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

Abstract

This thesis addresses the relationship between camera configuration and 3D Euclidean reconstruction. Simulations have been conducted and have shown that when error is present, the larger rotation angle, the worse the reconstruction quality. When rotation is avoided, errors in the intrinsic parameters do not affect the 3D reconstruction in a significant way. Therefore, it is suggested to minimize or avoid rotation when constructing a stereo vision system. Once this configuration is applied, inaccurate intrinsic parameters, even without the prior information of intrinsic parameters, can also yield good reconstruction quality. The configuration of pure translation also provides a framework, which can be used to compute elements of intrinsic parameters with an additional geometry constraint. The perpendicular constraint is selected as an example. Focal length can be recovered from this constraint by assuming the principal point is the centre of the image.

To the bright future.

Acknowledgements

I am hearily thankful to my supervisor, Dr Boubakeur Boufama, for his kind and patience during my study. Besides, I am grateful to my internal reader, Dr Imran Ahmad, and my external reader, Dr Majid Ahmadi, for their help to improve my thesis. Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of my thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computer vision is the theory of retrieving, interpreting and utilizing information from a single image, multiple images or videos. Most of the time cameras are used as the devices to acquire information. Camera calibration and scene reconstruction are two important tasks in computer vision, contributing to applications of robot navigation, stereo vision, pattern recognition, video surveillance, and others. In this thesis, the pin-hole camera model is adopted: 3D (three dimensional) scene points are projected through a single point (pin-hole) to an image plane.

Camera calibration is defined as the estimation or calculation of the intrinsic parameters of a camera. The intrinsic parameters consists of the followings: the principal point (image centre) coordinates; the focal length, which is the distance between principal point and camera centre (projection centre); the aspect ratio, which is the ratio of the horizontal size and the vertical size of a single pixel in a image; and the skew factor, which describes the distortion of pixels if they are not rectangular. Currently, widely-used commercial CCD cameras can provide the features which ensure that both the horizontal size and the vertical size of a single pixel are the same and the pixel is always rectangular. Therefore, in this

thesis it is convenient to assume that the aspect ratio is 1 and the skew factor is 0.

Scene reconstruction is defined as the process of recovering 3D scene information from an image or set of images. The projection of 3D scene to a 2D (two dimensional) image is a process in which we lose one dimension [1], and some useful information has been lost during the projection. The full reconstruction, which is very important for the upcoming application, is to recover the original Euclidean structure of the scene. In order to reconstruct the Euclidean structure of the scene, additional information is required. This is difficult because the projective structure lacks metric information [4]. In other words, one cannot fully reconstruct the scene from a single image without any prior information [3]. In this thesis, scene reconstruction means the recovery of the Euclidean structure of the scene.

Classical scene reconstruction methods require at least two images [31, 10, 43, 21]. Cameras need to be calibrated. Then the camera motion can be recovered. Once they are done, the Euclidean reconstruction is straightforward by solving a set of equations derived from the projection procedure. However, errors in intrinsic parameters are inevitable, and pixel errors always exist because feature correspondence between images is required during the camera calibration stage or the motion recovery stage. How errors in intrinsic parameters affect reconstruction quality has not been thoroughly studied. In this thesis we propose to study how camera motion affects the error influence to scene reconstruction when the intrinsic parameters are contaminated. Our study shows that rotation plays an important role in determining the effective level of error in intrinsic parameters: the larger the rotation angle, the larger the error influence. More interesting is if the rotation can be avoided, as in pure translation, the error of intrinsic parameters seems to be unable to affect the 3D reconstruction. This translation should be restricted to the vertical or horizontal direction. In practice, the rough value of intrinsic parameters are usually available from either the

manufacturer's data or previous experiments [4, 5]. Consequently once the pure translation is performed, the calibration can be totally avoided during the 3D reconstruction task.

If the motion is restricted as described before, the projection can be dramatically simplified since the rotation matrix is identity. It brings a neat expression of scene point coordinates in three directions, respectively, from intrinsic parameters and corresponding pixels. This turns the complicated camera calibration into a simple task: any geometry constraint in the scene can be applied if there is a constraint which can form sufficient equations to solve unknowns. In this thesis we propose a calibration method by applying the perpendicular constraint as an example. The output of such method is a single linear equation of three unknowns: focal length and two principal point coordinates. Any unknown parameter can be calculated if the other two parameters are known. It is suggested to use the perpendicular constraint for computing the focal length, since the principal point coordinates are almost equal to the centre of the image in modern CCD cameras. Some parts of this thesis have been published [51].

The remainder of this thesis is organized as follow: Chapter 2 introduces the fundamental concepts of camera calibration and scene reconstruction. Chapter 3 introduces the related works of this thesis. Chapter 4 shows how the idea of pure translation is derived and why this idea is effective, besides, this chapter also provides an innovative scheme of camera calibration if pure translation is applied and makes a perpendicular constraint as an example. Finally, Chapter 5 summarizes this thesis.

Chapter 2

Background

This chapter's aim is to introduce the prerequisite and fundamental concepts for camera calibration and scene reconstruction. It is divided into two parts: first, the illustration and the mathematical expression of the projection process is provided; next, stereo vision is introduced. In stereo vision, the general expression set to describe "stereo" is demonstrated. Then, because of the stereo feature, epipolar geometry is described in short. Derived from this epipolar geometry, an eight-point algorithm and a motion recovery algorithm are introduced. Once all of this is finished, it is straightforward to show how to recover the Euclidean structure of a scene.

2.1 Projection Process

A camera is a device which can project a 3D scene onto a 2D image. Since the pine-hole model is a good approximation of the real camera, the camera concept in this thesis is the pin-hole model assumption. Suppose all light rays are straight line, "pin-hole" means every ray should go through a single point, which lies in front of the inner part of the camera.

This single point is called ”pin-hole” or camera centre.

The whole projection process can be divided into three different steps:

The first step is a 3D rigid transformation. This transformation changes the 3D coordinate system from the scene coordinate system to the camera coordinate system, whose origin is the same as the camera origin. Such a centre is denoted as O. The camera origin O is also called “the projection centre”. This step can be avoided if the scene coordinate system is the same as the camera coordinate system, sometimes practical when the scene coordinate system is totally unknown.

The second step is a 3D to 2D transformation. The 3D scene is projected onto a camera frame through the projection centre O. A light ray between any 3D scene point and its corresponding camera frame point must go through O. Note that a camera frame point is a 2D point, and its coordinate system is the same as the camera coordinate without the depth axis (generally noted Z). The distance between the frame and the camera centre is the focal length. The boundary of a camera frame is infinite.

The last step is a 2D to 2D transformation. The goal of this step is to obtain the final image, which is acquired from the camera frame. This transformation changes the 2D coordinate, from the frame coordinate to the image coordinate. The image coordinate consists of u axis, v axis and the origin. If there is a line along the Z axis that goes through O, this line will intersect a camera frame. This intersection point, under the image coordinate, is called ”image centre” and denoted as o. Note that this transformation is not a rigid one: it is an affine transformation.

The whole projection process is illustrated in Figure 2.1, where O is the projection centre, o is the image centre, P is a scene point and p is the corresponding image pixel point. This figure shows how a scene point is projected into an image.

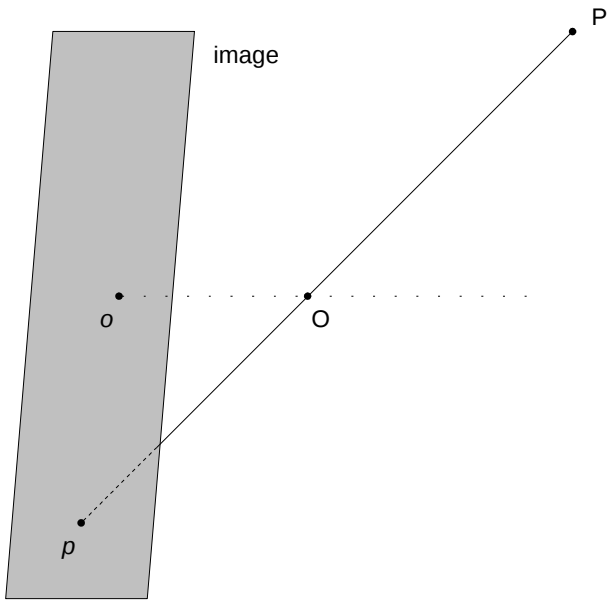

Figure 2.1: Projection Process

By using a homogeneous coordinate, the scene point P can be expressed as the vector $P = (X,Y,Z,T)^T$ and the image point p can be expressed as the vector $p = (u,v,\tau)^T$. From now on, if a scene point is described under the original scene coordinate system rather than the camera coordinate system, it is denoted as $\tilde{P}$; if a scene point is described under the camera coordinate system rather than the scene coordinate system, it is denoted as P. Similarly, if a 2D point is on a camera frame and is described under a camera frame coordinate, it is denoted as $\tilde{p}$; if a 2D point is on a camera frame and is described under an image coordinate, it is denoted as p.

The first transformation, which is a 3D rigid transformation, can be expressed with the equation:

$$P = D\tilde{P} \tag{2.1}$$

where D is a 3D transformation matrix that describes the transformation from the scene coordinate system to the camera coordinate system. When considering the second transformation, which is a 3D to 2D projection transformation, Equation (2.1) will be changed to:

$$\tilde{p} = IP = ID\tilde{P} \tag{2.2}$$

where I is a 3D to 2D projection matrix. Finally when the last step, a 2D to 2D coordinate system transformation is under consideration, the expression will be:

$$p = A\tilde{p} = AIP = AID\tilde{P} \tag{2.3}$$

where A is the matrix demonstrating the 2D to 2D coordinate transformation process.

This coordinate transformation is an affine transformation. A is also called "the intrinsic matrix". It describes the intrinsic parameters of the camera. The details of this matrix are shown below:

$$A = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.4}$$

where α_u, α_v are image scale factors and u_0, v_0 are image centre coordinates under the image coordinate system. What's more, $\alpha_u = -fk_u$ and $\alpha_v = -fk_v$, where f denotes the focal length of the camera, k_u denotes the ratio between pixel coordinate unit and camera coordinate unit along u axis, and k_v denotes the ratio between pixel coordinate unit and camera coordinate unit along u axis. Sometimes α_u and α_v are also called "focal length", and it is applied in the following context. If the scene coordinate can be set as the same as the camera coordinate, D will be an identical matrix and $P = \tilde{P}$. If T and τ are set to be 1, a scale factor λ will be introduced. Therefore, due to the introduction of λ, I can be expressed as

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.5}$$

The overall projection equation will be:

$$p = \lambda AIP \tag{2.6}$$

or

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.7}$$

2.2 Stereo Vision

This section deals with the situation when there are two cameras with a fixed configuration. It is convenient to assume that the coordinate of one camera can be adopted as the 3D scene coordinate. For simplicity, the reference camera is called the ”left camera” and another camera is called the ”right camera”. Assume that there is a scene point that can be projected into both cameras and can be observed in both images, then the 3D coordinate of this point can be recovered. This process is called scene reconstruction. See Figure 2.2. Note that images are placed in front of the camera centre, which is to facilitate analysis: the image acquired no longer represents an upside down scene.

In order to simplify this problem, it is appropriate to suggest that these two cameras are identical, meaning their intrinsic parameters are the same. Suppose that there is a scene point P, based on the previous discussion with one camera, the equation set can be easily formed to describe the projection process with two cameras:

$$\begin{cases} p = \lambda AIP \\ p' = \lambda' AI'P \end{cases} \tag{2.8}$$

where I' can be regarded as the camera's configuration: the motion from the first camera to the second. Another understanding is the transformation from the left camera's coordinate to the right camera's coordinate, which is a 3D to 3D rigid transformation. Therefore,

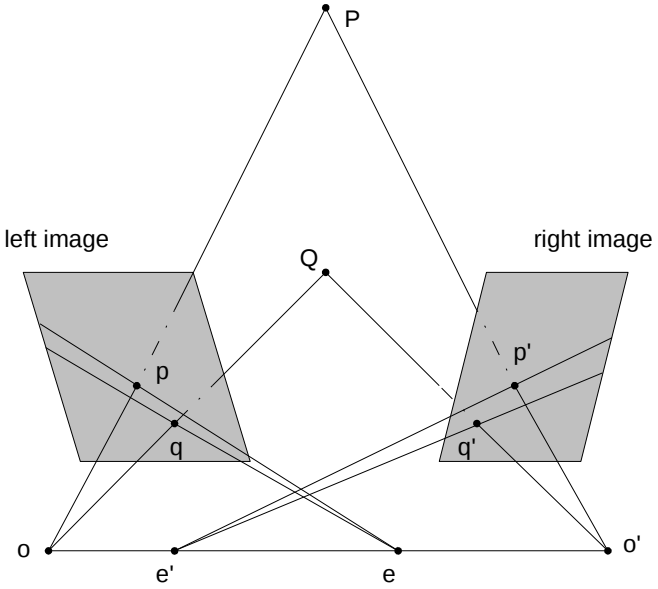

Figure 2.2: Stereo Vision

$I' = (R|t)$, where R is the rotation matrix and $t = (t_X, t_Y, t_Z)^T$ is the translation vector.

$$R = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \tag{2.9}$$

The details of Equation (2.8) is shown below:

$$\begin{cases} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \\ \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \lambda' \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & t_X \\ R_{21} & R_{22} & R_{23} & t_Y \\ R_{31} & R_{32} & R_{33} & t_Z \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \end{cases} \tag{2.10}$$

2.2.1 Epipolar Geometry

Figure 2.2 also illustrates the so-called "epipolar geometry" [40]. If there exists two cameras, along the epipolar line, their centres O and O', respectively, can be determined. Suppose there are two scene points P and Q, and both of them can be projected onto each image, the corresponding projected points can be noted as p, q on the left image, and p', q' on the right image. It is well-known that light rays travel in a straight line, so that p lies on the line PO. Similarly, q, p' and q' lie on the lines QO, PO' and QO', respectively.

Let's use P as an example. It is obvious that P, p, p', O and O' are coplanar. Therefore,

if lines are represented by vectors, the cross product of vectors $\vec{OO'}$ and $\vec{Op}$ is perpendicular to the vector $\vec{O'p'}$. The expression is:

$$\vec{O'p'} \cdot (\vec{O'O} \times \vec{Op}) = 0 \tag{2.11}$$

where $\cdot$ denotes scalar product and $\times$ denotes cross product. Equation (2.11) is correct if all the elements are defined in the same coordinate system. Suppose the coordinate system of one camera is selected as the reference, and denote $\vec{O'P'}$ as $\tilde{p'}^T$, $\vec{O'O}$ as t and $\vec{Op}$ as $\tilde{p}$. Equation (2.11) turns into:

$$\tilde{p'}^T \cdot (t \times R\tilde{p}) = \tilde{p'}^T TRP = 0 \tag{2.12}$$

where R is the rotation matrix and $t\times = T$, where T is the anti-symmetric matrix:

$$t\times = T = \begin{pmatrix} 0 & -t_Z & t_Y \\ t_Z & 0 & -t_X \\ -t_Y & t_X & 0 \end{pmatrix} \tag{2.13}$$

Here a new matrix $E = TR$ is adopted. E is called the "essential matrix". Then Equation (2.12) becomes

$$\tilde{p'}^T E\tilde{p} = 0 \tag{2.14}$$

Inside Equation (2.14), $E\tilde{p} = TR\tilde{p}$ is a vector with three elements which can be denoted as a, b and c, respectively. Then if $\tilde{p'}^T = (x, y, 1)$, Equation (2.14) can be transformed into a single equation:

$$ax + by + c = 0 \tag{2.15}$$

Equation (2.15) is a line equation. Therefore, it demonstrates that point p' lies on a line formed by $E\tilde{p}$. This is the inner connection between two image points projected from a single scene point: if each of the two points is determined, the position of another point can be limited on a line. This is also called "epipolar constraint" or "Longuet-Higgins constraint". This constraint can be used to find the corresponding point in another image if one feature point in one image is selected.

Point p and point p', which are discussed above, are defined within the normalised coordinates, which means these two points are defined under a camera frame coordinate system. In order to transform from normalised coordinates to pixel coordinates, which means these two points are defined under image coordinate system, the intrinsic matrix A should be introduced. If point p and point p' are defined under pixel coordinate system, since $\tilde{p} = A^{-}1p$ and $\tilde{p'} = A'^{-1}p'$, Equation (2.15) becomes

$$(A'^{-1}p')^T E(A^{-}1p) = 0 \tag{2.16}$$

Then

$$p'^T A'^{-T} E A^{-1} p = 0 \tag{2.17}$$

Here we introduce a new matrix $F = A'^{-T} E A^{-1}$ called "fundamental matrix". Then Equation (2.17) turns into:

$$p' F p = 0 \tag{2.18}$$

2.2.2 Eight-Point Algorithm

As shown in Equation (2.18), the fundamental matrix F can be determined by p and p'. Denote that $p = u, v, 1^T$ and $p' = u', v', 1^T$. Equation (2.18) turns into:

$$p'Fp = \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0 \tag{2.19}$$

That is:

$$uu'F_{11} + uv'F_{21} + uF_{31} + vu'F_{12} + vv'F_{22} + vF_{32} + u'F_{13} + v'F_{23} + F_{33} = 0 \tag{2.20}$$

Considering all image point pairs (an image point and its corresponding point in another image are called an image point pair), since one pair can form an equation by Equation (2.20), then a set of n equations can be rewritten as:

$$Af = \begin{pmatrix} u_1u'_1 & u_1v'_1 & u_1 & v_1u'_1 & v_1v'_1 & v_1 & u'_1 & v'_1 & 1 \\ u_2u'_2 & u_2v'_2 & u_2 & v_2u'_2 & v_2v'_2 & v_2 & u'_2 & v'_2 & 1 \\ u_3u'_3 & u_3v'_3 & u_3 & v_3u'_3 & v_3v'_3 & v_3 & u'_3 & v'_3 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_nu'_n & u_nv'_n & u_n & v_nu'_n & v_nv'_n & v_n & u'_n & v'_n & 1 \end{pmatrix} \begin{pmatrix} F_{11} \\ F_{21} \\ F_{31} \\ F_{12} \\ F_{22} \\ F_{32} \\ F_{13} \\ F_{23} \\ F_{33} \end{pmatrix} = 0 \qquad (2.21)$$

where f contains all elements of the matrix F and A is the equation matrix. Since F and f are defined up to an unknown scale, the additional constraint of forcing the norm of f to be 1 can be made. Therefore, thanks to this additional constraint, eight points, which can form eight equations, are sufficient to solve Equation (2.21). Some algorithms, such as Jacobi or Singular Value Decomposition (SVD), can be applied to find the least eigenvector of A^TA. The found eigenvector is the solution. Another method of solving Equation (2.21) is to set $F_{33} = 1$, which turns it into a linear least squares minimisation problem. It is claimed that general conclusions from these two algorithms, additional constraint and mandatory normalisation, are equally valid [20].

Because the classical eight-point algorithm is sensitive if any element of the coordinates has been changed, the normalisation process by expressing coordinates in fixed canonical frame is advised [20]. Besides, to improve the condition of the matrix A^TA, there are two methods suggested [20]. One is by scaling the coordinates to make the average of

homogeneous coordinates be unity. Another is by scaling the translation to minimise the value of coordinates. In practice, there are two approaches for improvement: isotropic scaling and non-isotropic scaling. As described in [20], these two approaches are listed below:

Isotropic Scaling:

1. The points are translated so that their centroid is at the origin.
2. The points are then scaled so that the average distance from the origin is equal to $\sqrt{2}$.
3. This transformation is applied to each of the two images independently.

Non-isotropic Scaling:

1. The points are translated so that their centroid is at the origin.
2. Both of two principal moments become unity by applying Choleski factorisation.

2.2.3 Motion Recovery

Once the essential matrix E is acquired, it is possible to decompose it and to recover the rotation and translation elements. Due to [31], $E = RT$, where R is the rotation matrix and T is the translation matrix. The outline of the algorithm can be found in [22]:

1. Find E.
2. Find the SVD of $E = UDV^T$, where $D = diag(a,b,c)$ and $a \geq b \geq c$.
3. If the centre of the first camera is the centre of the reference frame, the motion matrix is one of the four following matrices:

$(UZV^T|U(0,0,1)^T)$

$(UZV^T|-U(0,0,1)^T)$

$(UZ^TV^T|U(0,0,1)^T)$

$(UZ^TV^T|-U(0,0,1)^T)$

where

$$Z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.22}$$

To select the correct motion matrix, randomly select one point pair and test if the reconstructed scene point is in front of the camera.

2.2.4 Scene Reconstruction

Suppose the intrinsic parameters, the rotation matrix and the translation matrix are known, as shown in Equation (2.10). It is straightforward then to reconstruct the scene. Suppose there is a matrix $M = AI$. The projection equation, Equation (2.3), turns into:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.23}$$

Expanding Equation (2.23):

$$\begin{cases} u = \frac{m_{11}X+m_{12}Y+m_{13}Z+m_{14}}{m_{31}X+m_{32}Y+m_{33}Z+m_{34}} \\ v = \frac{m_{21}X+m_{22}Y+m_{23}Z+m_{24}}{m_{31}X+m_{32}Y+m_{33}Z+m_{34}} \end{cases} \tag{2.24}$$

For a pair of image points $p = (u, v, 1)^T$ and $p' = (u', v', 1)^T$:

$$\begin{cases} u = \frac{m_{11}X+m_{12}Y+m_{13}Z+m_{14}}{m_{31}X+m_{32}Y+m_{33}Z+m_{34}} \\ v = \frac{m_{21}X+m_{22}Y+m_{23}Z+m_{24}}{m_{31}X+m_{32}Y+m_{33}Z+m_{34}} \\ u' = \frac{m'_{11}X+m'_{12}Y+m'_{13}Z+m'_{14}}{m'_{31}X+m'_{32}Y+m'_{33}Z+m'_{34}} \\ v' = \frac{m'_{21}X+m'_{22}Y+m'_{23}Z+m'_{24}}{m'_{31}X+m'_{32}Y+m'_{33}Z+m'_{34}} \end{cases} \tag{2.25}$$

The final format of Equation (2.25) is:

$$\begin{pmatrix} um_{31}-m_{11} & um_{32}-m_{12} & um_{33}-m_{13} \\ vm_{31}-m_{21} & vm_{32}-m_{22} & vm_{33}-m_{23} \\ u'm'_{31}-m'_{11} & u'm'_{32}-m'_{12} & u'm'_{33}-m'_{13} \\ v'm'_{31}-m'_{21} & v'm'_{32}-m'_{22} & v'm'_{33}-m'_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} m_{14}-um_{34} \\ m_{24}-vm_{34} \\ m'_{14}-u'm'_{34} \\ m'_{24}-v'm'_{34} \end{pmatrix} \tag{2.26}$$

Equation (2.26) is an over-determined equation and is easily solved.

2.3 Conclusion

This chapter introduces the foundation of this field. If the intrinsic parameters are known and two images are given, the whole reconstruction process can be summarized as:

1. Find eight image point pairs.

2. Calculate the fundamental matrix by the improved eight-point algorithm.

3. Calculate the essential matrix.

4. Decompose the essential matrix to obtain rotation and translation information.

5. Reconstruct the scene.

When the intrinsic parameters are unknown and there are no more specific information of the scene, it is possible to recover the projective or affine (non-metric) structure of the scene. However, it is impossible to recover the Euclidean structure of the scene. So that in this thesis an assumption is adopted that the intrinsic parameters are known, even if they are inaccurate.

Chapter 3

Related Work

The camera calibration and scene reconstruction have been studies for more than two decades. This chapter is to introduce these previous works. They can be divided into several categories: classical calibration, self-calibration, motion recovery and scene reconstruction with known intrinsic parameters, which are directly related to our work. The calibration is to calculate intrinsic parameters of a camera. There are two directions: the classical calibration and self-calibration. The classical calibration uses a well known pattern while self-calibration uses matched pixels across many images instead. Motion recovery is to recover the configuration between two cameras. Scene reconstruction is to recover the scene information.

3.1 Classical Calibration

Camera calibration is the process of calculating the intrinsic parameters of a camera. Supposing that not only accurate scene point coordinates, but also accurate corresponding pixel point coordinates can be acquired, and the scene coordinate system is the same as the cam-

era coordinate system, one scene point and its projected image point can provide two equations. Because the intrinsic parameters contain four elements, at least two points can provide sufficient information to calculate the intrinsic parameters. The straightforward way to solve it is to use the Direct Linear Transformation method. However, it is not practical since measuring the coordinates is costly. Hence many alternative approaches have been proposed to deal with this problem. Some require more than one image; others require only one image to calibrate the camera.

3.1.1 Calibration by More Than One Images

One direction is to use a calibration pattern. Let the camera take two or more pictures from a pattern, then calculate the intrinsic parameters by measuring the predefined pattern features, which should appear in the image. This requires a, specially designed pattern, which limits its application range.

A coplanar grid can be used as a pattern to calibrate the camera. Tsai [48, 49] proposed a calibration method which recovers the intrinsic and extrinsic parameters by making them best fit the measured image coordinates of known target point coordinates. This method has two stages. The first stage is to estimate extrinsic parameters (the rotation and the translation parameters related to the scene coordinate system) by closed form least squares estimation. The second stage is to estimate intrinsic parameters by applying an iterative non-linear optimization.

Zhang [57, 58] proposed another calibration method by letting a camera view a planar pattern at several (at least two) different orientations. The information of the motion is not required. This method uses a closed-form solution. Then a non-linear refinement from a maximum-likelihood criterion is adopted. However, in practice this method requires

many images, at least 15 to 20 images with different orientation, to achieve the promising accuracy, which limits its usage. Besides, the special designed pattern also restricts applicable fields. The advantage of this method is its accuracy. This method currently has been widely applied and many people use this method as the reference to test the accuracy of their proposed new methods.

3.1.2 Calibration by Only One Image

Calibration by only one image has been in development for many years. It can be divided into several categories: calibration based on vanishing points, calibration based on circles, calibration based on symmetry and calibration based on four coplanar control lines. However, these methods require the specific structure of the scene and besides, the accuracy of these calibration methods is worse than the calibration method by patterns [57, 58].

The first idea of adopting vanishing points to calibrate the camera was proposed by Caprile [6]. It is based on the view of a cube. Three vanishing points can be retrieved from the image of the cube and the intrinsic parameters can be calculated from the attribute of vanishing points. Extended work has been proposed [17, 45, 53, 54, 55, 1, 9, 16]. For example, the use of an orthogonal wedge, which is defined as two rectangular planes intersecting at right angles, as a reference structure [45]; using paralelepipeds rather than cubes, which are a subclass of parallelepipeds, to calibrate the camera [53, 54, 55]; using a single image containing a rectangular prism, which is to generate two vanishing points for camera calibration [1].

The first paper which uses circles in a single image to calibrate the camera is the work of [12], who proposes the method of calibrating the camera using one image containing two concentric circles of known radii. The assumptions of zero skew, unit aspect ratio,

no distortion and square pixels are made. A similar attempt is proposed by [46]: camera calibration from a single view which contains three spheres. Another method was proposed by [7]: camera calibration with two arbitrary coplanar circles, even if the centres of the circles and/or part of the circles are occluded. Besides, the work of [27] was proposed as the improvement of [12]: camera calibration by using planar pattern of pairs of concentric circles. What's more, the work of [8] was proposed: camera calibration with two arbitrary coaxial circles.

Other attempts were also proposed. The idea of using symmetry in a single image to calibrate the camera was proposed by [26]. There are three different categories from this idea. The first is calibration from translational symmetry. The second is calibration from reflective symmetry. The last is calibration from rotational symmetry. Also a method of camera calibration from a single view with four coplanar control lines was proposed by [42], who use control lines rather than control points to perform constraints. These constraints can be used to compute intrinsic parameters.

3.2 Self-Calibration

Another direction to calibrate a camera is the so-called "self-calibration" or "auto-calibration". Most of the self-calibration methods rely on the calculation of virtual shapes and then derive the intrinsic parameters. Early self-calibration methods are based on the calculation of the Dual of the Image of the Absolute Conic (DIAC) using the so-called Kruppa's equations [11, 32, 29, 52]. Other methods rely on the calculation of the Absolute Quadric [13, 25, 47] which has the advantage of constricting both the DIAC and the plane at infinity on which the Absolute Conic lies in space. More recently, the so-called Absolute Line-Quadric was shown to exhibit some nice properties when dealing with a camera with varying pa-

rameters [50, 38].

Another approach for self-calibrating a camera consists in upgrading the projective structure of a scene into an affine one from which the metric reconstruction can be easily obtained. This can be achieved by calculating the position of the plane at infinity in space either through the so-called modulus constraint [37, 18] or from a quasi-affine reconstruction [24, 36]. Unfortunately, the accuracy of the estimated parameters, when using the above mentioned methods and others [14, 35], is undermined by the correspondence problem and by the numerous degenerate motion configurations [28, 44]. Since camera self-calibration is a non-linear problem, the problem of choosing initial values of parameters is often difficult.

In addition, initializing the optimization procedures which are close to the ground truth does not guarantee the convergence to the desired solution. For example, when the candidate plane at infinity contains one of the camera centres, the optimization procedure fails even if the motion of the camera is not degenerated. In order to circumvent these issues, [15] have proposed a globally convergent method that uses interval analysis in order to bound the values of the camera parameters. However, the prohibitive running time of this method makes it inappropriate for most applications.

3.3 Motion Recovery

Motion recovery is one of the crucial steps of scene reconstruction. Sometimes it is also included as one of the camera calibration targets. It aims at finding out the rotation matrix and the translation vector/matrix of the camera. The motion recovery step can be performed at the essential matrix, which can be obtained from the fundamental matrix and intrinsic parameters. The fundamental matrix can be derived from a point correspondence by the

eight-point algorithm. Once the essential matrix is obtained, by using an SVD algorithm, the rotation matrix and the translation matrix can be retrieved [31, 39, 22]. Nevertheless, this algorithm is to be criticized as extensively sensitive to noise in the specification of the matched points [20]. Therefore, other but more complicated algorithms have been proposed for calculating the fundamental matrix [56, 34, 23, 2].

Then Hartley [20] improved the eight-point algorithm by adopting the idea of normalization. This improvement does not only make the algorithm be less sensitive to pixel error, but also make the algorithm be less sensitive to errors on the intrinsic parameters. By adopting this improved algorithm, the scene reconstruction can achieve the considerable accuracy [4]. Hence this algorithm is applied in this thesis.

3.4 Scene Reconstruction

Scene reconstruction is to recover the Euclidean information of the scene, which is of great value in computer vision [4]. The classical and most popular approach is based on known or roughly known intrinsic parameters. Once intrinsic parameters are known, the 3D structure can be recovered from matched points only [31].

Such reconstruction method relies on knowing the intrinsic parameters and extrinsic parameters (motion) of the camera. Once such camera parameters are obtained, the scene reconstruction is straightforward by using triangulation [31] or a least-squares algorithm. This strategy, however, is impractical under some situations, due to the previous calibration process. To solve this problem, scene reconstruction without calibration has been proposed [13, 30, 41, 59, 19, 23]. When the intrinsic parameters are unknown, the reconstruction problem becomes more difficult in the Euclidean case. Given two images of the same rigid scene, we have 15 parameters to be estimated. That is, five for the camera

motion (rotation and translation) and ten for the intrinsic parameters of the two cameras (assuming different cameras). On the other hand, using pixel correspondences yields only 7 independent constraints. [33]. Adding more images will not provide enough constraints for this problem when each new image has different intrinsic parameters. Therefore, in the general case, it is not possible to recover the Euclidean structure using only pixel correspondences across images. In accordance with the above this thesis uses classical scene reconstruction, which requires a priori calibration process to acquire intrinsic parameters.

When dealing with the situation where the intrinsic parameters are approximately known, it is still possible to solve the reconstruction problem [4]. The 3D reconstruction obtained this way is affected by both pixel accuracy and errors on the intrinsic parameters' values. Although inaccurate intrinsic parameters might not be as serious as high-level pixel errors, the effect of errors in intrinsic parameters still needs to be analyzed.

3.5 Conclusion

This chapter investigates the literature of related topics. Different calibration approaches and two categories of scene reconstruction are proposed. In order to achieve a Euclidean reconstruction when a general case is applied, a previous calibration process is needed. There are many calibration methods which have been proposed. But the most reliable method with two images is the improved eight-point algorithm. Therefore, in this thesis the process of scene reconstruction consists of two stages: calculate the fundamental matrix and the essential matrix by the improved eight-point algorithm, recovers the motion and then reconstruct the scene by reversing the projection process.

Chapter 4

Error Effects on 3D Reconstruction

From the previous discussion, it is obvious that the scene reconstruction process is affected by errors from pixels and intrinsic parameters. Both of these two kinds of errors are inevitable. Besides, similar to the white noise, the values of pixel error are random and impossible to predict, which makes their influence permanent. When regarding errors in intrinsic parameters, sometimes the intrinsic parameters approximation is known, either from manufacturer's data or previous experiments. Because intrinsic parameters do not directly and independently affect the scene reconstruction result, it is possible to reduce or even eliminate intrinsic parameter errors, for the reconstruction result.

The first step in solving this problem is reviewing the scene reconstruction equation: Equation (2.6):

$$\tilde{p} = \lambda A I \tilde{P}$$

To help facilitate analysis, it is convenient to assume that image pixels are error free. Considering the effect of errors in intrinsic parameters, the equation above can be reformed

to:

$$\begin{aligned} p + p_e &= \lambda(A + A_e)IP \\ &= \lambda AIP + \lambda A_e IP \end{aligned}$$

where A_e is the error matrix of intrinsic parameters and p_e is the error vector of pixel coordinates. The above equation can be reformed as:

$$p_e = \lambda A_e IP \tag{4.1}$$

It's obvious that I, the projection matrix, plays an essential role in the intrinsic parameters error influence. In order to demonstrate how motion and intrinsic parameters error corrupt the recovered scene, some simulations have been conducted. The reason for using simulations rather than real experiments is that experimental coefficients can be controlled. Without such absolute control, due to the unpredictable nature of pixel error, one cannot tell which coefficient conducts or dominates the final imperfect result. Therefore, no conclusions can be made.

It is well-known that the ratio α_u/α_v is usually stable and is equal to 1, which is true for real cameras. With this fact, intrinsic parameter errors can be restricted to be inside the two coordinates of the principal point and to the focal length. Note that here that the focal length is the alias of α_u and α_v. It is crucial to clarify the camera coordinate system and the image coordinate system: the X-Axis and Y-Axis of the camera coordinate are within the same direction as the u-Axis and v-Axis of the image coordinate system, respectively; Z-Axis of the camera coordinate system is with the direction from the camera centre to the

image centre and is perpendicular to the image plane.

To make an acceptable approximation, the coefficients of simulation are listed as follows: the pre-defined virtual scene consists of 50 random space points, which are inside a volume of 30cm(X) by 30cm(Y) by 5cm(Z); two virtual cameras are located at 20cm(Z) off the scene; for each designated inter-camera configuration, two images are created by projecting the scene into two cameras, respectively. These two images are then used as inputs for the scene reconstruction system. In addition, pixel error with 0.5 pixel level or 1.0 pixel level are added to projected image points. For every test, a total of 100 trials is carried out. For every trial, 3D space points are selected randomly and errors from pixels and intrinsic parameters are generated randomly, too. The results shown on the different graphs are obtained by taking the mean value for each case. The scenario of the simulation is the same as described in Conclusion part of Chapter 2.

Note that during the experiments, the value of any single 3D relative error is meaningless since the error influence can be scaled by modifying the value of the focal length, by resizing the scene or by changing the distance between the camera and the scene. Besides, since an image centre o is set to be at the central position of a specific image, the size of an image also affects the value of the 3D relative error. Therefore, attention should be paid only on the emerging trend caused by different rotation angles under certain circumstances. In simulation, the translation along Z-Axis is not under consideration. The reason is as follows: Assume there are two cameras placed along Z-Axis. Even if the front camera is transparent and does not affect the function of another camera, it is still impossible to analysis it. This is because distances between the scenes are different, which is similar to the situation if these two cameras have two different focal lengths. It is against the former assumption that two cameras are identical.

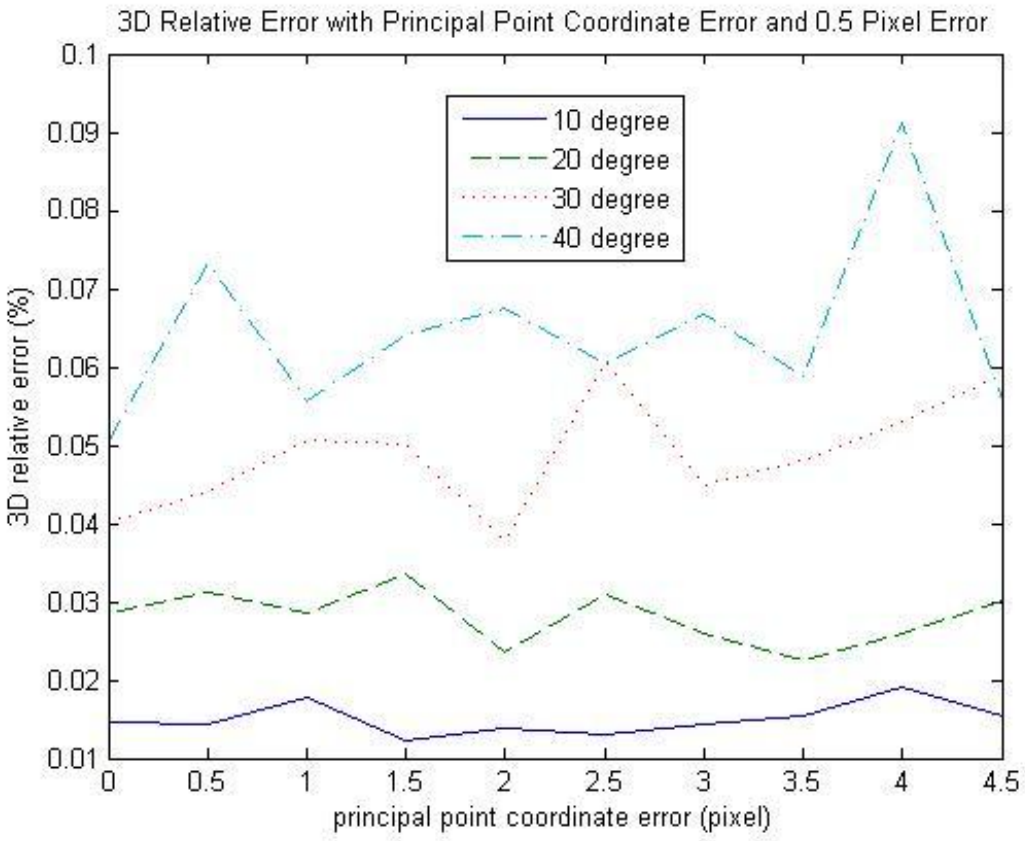

Figure 4.1: Translation along X-Axis and regular rotation, with principal point coordinate error and 0.5 pixel error level

4.1 General Case: Rotation and Translation Between the Two Cameras

First, the translation between each camera is set to be 20*cm* along the X-Axis. The results are listed below:

Figure 4.1 and Figure 4.2 show experimental results with various rotation angles and a constant translation along the X-Axis if principal point coordinate error is applied. Because the error is randomly selected, the position of the principal point is not stable and will

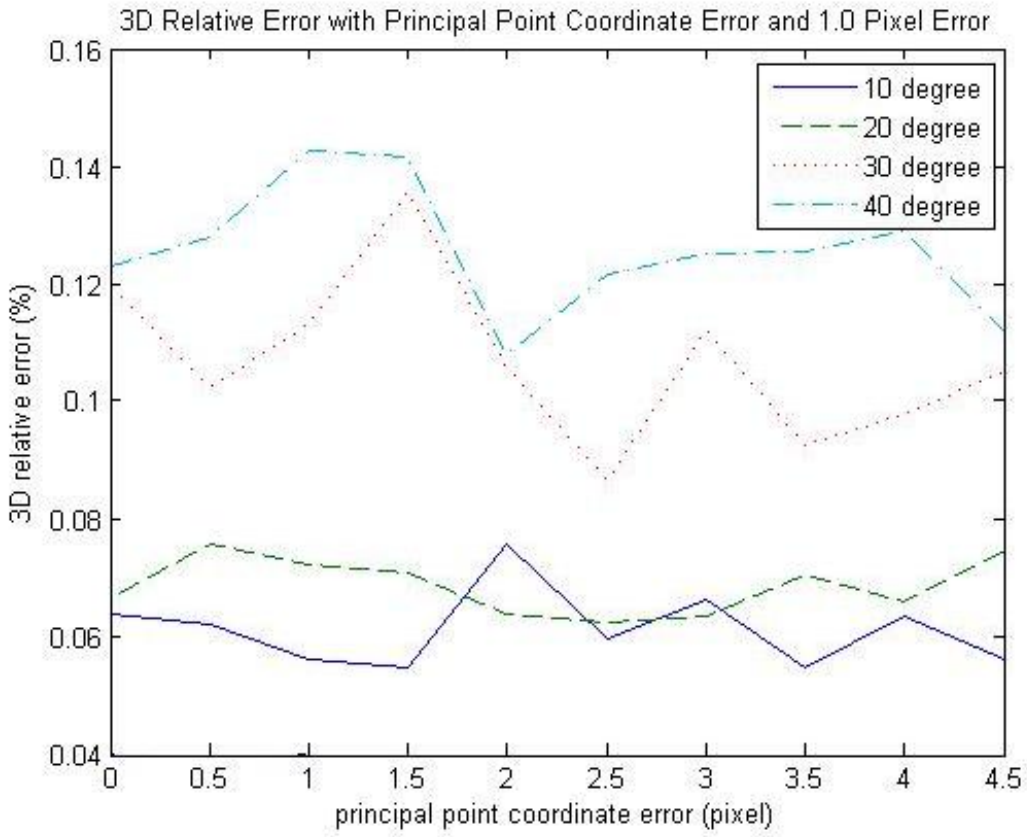

Figure 4.2: Translation along X-Axis and regular rotation, with principal point coordinate error and 1.0 pixel error level

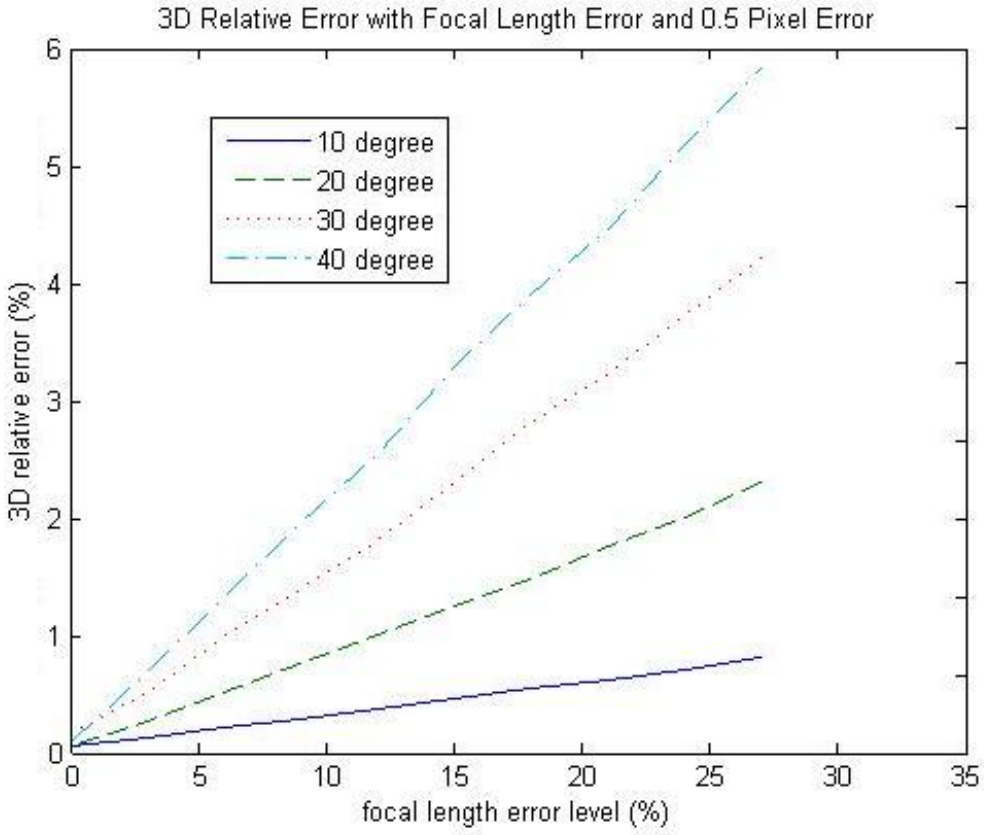

Figure 4.3: Translation along X-Axis and regular rotation, with focal length error and 0.5 pixel error level

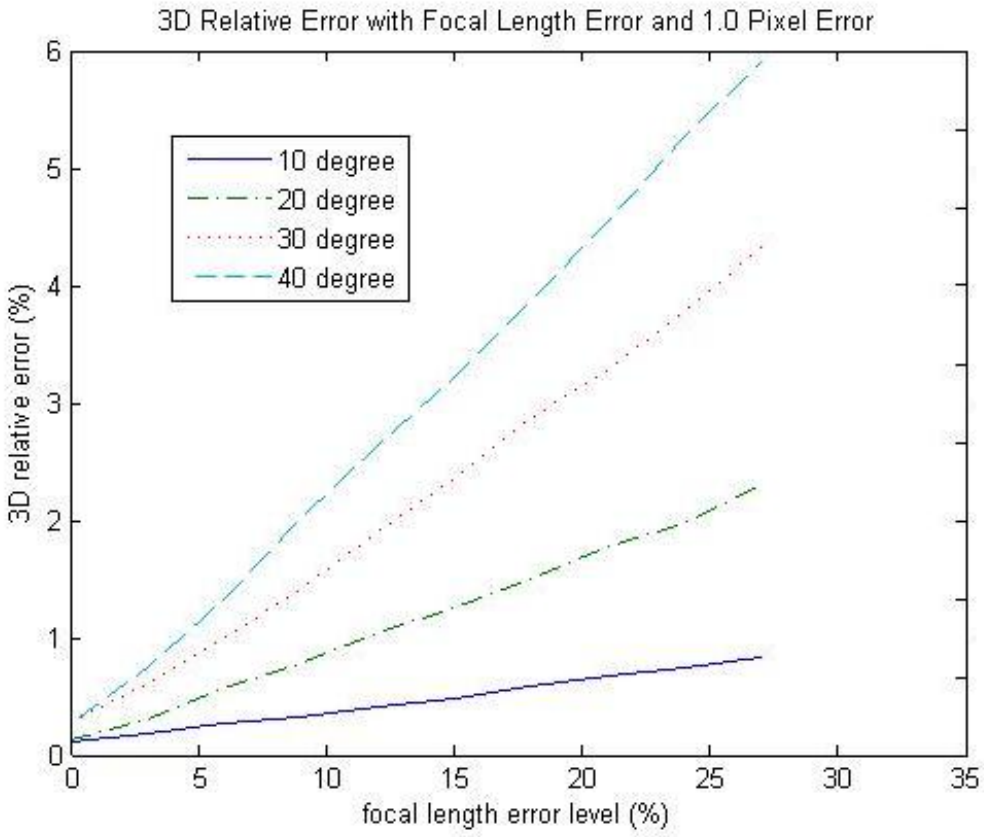

Figure 4.4: Translation along X-Axis and regular rotation, with focal length error and 1.0 pixel error level

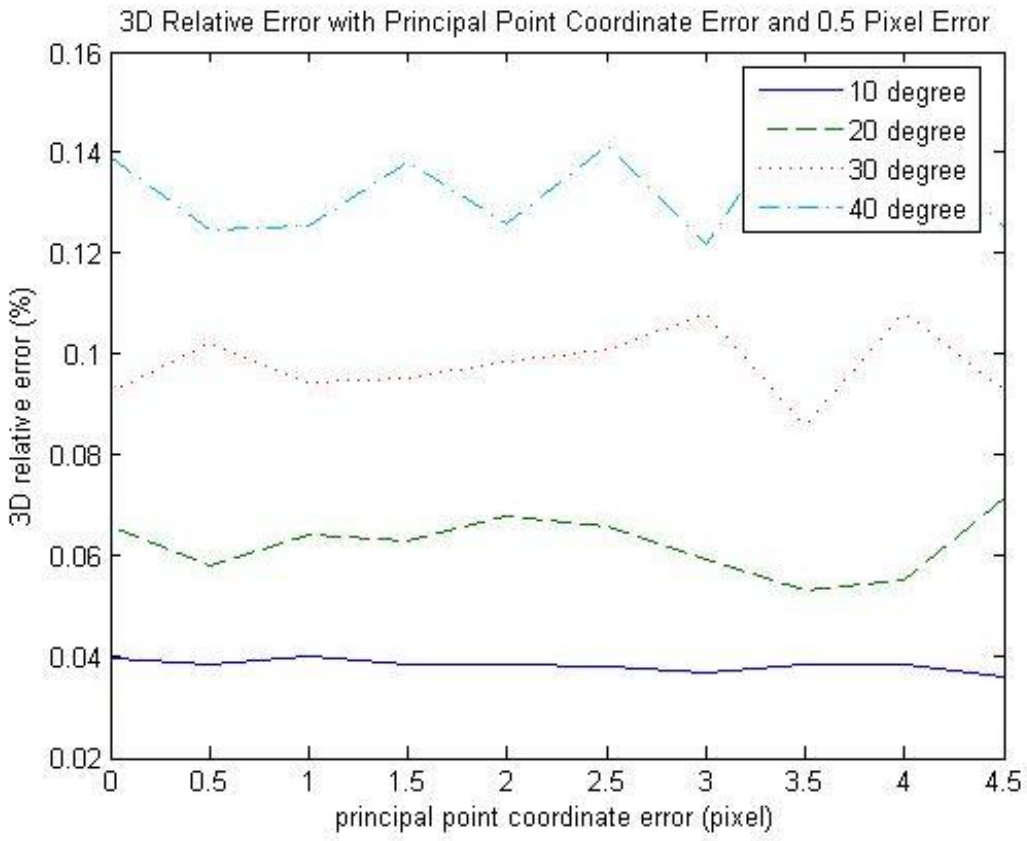

Figure 4.5: Translation along Y-Axis and regular rotation, with principal point coordinate error and 0.5 pixel error level

affect the overall reconstruction quality. However, it still can be concluded that most of the time smaller rotation angles yield better reconstruction quality. Figure 4.3 and Figure 4.4 show experimental results with various rotation angles and a constant translation along the X-Axis if focal length error is applied. For these two figures, the trend is very clear: the larger the rotation angle, the higher the error on the 3D reconstruction. In conclusion, the reconstruction quality drops if the rotation angle increases.

The next set of figures show the results when the translation is $20cm$ and along the Y-Axis.

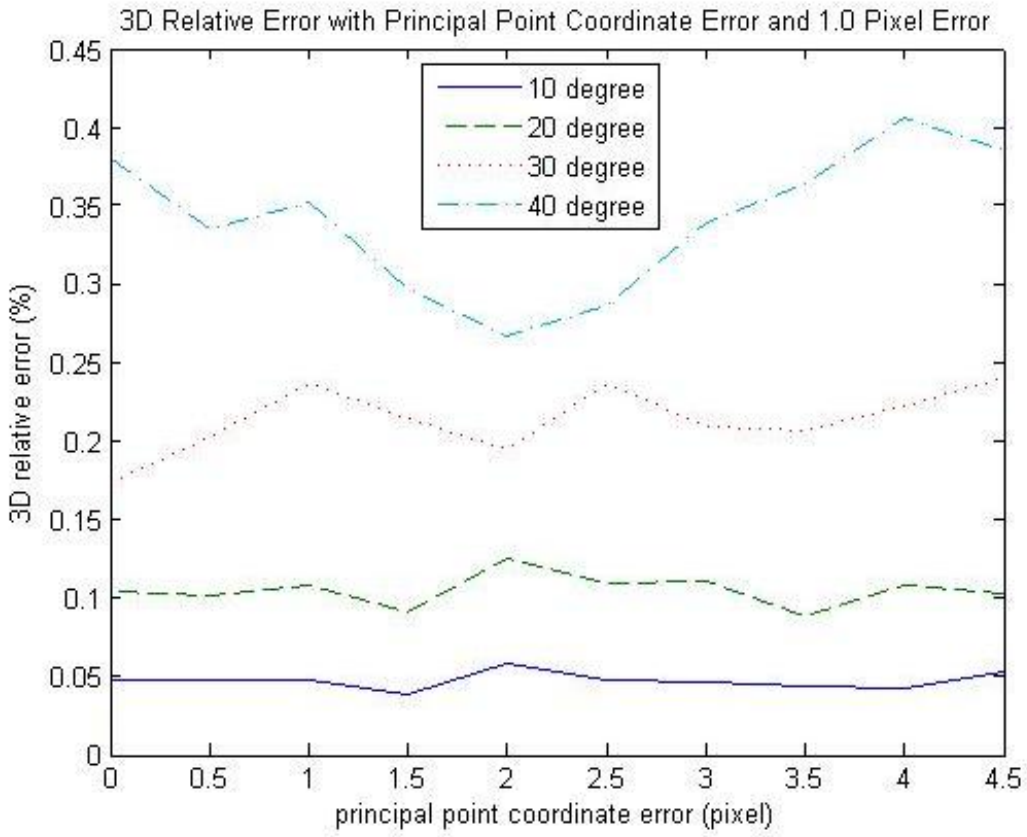

Figure 4.6: Translation along Y-Axis and regular rotation, with principal point coordinate error and 1.0 pixel error level

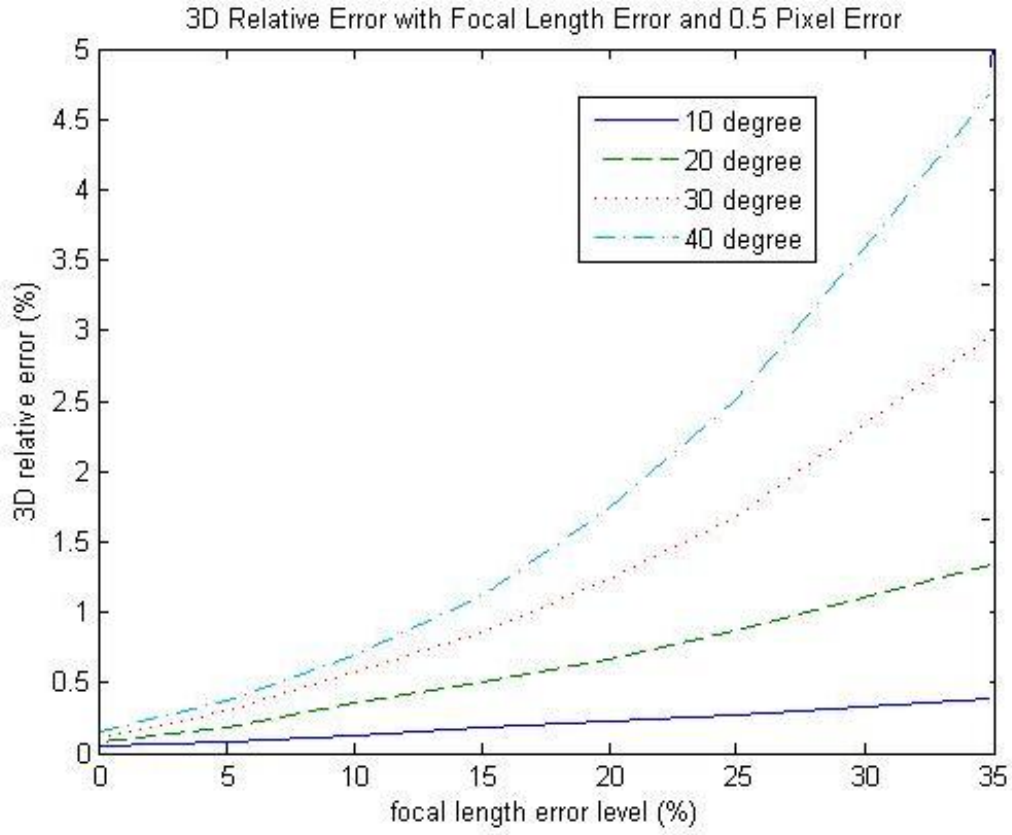

Figure 4.7: Translation along Y-Axis and regular rotation, with focal length error and 0.5 pixel error level

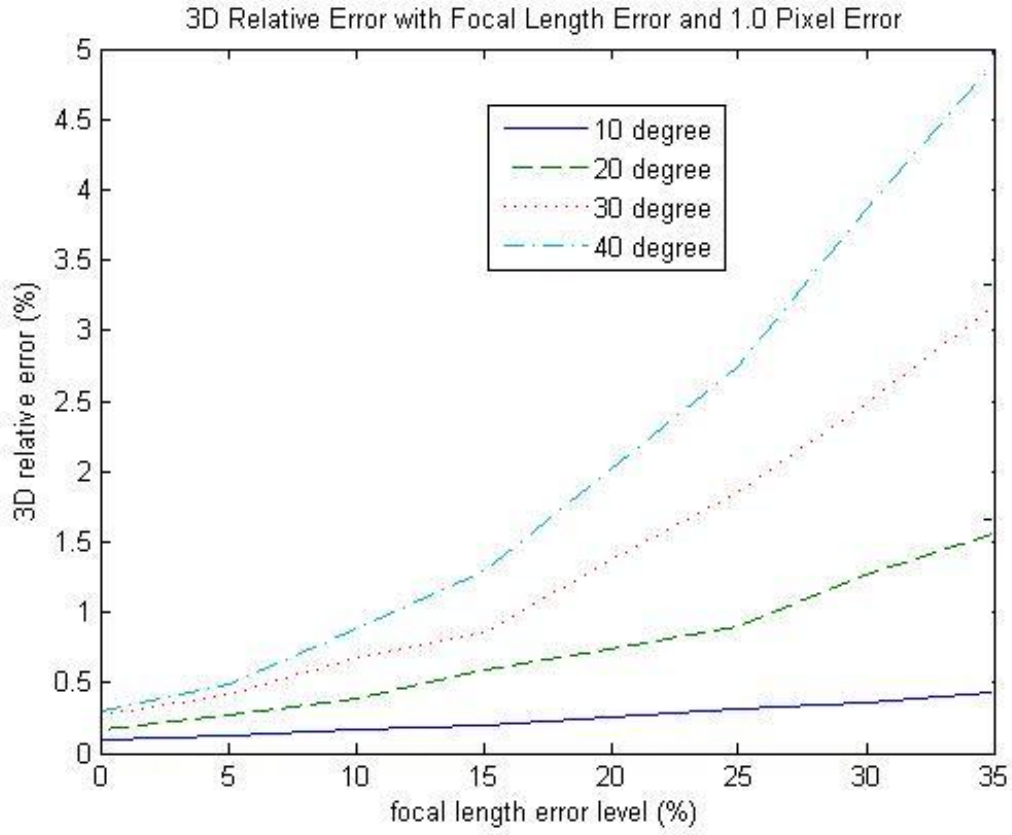

Figure 4.8: Translation along Y-Axis and regular rotation, with focal length error and 1.0 pixel error level

Figure 4.5, Figure 4.6, Figure 4.7 and Figure 4.8 also verify the conclusion that the bigger the rotation, the worse the reconstruction quality.

4.2 Simplified Case: Pure Translation Between the Two Cameras

As discussed above, the rotation angles between the two cameras amplify the effect of errors on reconstruction quality. It is quite straight-forward to think about the situation when the motion is pure translation. Simulations have been performed under this situation. Coefficients are kept the same.

Figure 4.9, Figure 4.10, Figure 4.13, Figure 4.14, Figure 4.11, Figure 4.12, Figure 4.15 and Figure 4.16 again demonstrate the discipline: the smaller the rotation angle, the better the reconstruction quality. More interesting is if the rotation is totally avoided, the reconstruction quality is best and errors in the intrinsic parameters seem to be not affected by the reconstruction quality. It is illustrated in all of these figures. It is because the projection matrix I plays an essential role in scene reconstruction and this matrix is directly determined by the inter-camera configuration. When there is no rotation, I is:

$$I = \left(\begin{array}{ccc|c} 1 & 0 & 0 & \\ 0 & 1 & 0 & |t \\ 0 & 0 & 1 & \end{array} \right) \tag{4.2}$$

The way of how translation affects the reconstruction error, when there are only errors from intrinsic parameters, is the same as when introducing inaccurate scale factors. Since scale factors can be eliminated during the 3D reconstruction process, errors from intrinsic

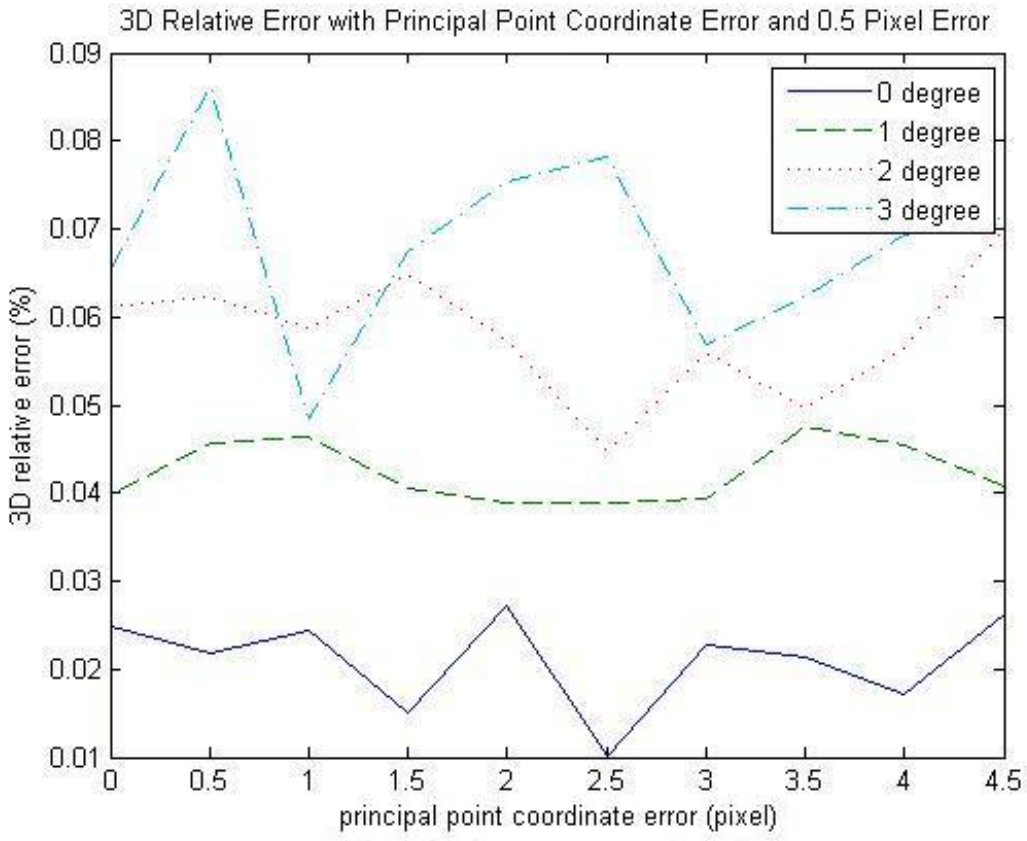

Figure 4.9: Translation along X-Axis and small rotation, with principal point coordinate error and 0.5 pixel error level

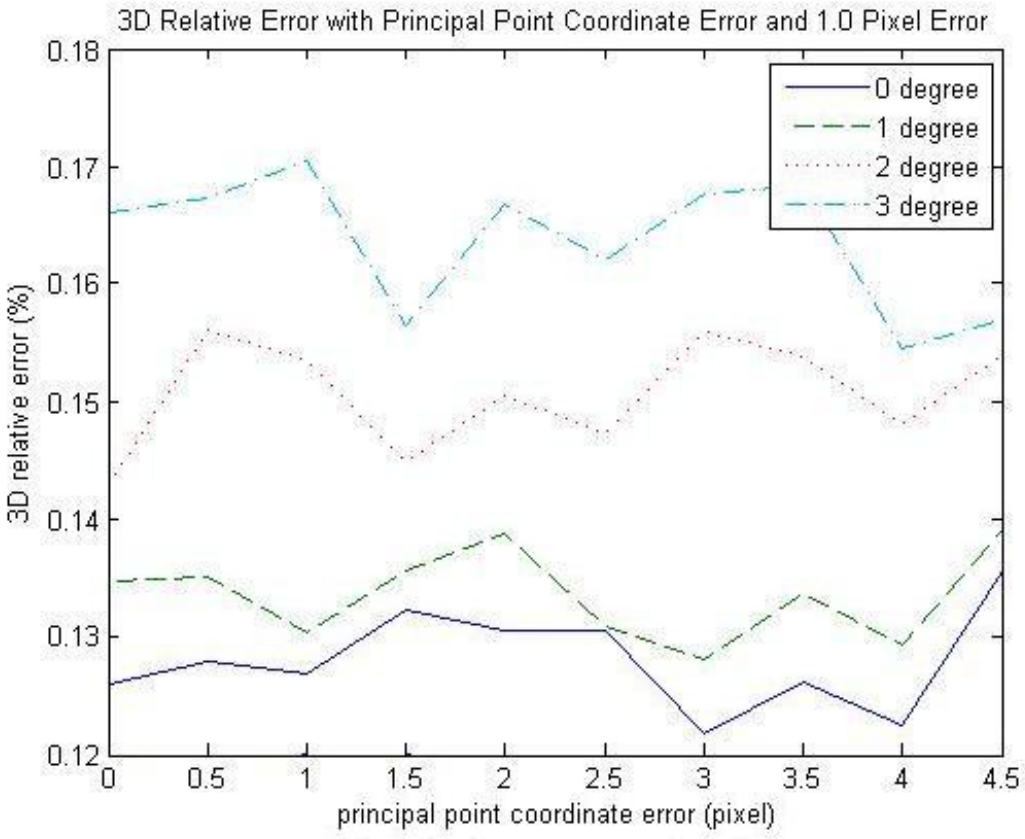

Figure 4.10: Translation along X-Axis and small rotation, with principal point coordinate error and 1.0 pixel error level

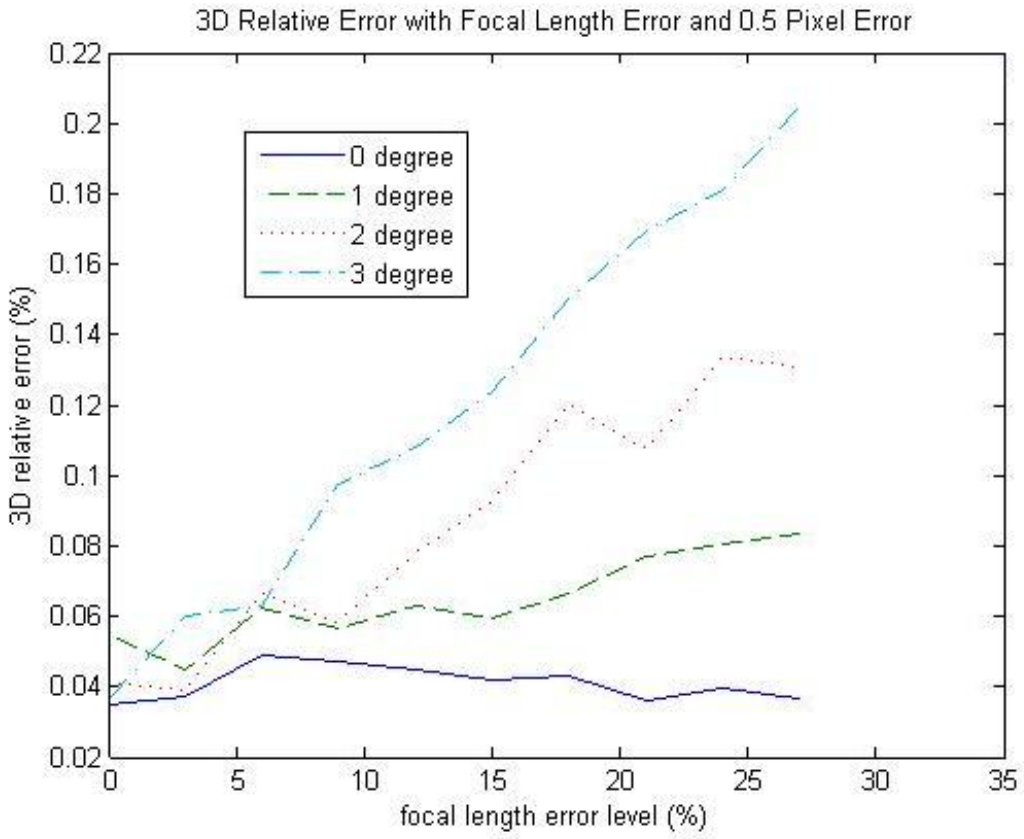

Figure 4.11: Translation along X-Axis and small rotation, with focal length error and 0.5 pixel error level

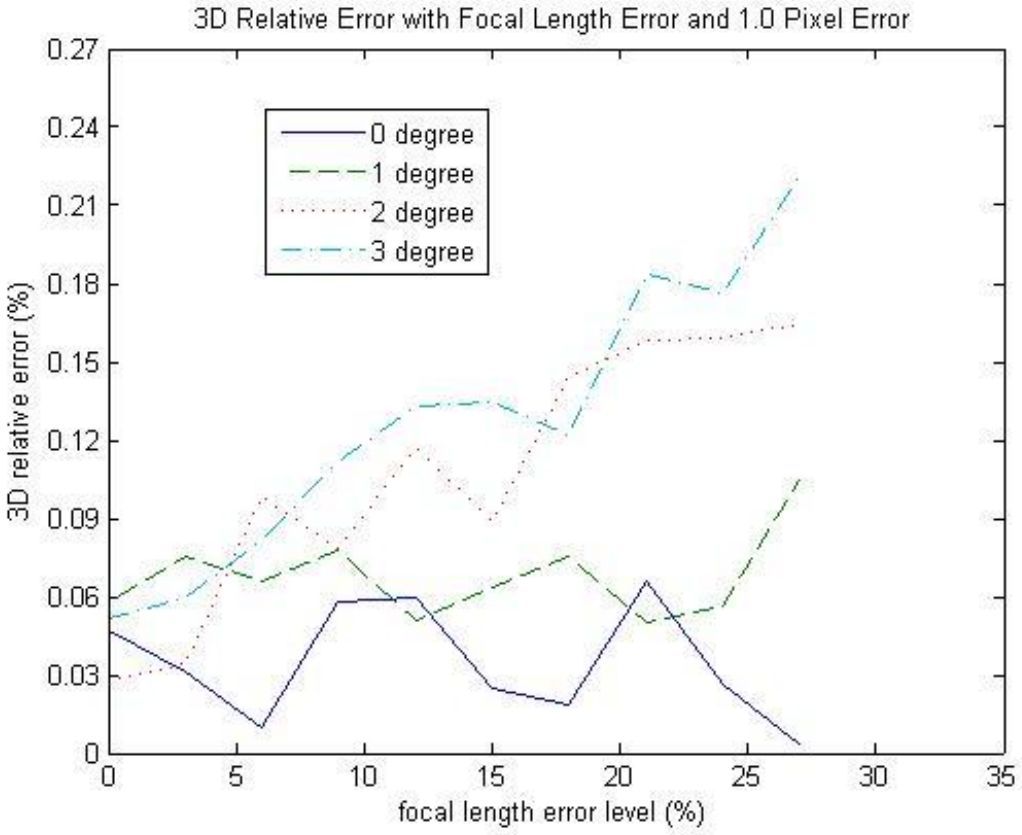

Figure 4.12: Translation along X-Axis and small rotation, with focal length error and 1.0 pixel error level

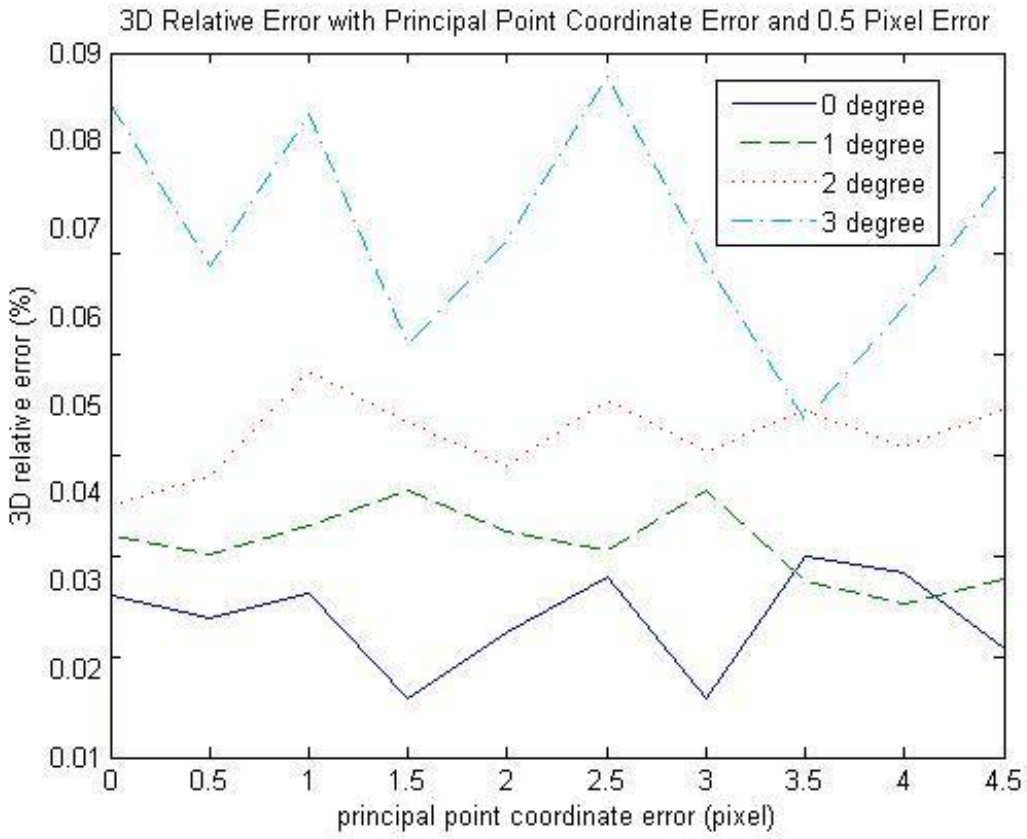

Figure 4.13: Translation along Y-Axis and small rotation, with principal point coordinate error and 0.5 pixel error level

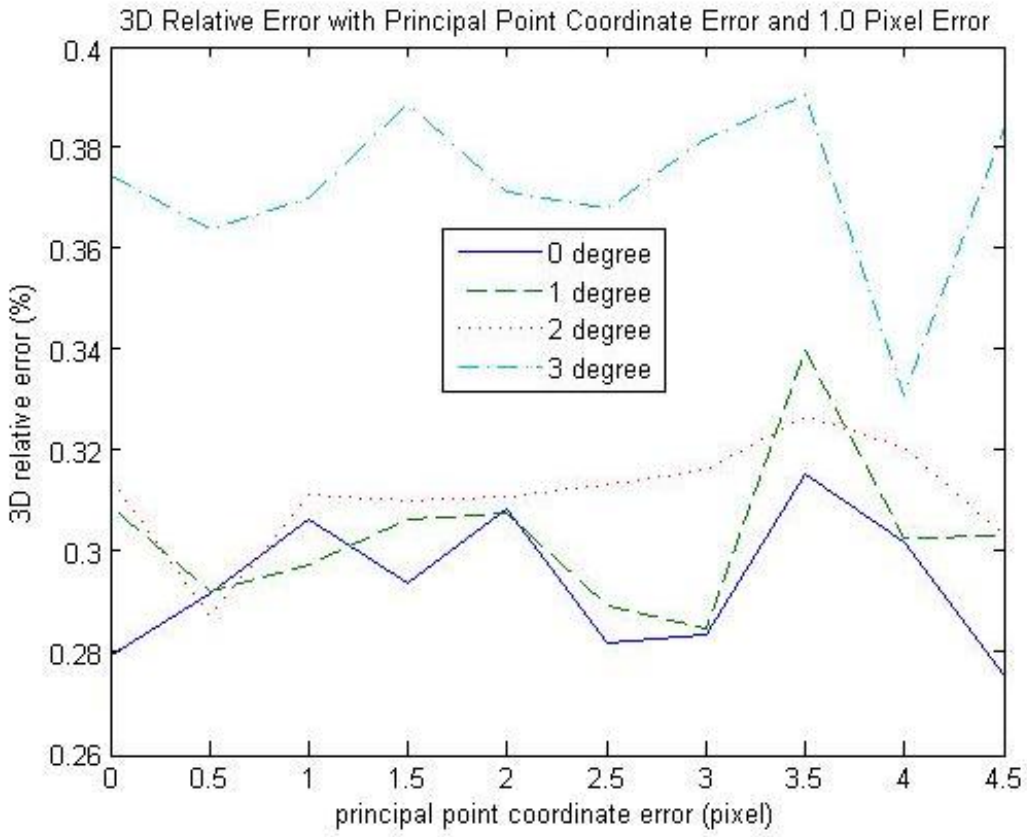

Figure 4.14: Translation along Y-Axis and small rotation, with principal point coordinate error and 1.0 pixel error level

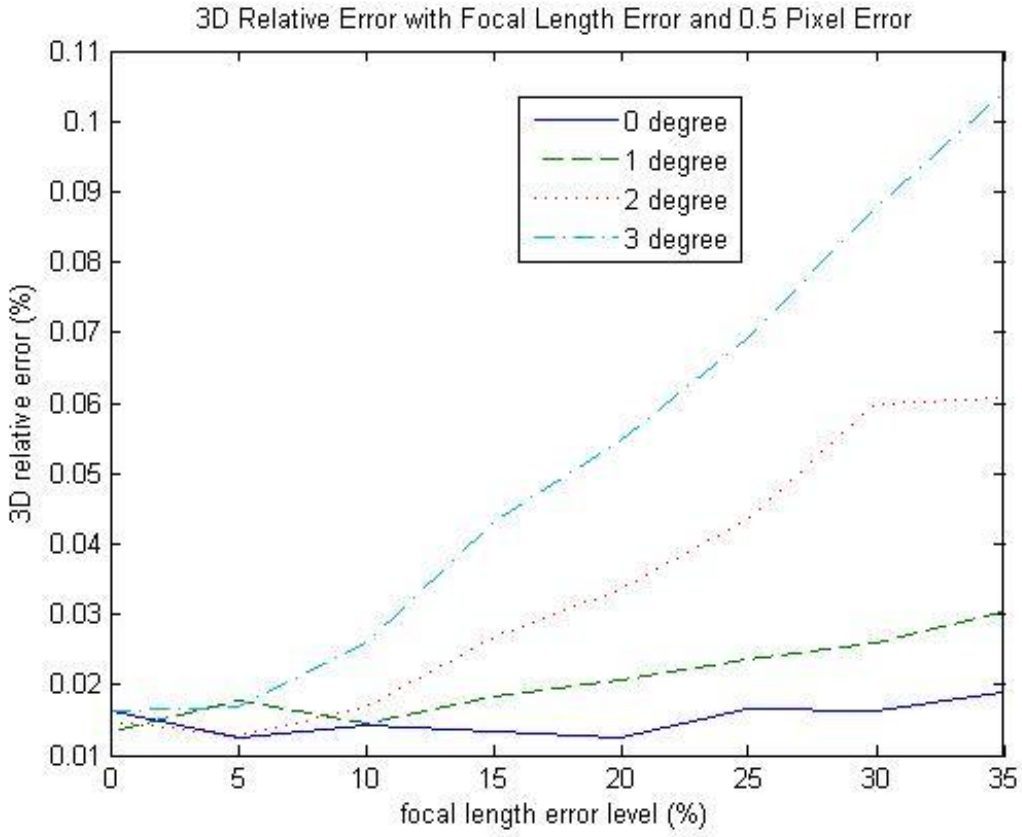

Figure 4.15: Translation along Y-Axis and small rotation, with focal length error and 0.5 pixel error level

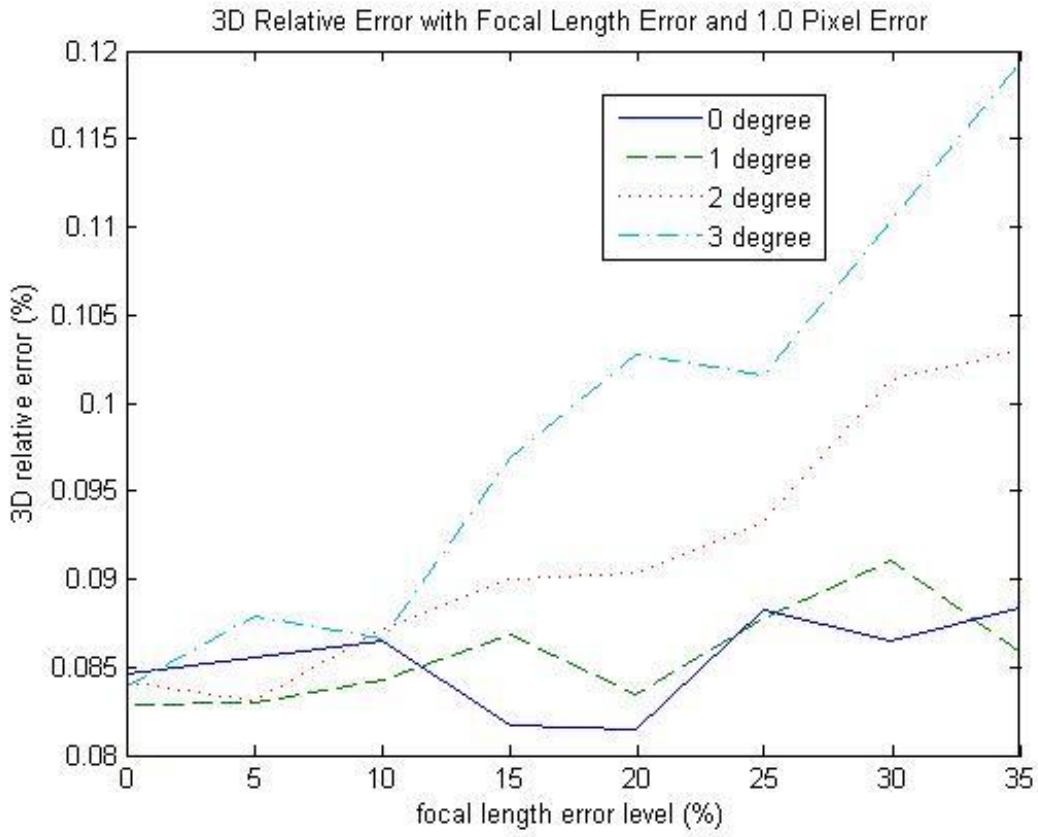

Figure 4.16: Translation along Y-Axis and small rotation, with focal length error and 1.0 pixel error level

parameters also do not affect the reconstruction quality if pure translation is applied.

4.3 Discussion

The simulation result verifies the initial guess: camera motion (inter-camera configuration) plays an important role in determining how much error will corrupt the reconstruction quality. This is because, as shown from above, that a large rotation will significantly amplify the noise effect derived from the pixel error or errors in intrinsic parameters. When a pure translation along the X-Axis or Y-Axis or their combination is applied, errors in intrinsic parameters will no longer affect the reconstruction quality.

Simulations were also conducted when the translation is set to the combination of movement along the X-Axis and movement along the Y-Axis, and the results verify the same trend. However the reconstruction quality is worse when the combined translation is applied under the same conditions. This is reasonable since pure translation along the X-Axis or the Y-Axis only introduces one unknown into the matrix I, but the combined translation introduces two unknowns. Hence in the next chapter, when referring to the pure translation, it means the translation along the X-Axis or the Y-Axis separately.

In order to construct an optimized system, pure translation, or translation and a very small rotation is suggested. Once pure translation is applied, accurate intrinsic parameters will no longer be needed and a very good 3D reconstruction quality can still be acquired. In practice, pure translation is easy to implement: align two cameras together or move a single camera along a straight line. Even if absolute translation is hard to achieve, minimized rotation still benefits the accuracy of the system.

4.4 The Use of Perpendicularity to Obtain the Focal Length

4.4.1 Mathematical Analyze

As discussed previously, pure translation should be applied. This section introduces another benefit when pure translation is introduced. The assumptions of $\tilde{f} = \alpha_u = \alpha_v$ and identical cameras are made. Since translation along the X-Axis or the Y-Axis is symmetrical, only translation along the X-Axis and Y-Axis is analyzed in detail. Equation (2.10) becomes:

$$\begin{cases} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \\ \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \lambda' \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & t_X \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \end{cases} \tag{4.3}$$

After transformation, the above equation becomes:

$$\begin{cases} X = t_X(a + bu_0) \\ Y = c + bv_0 \\ Z = -bt_X\tilde{f} \end{cases} \tag{4.4}$$

where

$$a = -\frac{2u^2 - 2uu' + (v - v')^2}{2(u - u')^2 + 2(v - v')^2}$$

$$b = \frac{u - u'}{(u-u')^2 + (v-v')^2}$$

$$c = -\frac{(u-u')(v+v')}{2(u-u')^2 + 2(v-v')^2}$$

Equation (4.4) is the overall expression of a scene point. To find out a scene point its image pair, the translation distance and intrinsic parameters are required. The scene point can also be used as a framework to calibrate the camera. Using certain space geometry constraint can form new equations, whose coefficients are image pairs, translation distance and intrinsic parameters. Since image pairs are easily known, the relationship among translation distance and elements of intrinsic parameters can be obtained. Sometimes translation distances can be eliminated and the unknowns are elements of intrinsic parameters only, which provide some methods of calculating intrinsic parameters (camera calibration).

Here the perpendicular constraint is adopted as an example. Given three scene points P_1, P_2 and P_3, each with different depth, and $\vec{P_1P_2}$ and $\vec{P_1P_2}$ are perpendicular, then

$$\vec{P_1P_2} \cdot \vec{P_2P_3} = 0$$

which is

$$(X_2 - X_1)(X_3 - X_2) + (Y_2 - Y_1)(Y_3 - Y_2) + (Z_2 - Z_1)(Z_3 - Z_2) = 0 \tag{4.5}$$

Use Equation (4.4) and Equation (4.5) turns to:

$$a'\alpha_u^2 + b'v_0 + c'(u_0^2 + v_0^2) + d'u_0 + e' = 0 \tag{4.6}$$

where

$$a' = (b_1 - b_2)(b_3 - b_2)$$

$$b' = (c_1 - c_2)(b_3 - b_2) + (c_3 - c_2)(b_1 - b_2)$$

$$c' = (b_1 - b_2)(b_3 - b_2)$$

$$d' = (a_1 - a_2)(b_3 - b_2) + (a_3 - a_2)(b_1 - b_2)$$

$$e' = (a_1 - a_2)(a_3 - a_2) + (c_1 - c_2)(c_3 - c_2)$$

Equation (4.6) has three unknowns: the focal length and two principal point coordinates. This shows that with a right angle in a scene, any intrinsic parameter can be calculated if the other two are previously known. As discussed before, it is convenient to assume that the principal point lies at the centre of the image. Under this assumption once the image size, which is easily fetched, is measured, the focal length can be calculated if there is a right angle existing in the scene. In the real world, right angles are common, such as windows, desks, walls and so on. Therefore, this method can be widely applied.

4.4.2 Experiment

Real experiments were conducted to test this algorithm. A pattern cube is shot by two aligned SONY DSC-S930 cameras and two sets of right angles are extracted manually

from images, respectively. The camera is previously calibrated by Zhang's Method [58]. The result is shown in Table 4.1. Note that values in entries have been rounded off to one decimal.

Table 4.1: Calibration Test

Test	Image Position	Right Angle	Focal Length	Reconstruction Average Error
1	Left	∠ABC	2132.4	11.1mm
	Right	∠abc		
2	Left	∠DEF	2031.3	11.7mm
	Right	∠def		
Zhang's Method			2242.4	10.3mm

The proposed calibration method provides similar results when compared to the wide-accepted Zhang's Method. The Zhang's method provides a more accurate focal length since its reconstruction quality is the best. It is because the focal length obtained from Zhang's method, which is shown in the table, is calculated by using 20 images.

Our method also works when there is only one camera: move the camera along a straight line and take two images for the scene. The length of the movement does not matter. However, in the future the proposed method still needs to be improved since its effectiveness relies on the accuracy of image point coordinates.

4.5 Conclusion

This chapter introduces several experiments which are to illustrate the relationship between rotation angles and error effects. These experiments show that when pixel error is applied, the larger the rotation angle, the worse the error influence from intrinsic parameters. Therefore, in order to minimize the error influence from intrinsic parameters, rotation should be decreased or avoided. The optimized configuration of a stereo system is the pure translation.

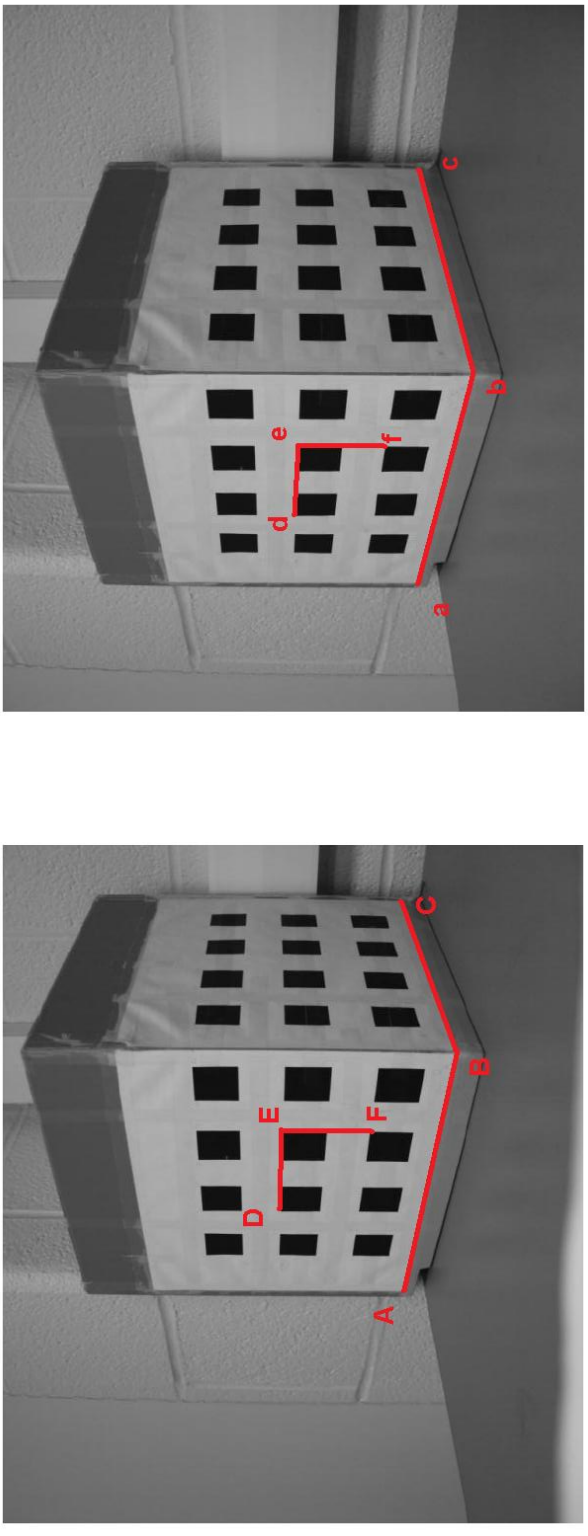

Figure 4.17: Two Corresponding Images of a Same Cube

Once this configuration is fixed, the three coordinates of a scene point can be expressed individually and explicitly. Thanks to those expressions, the framework of calibration then can be established. In this chapter a perpendicular constraint is selected to calibrate the camera by solving a linear equation and then perform the 3D reconstruction. Experiments show the effectiveness of this method.

Chapter 5

Conclusion

In this thesis we have addressed camera calibration, 3D Euclidean scene reconstruction and how inter-camera geometric configuration of a stereo vision system affects the accuracy of 3D reconstruction. We have reviewed two different calibration approaches together with motion recovery and two categories of scene reconstruction. When the intrinsic parameters and the stereo vision system geometry are unknown and there is no additional information about the scene, it is impossible to recover the 3D Euclidean structure of the scene. Note that it is possible to recover the 3D projective structure of a scene using only matched pixels. However, such 3D reconstruction lacks any metric information making useless for most applications. In order to obtain the 3D Euclidean reconstruction, one needs to know both the intrinsic parameters and the stereo vision system's geometry, namely, total calibration. Typically, one needs to first calibrate each camera of the stereo vision system before being able to perform the Euclidean 3D reconstruction of an observed scene. In the case of known intrinsic parameters, numerous methods have been proposed in the literature for 3D reconstruction using stereo images. However, the most reliable method that uses two images is the normalized (improved) eight-point algorithm. Therefore, in this thesis the process of

scene reconstruction consisted of the following steps. The calculation of the fundamental matrix and the essential matrix is done by the improved eight-point algorithm. Then, we recover the camera configuration before performing the reconstruction of the scene by reversing the projection process. Most of the time an approximation of the intrinsic parameters is easy to be acquired. Hence, we have investigated how errors in the intrinsic parameters affect the reconstruction quality and what kind of inter-camera geometry would be desirable in order to minimize the effects of these errors on the reconstructed scene.

In order to illustrate the relationship between input errors and 3D Euclidean reconstruction quality, extensive simulations have been conducted under different camera motions (inter-camera geometry). Input errors consist of pixel error and errors from intrinsic parameters. As a result, simulations show that when the range of pixel error is fixed, rotations amplify the error influence: the larger the rotation angle, the worse the reconstruction quality. What's more, when rotation is avoided and there is only translation between the two cameras, errors in intrinsic parameters do not affect the reconstruction quality significantly and the scene reconstruction yields the best accuracy. Therefore, it is suggested to avoid rotation, or at least minimize the rotation angle when constructing a stereo vision system in order to achieve a robust system against errors in intrinsic parameters. Note that pure translation should be along the X-Axis or the Y-Axis separately, since the combined translation brings worse reconstruction quality. This special configuration is close to a human's vision system: two eyes. This specific configuration also demonstrates that inaccurate intrinsic parameters can still yield a good reconstruction quality. Another benefit of pure translation is the ability to simplify the whole projection process mathematically since the rotation, which consists of three independent parameters (unknowns) in a projection equation, can be ignored. This simplified expression provides a framework which can be used to compute

the elements of the intrinsic parameters in assistance with additional geometrical constraint. The perpendicular constraint is selected as an example. When there is a right angle in the scene, the corresponding perpendicularity can form an equation, whose unknowns are intrinsic parameters. Focal length can be recovered from this equation by assuming that the principal point is the centre of the image, and then the whole Euclidean scene reconstruction can be done. The future work will be focused on the mathematical analysis of the projection process when both rotation and translation are applied, and on the improvement of the stability of the focal length calculation.

Bibliography

[1] N. Avinash and S. Murali. Perspective geometry based single image camera calibration. *Journal of Mathematical Imaging and Vision*, 30(3):221–230, 2008.

[2] P.A. Beardsley, A. Zisserman, and D.W. Murray. Navigation using affine structure from motion. *Lecture Notes in Computer Science*, 801:85–96, 1994.

[3] S.B. Boubakeur. On the recovery of motion and structure when cameras are not calibrated. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(5):735–759, 1999.

[4] B. Boufama and A. Habed. Three-dimensional structure calculation: achieving accuracy without calibration. *Image and Vision Computing*, 22(12):1039–1049, 2004.

[5] B. Boufama and A. Habed. Three-dimensional reconstruction using the perpendicularity constraint. In *Sixth International Conference on 3-D Digital Imaging and Modeling*, pages 241–248, 2007.

[6] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–139, 1990.

[7] Q. Chen, H. Wu, and T. Wada. Camera calibration with two arbitrary coplanar circles. *Lecture Notes in Computer Science*, pages 521–532, 2004.

[8] C. Colombo, D. Comanducci, and A. Del Bimbo. Camera calibration with two arbitrary coaxial circles. *Lecture Notes in Computer Science*, 3951:265, 2006.

[9] J. Deutscher, M. Isard, and J. MacCormick. Automatic camera calibration from a single manhattan image. *Lecture Notes in Computer Science*, pages 175–188, 2002.

[10] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[11] O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 321–334. Springer-Verlag, May 1992.

[12] V. Fremont and R. Chellali. Direct camera calibration using two concentric circles from a single view. In *International Conference on Artificial Reality and Telexistence*, pages 93–98, 2002.

[13] A. Fusiello. Uncalibrated euclidean reconstruction: A review. *Image and Vision Computing*, 18(6-7):555–563, May 2000.

[14] A. Fusiello. A new autocalibration algorithm: Experimental evaluation. In W. Skarbek, editor, *Computer Aanalysis of Images and Patterns*, volume 2124 of *Lecture Notes in Computer Science*, pages 717–724. Springer-Verlag, 2001.

[15] A. Fusiello, A. Benedetti, M. Farenzena, and A. Busti. Globally convergent autocalibration using interval analysis. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1633–1638, 2004.

[16] L. Grammatikopoulos, G. Karras, and E. Petsa. An automatic approach for camera calibration from vanishing points. ISPRS *Journal of Photogrammetry and Remote Sensing*, 62(1):64–76, 2007.

[17] E. Guillou, D. Meneveaux, E. Maisel, and K. Bouatouch. Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. *The Visual Computer*, 16(7):396–410, 2000.

[18] A. Habed and B. Boufama. Camera self-calibration from bivariate polynomial equations and the coplanarity constraint. *Image and Vision Computing*, 24(5):498–514, 2006.

[19] M. Han and T. Kanade. Creating 3D models with uncalibrated cameras. In *Fifth IEEE Workshop on Applications of Computer Vision*, pages 178–185, 2000.

[20] R. Hartley. In defence of the eight-point algorithm. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.

[21] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[22] R.I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. European Conference on Computer Vision*, volume 92, pages 579–587. NLCS, 1992.

[23] R.I. Hartley. Euclidean Reconstruction from Uncalibrated Views. In *Applications of invariance in computer vision: second joint European-US workshop, proceedings*, page 237. Springer, 1994.

[24] Richard I. Hartley, Lourdes de Agapito, Ian D. Reid, and Eric Hayman. Camera calibration and the search for infinity. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, volume 1, pages 510–517, 1999.

[25] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 339–343. IEEE Computer Society Press, August 1996.

[26] W. Hong, A.Y. Yang, K. Huang, and Y. Ma. On symmetry and multiple-view geometry: Structure, pose, and calibration from a single image. *International Journal of Computer Vision*, 60(3):241–265, 2004.

[27] G. Jiang and L. Quan. Detection of concentric circles for camera calibration. In *Proceedings of 10th IEEE International Conference on Computer Vision*, volume 1, pages 333–340, 2005.

[28] F. Kahl and B. Triggs. Critical motions in euclidean structure from motion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, volume 2, pages 366–372, June 1999.

[29] C. Lei, F. Wu, Z. Hu, and H.T. Tsui. A new approach to solving kruppa equations for camera self-calibration. In *Proceedings of the 16th International Conference on Pattern Recognition, Québec city, Canada*, volume 2, pages 308–311, 2002.

[30] YF Li and RS Lu. Uncalibrated Euclidean 3-D reconstruction using an active vision system. *IEEE Transactions on Robotics and Automation*, 20(1):15–25, 2004.

[31] H.C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[32] M.I.A. Lourakis and R. Deriche. Camera self-calibration using the singular value decomposition of the fundamental matrix. In *Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan*, volume 1, pages 403–408, January 2000.

[33] Q.T Luong. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of Computer Vision*, 22(3):261–289, 1993.

[34] Q.T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. *Report RR-1894, INRIA*, 1993.

[35] P.R.S. Mendon0Ł4a and R. Cipolla. A simple technique for self-calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, volume 1, pages 500–505, June 1999.

[36] D. Nistér. Untwisting a projective reconstruction. *International Journal of Computer Vision*, 60(2):165–183, 2004.

[37] M. Pollefeys and L. van Gool. Stratified self-calibration with the modulus constraint. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 21(8):707–724, 1999.

[38] J. Ponce, K. McHenry, T. Papadopoulo, M. Teillaud, and B. Triggs. On the absolute quadratic complex and its application to autocalibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Washington, DC, USA*, volume 1, pages 780–787, 2005.

[39] J. Porrill and S. Pollard. Curve Fitting and Stereo Calibration. In *Proceedings of the British Machine Vision Conference*, pages 37–42, 1990.

[40] C. Rothwell, G. Csurka, and O. Faugeras. A comparison of projective reconstruction methods for pairs of views. *Computer Vision and Image Understanding*, 68(1):37–58, 1997.

[41] M. Sainz, N. Bagherzadeh, and A. Susin. Recovering 3D metric structure and motion from multiple uncalibrated cameras. In *IEEE Proceedings of International Conference on Information Technology: Coding and Computing*, pages 268–273, 2002.

[42] Y. Shang, Q. Yu, and X. Zhang. Analytical method for camera calibration from a single image with four coplanar control lines. *Applied Optics*, 43(28):5364–5369, 2004.

[43] A. Shashua. Projective structure from two uncalibrated images : Structure from motion and recognition. Technical Report A.I. Memo No. 1363, Massachusetts Institute of Technology, September 1992.

[44] Peter Sturm. A case against kruppa's equations for camera self-calibration. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1199–1204, October 2000.

[45] D. Svedberg and S. Carlsson. Calibration, pose and novel views from single images of constrained scenes. *Pattern Recognition Letters*, 21(13-14):1125–1133, 2000.

[46] H. Teramoto and G. Xu. Camera calibration by a single image of balls: From conics to the absolute conic. In *Proccedings of 5th Asian Conference on Computer Vision*, pages 499–506, 2002.

[47] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, June 1997.

[48] R.Y. Tsai. A efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, 1986.

[49] R.Y. Tsai. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the Shelf TV Cameras and Lenses. *Journal of Robotics and Automation*, 3(4):323–334, 1987.

[50] A. Valdes, J.I. Ronda, and G. Gallego. The absolute line quadric and camera autocalibration. *International Journal of Computer Vision*, 66(3):283–303, 2006.

[51] Z. Wang and B. Boufama. Using stereo geometry towards accurate 3d reconstruction. In *IEEE International Conference on Electro/Information Technology*, pages 134–140, 2009.

[52] A. Whitehead and G. Roth. Estimating intrinsic camera parameters from the fundamental matrix using an evolutionary approach. *EURASIP Journal of Applied Signal Processing*, 8(2004), 2004.

[53] M. Wilczkowiak, E. Boyer, and P. Sturm. Camera calibration and 3D reconstruction from single images using parallelepipeds. In *International Conference on Computer Vision*, pages 142–148, 2001.

[54] M. Wilczkowiak, E. Boyer, and P. Sturm. 3D modelling using geometric constraints: A parallelepiped based approach. *Lecture Notes in Computer Science*, pages 221–236, 2002.

[55] M. Wilczkowiak, P. Sturm, and E. Boyer. Using geometric constraints through parallelepipeds for calibration and 3D modeling. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, pages 194–207, 2005.

[56] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.

[57] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *International Conference on Computer Vision*, volume 1, pages 666–673, 1999.

[58] Z. Zhang. A flexible new technique for camera calibration. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[59] Y.J. Zhou and X.J. Kou. A practical iterative two-view metric reconstruction with uncalibrated cameras. *Journal of Zhejiang University-Science A*, 8(10):1614–1623, 2007.

Vita Auctoris

Zhuo Wang was born in 1985 in Nantong, Jiangsu Province, People's Republic of China. He earned his Bachelor of Eletrical Engineering in 2007 from Nanjing University of Science and Technology, Nanjing. Zhuo Wang is currently a candidate for the Master's degree under the supervision of Dr. Boubakeur Boufama in the School of Computer Science at the University of Windsor, Ontario, Canada and expecting to graduate in Fall 2009.