CHAPTER III

Theory

3.1 Camera Model

Hartley and Zisserman (2003: 153) said "A camera is a mapping between the 3D world (object space) and a 2D image". Then the camera models are matrices with particular properties that represent the camera mapping.

There are a various types of camera model. In this research, basic pinhole model has been used. By consider the central projection of points in space onto a plane. Let the center of projection be the origin of a Euclidean coordinate system, and consider a plane $Z = f$, which is called the image plane or focal plane. Under the pinhole camera model, a point in space with coordinates $\mathbf{X} = (x_c, y_c, z_c)^T$ in camera coordinate frame is mapped to the point on the image plane where a line joining the point $\mathbf{X}$ to the center of projection meets the image plane. This is shown in figure 3.1.

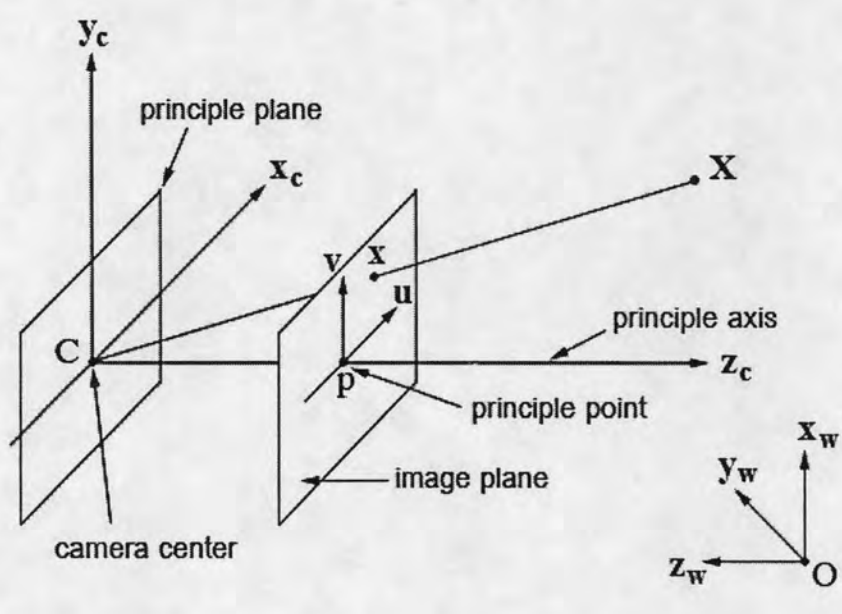

Figure 3.1: Pinhole camera geometry

The center of projection is called the camera center. It is also known as the optical center. The line from the camera center perpendicular to the image plane is called the principle axis or principle ray of the camera and the point where the principle axis meets the image plane is called the principle point. The plane through the camera center parallel to the image plane is called the principle plane of camera.

By similar triangles as shown in figure 3.2, one quickly computes that the point (x_c, y_c, z_c) is mapped to the point $(fx_c / z_c, fy_c / z_c, f)$ on the image plane. Then 2D image point, which obtained by ignoring the final image coordinate, is $\mathbf{x}_c = (fx_c / z_c, fy_c / z_c)$. This is a mapping from Euclidean 3-space $\mathbb{R}^3$ to Euclidean 2-space $\mathbb{R}^2$.

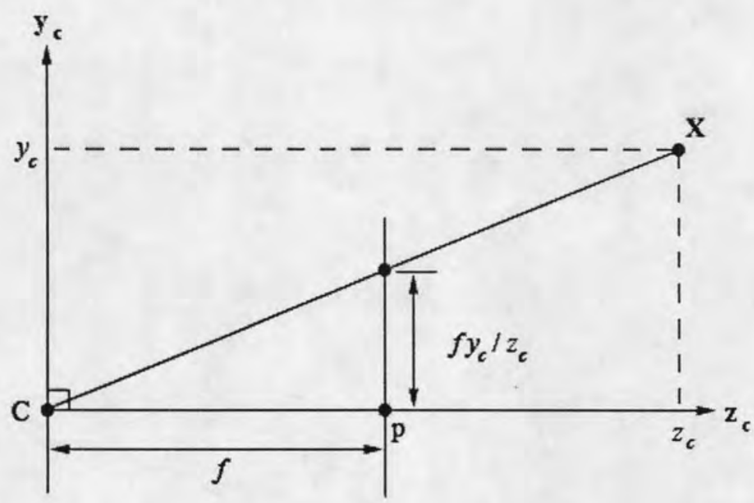

Figure 3.2: Similar triangles of pinhole camera geometry

If the world and image points are represented by homogeneous vectors, then central projection is very simply expressed as a line mapping between their homogeneous coordinates. In particular, this mapping may be written in terms of matrix multiplication as:

$$z_c \begin{bmatrix} fx_c / z_c \\ fy_c / z_c \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{3.1}$$

The expression (3.1) assumed that the origin of coordinates in the image plane is at the principle point, $\mathbf{p}$. In practice, it may not be, so that in general the origin of coordinate in the image plane is at $\mathbf{p}' = (u_0, v_0)$ as shown in figure 3.3, and expression (3.1) become

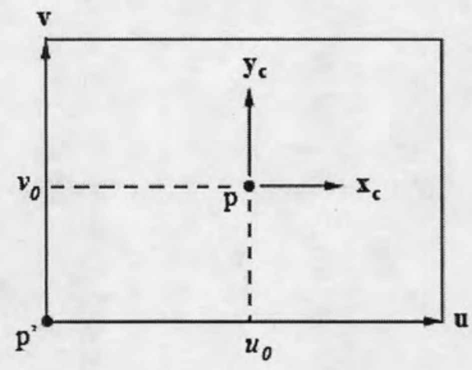

Figure 3.3: Image and camera coordinate system

$$z_c\begin{bmatrix}(fx_c/z_c)+u_0\\(fy_c/z_c)+v_0\\1\end{bmatrix}=\begin{bmatrix}f&0&u_0&0\\0&f&v_0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}x_c\\y_c\\z_c\\1\end{bmatrix}\tag{3.2}$$

Now, writing

$$\mathbf{K}=\begin{bmatrix}f&0&u_0\\0&f&v_0\\0&0&1\end{bmatrix}\tag{3.3}$$

The (3.2) has the concise form

$$\mathbf{x}=\mathbf{K}[\mathbf{I}\mid\mathbf{0}]\mathbf{X}_c\tag{3.4}$$

The matrix $\mathbf{K}$ is called the camera calibration matrix. The pinhole camera model just derived assumes that the image coordinates are Euclidean coordinate having equal scales in both axial directions. In the case of CCD cameras, there is the additional possibility of having non-square pixels. If the image coordinates are measured in pixels, then this has the extra effect of introducing unequal scale factors in each direction. In particular if the number of pixels per unit distance in image coordinates are m_x and m_y in the x and y directions, then the camera calibration can be written as

$$\mathbf{K}=\begin{bmatrix}\alpha_x&s&u_0\\0&\alpha_y&v_0\\0&0&1\end{bmatrix}\tag{3.5}$$

where

$$\alpha_x = fm_x \tag{3.6}$$

$$\alpha_y = fm_y \tag{3.7}$$

s is referred to as the skew parameter. The skew parameter will be zero for most normal cameras. However, in certain unusual instances, such as when lens's axis is not exactly perpendicular to the image plane, it can tack non-zero values.

In general, points in space will be expressed in terms of a different Euclidean coordinate frame, known as the world coordinate frame. The two coordinate frames are related via rotation and translation. See figure 3.4. If $\mathbf{X}_w$ is an inhomogeneous 3-vector representing the coordinates of a point in the world coordinate frame, and $\mathbf{X}_c$ represents the same point in the camera coordinate frame, then

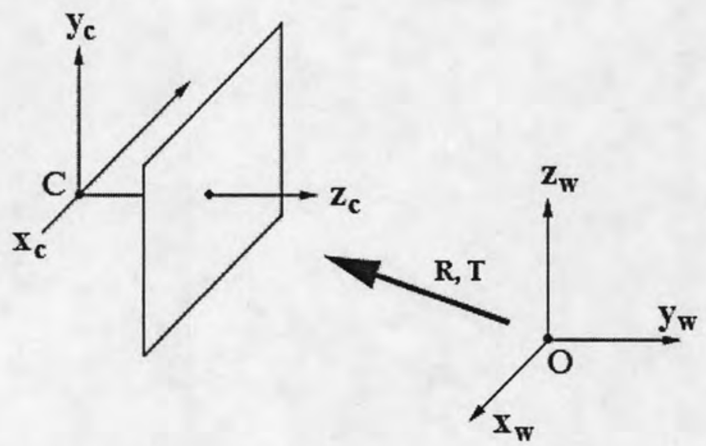

Figure 3.4: The transformation between the world and camera coordinate frame

$$\mathbf{X}_c = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{X}_w \tag{3.8}$$

Where $\tilde{\mathbf{C}}$ represents the coordinates of the camera center in the world coordinate frame and $\mathbf{R}$ is 3×3 rotation matrix representing the orientation of the camera coordinate frame. The expression (3.8) can be written in terms of translation vector $\mathbf{t}$ as:

$$\mathbf{X}_c = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{X}_w \tag{3.9}$$

where

$$\mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}} \tag{3.10}$$

Putting this together with (3.4) leads to the formula

$$\mathbf{x} = \mathbf{KR}\left[\mathbf{I} \mid -\tilde{\mathbf{C}}\right]\mathbf{X}_w \tag{3.11}$$

or

$$\mathbf{x} = \mathbf{K}\left[\mathbf{R} \mid \mathbf{t}\right]\mathbf{X}_w \tag{3.12}$$

This is the general mapping given by a pinhole camera.

given

$$\mathbf{P} = \mathbf{K}\left[\mathbf{R} \mid \mathbf{t}\right] = \mathbf{KR}\left[\mathbf{I} \mid -\tilde{\mathbf{C}}\right] \tag{3.13}$$

then

$$\mathbf{x} = \mathbf{P}\mathbf{X}_w \tag{3.14}$$

Matrix $\mathbf{P}$ is called camera matrix. It has 11 degrees of freedom: 5 for $\mathbf{K}$ (the elements α_x, α_y, s, u_0, v_0), 3 for $\mathbf{R}$ and 3 for $\tilde{\mathbf{C}}$ (or $\mathbf{t}$). The parameters contained in $\mathbf{K}$ are called the internal camera parameters, or the intrinsic parameters. The parameters of $\mathbf{R}$ and $\tilde{\mathbf{C}}$ which relate the camera orientation and position to a world coordinate system are called the external parameters or the extrinsic parameters. These camera parameters can be evaluated in the camera calibration process.

3.2 Lens Distortion

The assumption throughout previous section has been that a linear model is an accurate model of image process. Thus the world point, image point and optical center are collinear, and world lines are imaged se lines and so on. For real (non-pinhole) lenses this assumption will not hold. The most important deviation is generally a lens distortion. In practical this error becomes more significant as the focal length of the lens decrease. The effect of lens distortion is shown in figure 3.5.

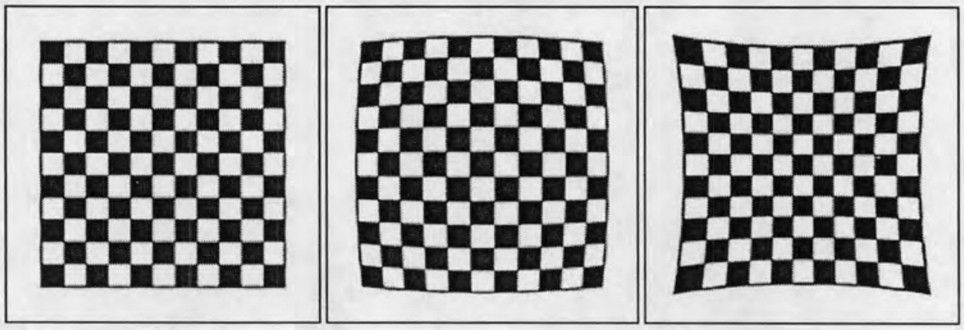

Figure 3.5: (a) is an image of squares pattern that have been obtained under a perfect linear lens. (b) and (c) are images that have been obtained under a real lens.

There are two type of lens distortion including tangential distortion and radial distortion. However, Tsai (1987) shows that only radial distortion needs to be considered because the tangential distortion is very small effect to deviation and it can be neglected.

The cure of this distortion is to correct the image measurements to those that would have been obtained under a perfect linear camera action. The camera is then effectively again a linear device.

Let $\mathbf{x_u} = \begin{bmatrix} x_u & y_u & 1 \end{bmatrix}^T$ is a homogeneous 3-vector represents the ideal image position which obeys linear projection of world point, whose coordinates are $\mathbf{X_c}$ measured in camera coordinate system.

$$\mathbf{x_u} = \begin{bmatrix} \mathbf{I} \mid \mathbf{0} \end{bmatrix} \mathbf{X_c} \tag{3.15}$$

The actual image position (after radial distortion) $\mathbf{x_d} = \begin{bmatrix} x_d & y_d & 1 \end{bmatrix}^T$ is related to the ideal image position by radial displacement as:

$$\mathbf{x_u} = F(r_d)\mathbf{x_d} \tag{3.16}$$

or

$$\mathbf{x_d} = G(r_u)\mathbf{x_u} \tag{3.17}$$

when

$$r_d = \sqrt{x_d^2 + y_d^2} \tag{3.18}$$

$$r_u = \sqrt{x_u^2 + y_u^2} \tag{3.19}$$

$F(r_d)$ is a radial distortion factor, which is a function of r_d.

$G(r_u)$ is a radial distortion factor, which is a function of r_u.

In order to obtain 3-D position of the object, the actual image positions of the object are known and the ideal image positions need to be calculated. It is difficult to use (3.17) for calculation of the ideal image positions from the given actual image positions. The polynomial of degree 5 need to be solved and it will consume a lot of computational time. Thus, (3.16) has been used in this research.

The function $F(r_d)$ is only defined for positive values of r_d and $F(0)=1$. An approximation to an arbitrary function $F(r_d)$ may be given by a Taylor expansion as:

$$F(r_d) = 1 + k_1 r_d + k_2 r_d^2 + k_3 r_d^3 + \ldots \tag{3.20}$$

The coefficients of radial correction ($k_1, k_2, k_{3,\ldots}$) are considered part of the intrinsic parameter of the camera. They can be evaluated in camera calibration process.

3.3 Camera Calibration

Camera calibration means to compute the camera intrinsic and extrinsic parameters which relate world point to the 2-D image point. Many methods have been developed for camera calibration. They can be classified in to two major categories, photogrammetric calibration and self-calibration (auto-calibration).

Photogrammetric calibration, which has been used in this research, is performed by observing a calibration pattern whose geometries in 3D space are known accurately. The corresponding 2-D coordinates obtained from each camera, are used to compute the camera parameters.

Self-calibration methods do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides constraints to intrinsic parameters. The correspondences between images, which are captured by the same camera in different view point, are sufficient to recover both intrinsic and extrinsic parameters. 3-D pose can be reconstructed up to a similarity. However, self-calibration method cannot always evaluate reliable results because there are many parameters to be estimated.

3.3.1 Simple Photogrammetric Calibration

By giving a number of world points whose coordinates in the world coordinate are known accurately, with corresponding image coordinates. From (3.14), a relationship can be derived as:

$$\begin{bmatrix} \mathbf{0}^{\mathrm{T}} & -\mathbf{X}_{\mathrm{i}}^{\mathrm{T}} & v\mathbf{X}_{\mathrm{i}}^{\mathrm{T}} \\ \mathbf{X}_{\mathrm{i}}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & -u\mathbf{X}_{\mathrm{i}}^{\mathrm{T}} \\ -v\mathbf{X}_{\mathrm{i}}^{\mathrm{T}} & u\mathbf{X}_{\mathrm{i}}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \end{bmatrix} \begin{pmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{pmatrix} = \mathbf{0} \tag{3.21}$$

Where each $\mathbf{p}^j$ is a 4-vector, the j-th row of camera matrix $\mathbf{P}$. From (3.21), only 2 equations are linearly independent and choose to use:

$$\begin{bmatrix} \mathbf{0}^{\mathrm{T}} & -\mathbf{X}_i^{\mathrm{T}} & v_i\mathbf{X}_i^{\mathrm{T}} \\ \mathbf{X}_i^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & -u_i\mathbf{X}_i^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{bmatrix} = \mathbf{0} \tag{3.22}$$

From a set of n point correspondences, $2n$ equations can be obtained by stacking up the equation (3.22) for each correspondence as:

$$\begin{bmatrix} \mathbf{0}^{\mathrm{T}} & -\mathbf{X}_1^{\mathrm{T}} & v_1\mathbf{X}_1^{\mathrm{T}} \\ \mathbf{X}_1^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & -u_1\mathbf{X}_1^{\mathrm{T}} \\ \vdots & \vdots & \vdots \\ \mathbf{0}^{\mathrm{T}} & -\mathbf{X}_n^{\mathrm{T}} & v_n\mathbf{X}_n^{\mathrm{T}} \\ \mathbf{X}_n^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & -u_n\mathbf{X}_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{p}^3 \end{bmatrix} = \mathbf{0} \tag{3.23}$$

Solving (3.23) with SVD or Pseudo-invert matrix, then camera matrix has been evaluated. The camera calibration matrix $\mathbf{K}$ and matrix $[\mathbf{R} \mid \mathbf{t}]$ can be evaluated from camera matrix $\mathbf{P}$ by using RQ decomposition.

3.3.2 Tsai's Calibration Method

Tsai (1987) proposes a high accuracy photogrammetric calibration method. Tsai uses monoview with coplanar or non-coplanar set of world points, of the known pre-specify object to evaluate camera parameters.

Tsai uses a camera model which has zero skew parameter. This method relates world coordinate system and image coordinate using image size and number of CCD cells as equation (3.26) and (3.27).

$$u = fs_x d_x^{-1} \frac{N_{cx}}{N_{fx}} x_d + u_0 \tag{3.24}$$

$$v = fd_y^{-1} y_d + v_0 \tag{3.25}$$

where

s_x is the uncertainty image scale factor.

d_x is center to center distance between adjacent sensor elements in x direction.

d_y is center to center distance between adjacent CCD sensor in y direction.

N_{cx} is number of sensor elements in the x direction.

N_{fx} is number of pixels in a line as sampled by the computer.

According to lens distortion, Tsai considers only radial distortion. He model his radial distortion by using (3.16) with the first order of Taylor expansion.

$$\mathbf{x_u} = \mathbf{x_d}\left(1 + k_1 r_d^2\right) \tag{3.26}$$

$$r_d = \sqrt{x_d^2 + y_d^2} \tag{3.27}$$

From (3.24) and (3.25), parameter α_x and α_y of calibration matrix $\mathbf{K}$ in (3.5), can be written as:

$$\alpha_x = fs_x d_x^{-1} \frac{N_{cx}}{N_{fx}} \tag{3.28}$$

$$\alpha_y = fd_y^{-1} \tag{3.29}$$

The intrinsic parameters, which obtained by Tsai's calibration method, consist of f, s_x, k_1, u_0 and v_0. Tsai usually set the value of u_0 and v_0 to the center pixel of frame memory.

3.3.3 Zhang's Calibration Method

Zhang (1998) uses multiple views with coplanar set of world points, of the known pre-specify object to evaluate the camera parameters. By assume that coplanar points are on plane $Z = 0$ in world coordinate system. Thus,

$$\beta \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \tag{3.30}$$

where

β is arbitrary scale factor.

$\mathbf{r}_i$ is i-th column of rotation matrix $\mathbf{R}$.

The plane $Z = 0$ is assumed to be the first image plane and image plane of the capturing image is assumed to be the second image plane. Given $\mathbf{M} = \begin{bmatrix} x_w & y_w & 1 \end{bmatrix}^T$ and $\mathbf{m} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$ are homogeneous 3-vector represent the coordinates of world point on first and second image plane respectively. These two planes are relate together with homography $\mathbf{H}$, which defined up to a non-zero scale factor.

$$\beta \mathbf{m} = \mathbf{HM} \tag{3.31}$$

Homography can be evaluated by giving an enough world points and their corresponding image points. From (3.14) and (3.31), relation between homography and camera parameters can be obtained from:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{K}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \tag{3.32}$$

where

λ is an arbitrary scalar value

$\mathbf{h}_i = \begin{bmatrix} h_{i1} & h_{i2} & h_{i3} \end{bmatrix}^T$ is i-th column of homography $\mathbf{H}$.

Using orthonormal constraint between $\mathbf{r}_1$ and $\mathbf{r}_2$ with (3.32), following equations have been used to evaluate calibration matrix $\mathbf{K}$.

$$\mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0 \tag{3.33}$$

$$\mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 - \mathbf{h}_2^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0 \tag{3.34}$$

$$\mathbf{K}^{-T}\mathbf{K}^{-1} = \begin{bmatrix} \frac{1}{\alpha_x^2} & -\frac{s}{\alpha_x^2\alpha_y} & \frac{v_0 s - u_0\alpha_y}{\alpha_x^2\alpha_y} \\ -\frac{s}{\alpha_x^2\alpha_y} & \frac{s^2}{\alpha_x^2\alpha_y^2} + \frac{1}{\alpha_y^2} & -\frac{s\left(v_0 s - u_0\alpha_y\right)}{\alpha_x^2\alpha_y^2} - \frac{v_0}{\alpha_y^2} \\ \frac{v_0 s - u_0\alpha_y}{\alpha_x^2\alpha_y} & -\frac{s\left(v_0 s - u_0\alpha_y\right)}{\alpha_x^2\alpha_y^2} - \frac{v_0}{\alpha_y^2} & \frac{\left(v_0 s - u_0\alpha_y\right)^2}{\alpha_x^2\alpha_y^2} + \frac{v_0^2}{\alpha_y^2} + 1 \end{bmatrix} \tag{3.35}$$

$\mathbf{K}^{-T}\mathbf{K}^{-1}$ is called image of absolute conic which is a symmetric matrix. It can be defined by 6-column vector $\mathbf{b}$.

given

$$\mathbf{B} = \mathbf{K}^{-T}\mathbf{K}^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \tag{3.36}$$

$$\mathbf{b} = \begin{bmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{bmatrix}^T \tag{3.37}$$

Now, writing

$$\mathbf{h}_i^T\mathbf{B}\mathbf{h}_j = \mathbf{v}_{ij}^T\mathbf{b} \tag{3.38}$$

where

$$\mathbf{v}_{ij} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}^T \tag{3.39}$$

from (3.33) and (3.34)

$$\mathbf{V}\mathbf{b} = \begin{bmatrix} \mathbf{v}_{12}^T \\ \left(\mathbf{v}_{11} - \mathbf{v}_{22}\right)^T \end{bmatrix}\mathbf{b} = 0 \tag{3.40}$$

For n images of calibration pattern ($n \geq 3$), the n homographies can be obtained. Then $2n \times 6$ matrix $\mathbf{V}$ can be formed by stacking up (3.40) for each corresponding homography. Solve these set of equations to obtain $\mathbf{B}$, which relates to intrinsic parameters as following:

$$v_0 = \frac{\left(B_{12}B_{13} - B_{11}B_{23}\right)}{\left(B_{11}B_{22} - B_{12}^2\right)} \tag{3.41}$$

$$\zeta = B_{33} - \frac{B_{13}^2 + v_0\left(B_{12}B_{13} - B_{11}B_{23}\right)}{B_{11}} \tag{3.42}$$

$$\alpha_x = \sqrt{\frac{\zeta}{B_{11}}} \tag{3.43}$$

$$\alpha_y = \sqrt{\frac{\zeta B_{11}}{\left(B_{11}B_{22} - B_{12}^2\right)}} \tag{3.44}$$

$$s = -\frac{B_{12}\alpha_x^2\alpha_y}{\zeta} \tag{3.45}$$

$$u_0 = \frac{s v_0}{\alpha_x} - \frac{B_{13}\alpha_x^2}{\zeta} \tag{3.46}$$

from (3.32)

$$\mathbf{r}_1 = \lambda\mathbf{K}^{-1}\mathbf{h}_1 \tag{3.47}$$

$$\mathbf{r}_2 = \lambda\mathbf{K}^{-1}\mathbf{h}_2 \tag{3.48}$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \tag{3.49}$$

$$\mathbf{t} = \lambda\mathbf{K}^{-1}\mathbf{h}_3 \tag{3.50}$$

where

$$\lambda = \frac{1}{\left\|\mathbf{K}^{-1}\mathbf{h}_1\right\|} = \frac{1}{\left\|\mathbf{K}^{-1}\mathbf{h}_2\right\|} \tag{3.51}$$

So, from Zhang's calibration method, the equation (3.41) to (3.51) can be used to obtained both intrinsic and extrinsic parameters. However, the rotation matrix may not be satisfied orthonormal constraint because of the noise. Zhang suggests that to use Levenberg-Marquardt method for optimizing these parameters. The minimization function, for Levenberg-Marquardt optimization, is the summation of geometric distances between measured image points and reprojected image points.

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\mathbf{x}_{ij}-\hat{\mathbf{x}}(\mathbf{K},\mathbf{R}_i,\mathbf{t}_i,\mathbf{X}_j)\right\|^2 \quad (3.52)$$

where

n is number of images of calibration pattern.

m is number of points on a calibration pattern.

$\mathbf{R}_i$ is rotation matrix which corresponding to image i

t_i is translation vector which corresponding to image i

$\hat{\mathbf{x}}(\mathbf{K},\mathbf{R}_i,\mathbf{t}_i,\mathbf{X}_j)$ is projected image point of $\mathbf{X}_j$ on image i

According to lens distortion, Zhang models lens distortion by consider only radial distortion can be expressed as the equation (3.17) with second order Taylor expansion. But in this research, instead of using (3.17), we use (3.16) instead. Linear equation system for solving radial correction coefficients is:

$$\begin{bmatrix}(u_d-u_0)(x_d^2+y_d^2) & (u_d-u_0)(x_d^2+y_d^2)^2\\ (v_d-v_0)(x_d^2+y_d^2) & (v_d-v_0)(x_d^2+y_d^2)^2\end{bmatrix}\begin{bmatrix}k_1\\ k_2\end{bmatrix}=\begin{bmatrix}(u_u-u_d)\\ (v_u-v_d)\end{bmatrix} \quad (3.53)$$

Where (u_u,v_u) and (u_d,v_d) are image coordinates which corresponding to $\mathbf{x}_u$ and $\mathbf{x}_d$ respectively. If there are n images of calibration pattern used in this calibration method and each image contains m points of correspondence, then $2mn$ equations can be obtained and the Least-Squares method can be used for solving k_1 and k_2.

3.4 3-D Reconstruction

3-D reconstruction is to find 3-D world coordinates of an object from the given corresponding image coordinates which are captured from group of cameras. Normally, 3-D world coordinates of object can be reconstructed from the intersection in space of back projection rays. Each ray passes through the optical center and the 2-D image point in the image plane of the corresponding camera. These rays will intersect at the same point as show in figure 3.6. Due to the presence of noise, these rays are not guarantee to intersect at a single point as show in figure 3.7. There are some commonly-suggested methods to overcome this problem as:

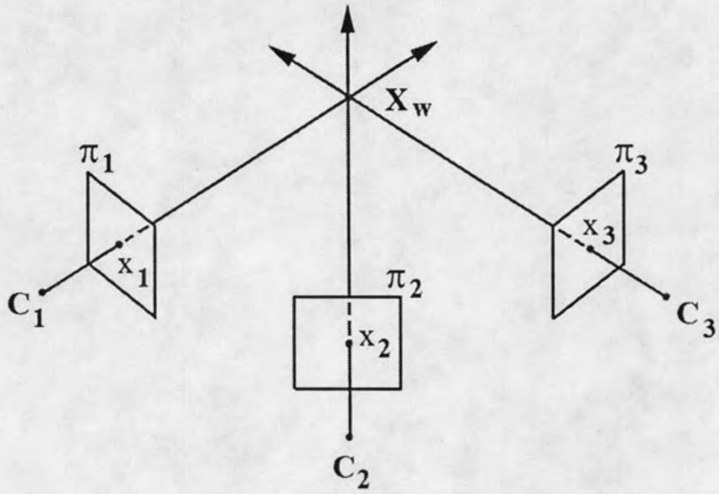

Figure 3.6: Ideal reconstruction.

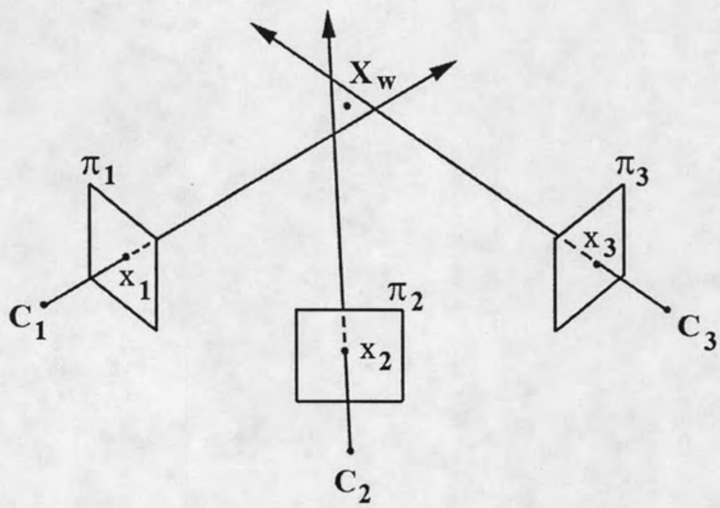

Figure 3.7: Real world reconstruction.

3.4.1 Midpoint

This method uses average of midpoint of common perpendicular to two back projection rays. The equation of back projection ray can be found:

from (3.12) and (3.14)

$$\lambda \mathbf{x} = \mathbf{P}\mathbf{X}_w = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\mathbf{X}_w$$

then

$$\mathbf{X}_w(\lambda) = \lambda\left[\mathbf{R}^T \mid -\mathbf{R}^T\mathbf{t}\right]\begin{bmatrix}\mathbf{K}^{-1}\mathbf{x} \\ 1\end{bmatrix} \tag{3.54}$$

where

λ is an arbitrary scalar value.

From figure 3.6 and 3.7, back projection ray is the ray that starts from camera center, which can be obtained from (3.10), and passes through an image point on image plane. Let $L(t)$ is a back projection ray which is a function of t, $\mathbf{x}_\mathrm{w} = \mathbf{X}_\mathrm{w}(1)$ is homogeneous 4-vector represent 3-D world coordinate of image point on image plane, $\tilde{\mathbf{x}}_\mathrm{w}$ is 3-vector corresponding to $\mathbf{x}_\mathrm{w}$.

$$L(t) = \frac{\tilde{\mathbf{x}}_\mathrm{w} - \tilde{\mathbf{C}}}{\left\| \tilde{\mathbf{x}}_\mathrm{w} - \tilde{\mathbf{C}} \right\|} t + \tilde{\mathbf{C}} \tag{3.55}$$

3.4.2 Linear Triangulation

Given $\mathbf{P}_i$, the camera matrix of i-th camera, and $\mathbf{x}_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T$ is homogeneous 3-vector represents 2-D image coordinates of object which measured from i-th camera. From (3.14), we can write

$$\mathbf{x}_i \times \left(\mathbf{P}_i \mathbf{X}_\mathrm{w} \right) = 0 \tag{3.56}$$

then we will have now,

$$\left. \begin{aligned} u_i \left(\mathbf{p}_i^{3T} \mathbf{X}_\mathrm{w} \right) - \left(\mathbf{p}_i^{1T} \mathbf{X}_\mathrm{w} \right) = 0 \\ v_i \left(\mathbf{p}_i^{3T} \mathbf{X}_\mathrm{w} \right) - \left(\mathbf{p}_i^{2T} \mathbf{X}_\mathrm{w} \right) = 0 \\ u_i \left(\mathbf{p}_i^{2T} \mathbf{X}_\mathrm{w} \right) - v_i \left(\mathbf{p}_i^{1T} \mathbf{X}_\mathrm{w} \right) = 0 \end{aligned} \right\} \tag{3.57}$$

where

$\mathbf{p}_i^{jT}$ is column vector of j-th row of camera matrix $\mathbf{P}_i$.

From (3.57), it is easily to show that only two equations are linearly independent. There are three variables need to be solved. For n cameras ($n \geq 2$), homogeneous equation system can then be performed as:

$$\mathbf{A}\mathbf{X}_\mathrm{w} = \mathbf{0} \tag{3.58}$$

where

$$\mathbf{A} = \begin{bmatrix} u_1\mathbf{p}_1^{3T} - \mathbf{p}_1^{1T} \\ v_1\mathbf{p}_1^{3T} - \mathbf{p}_1^{2T} \\ u_2\mathbf{p}_2^{3T} - \mathbf{p}_2^{1T} \\ v_2\mathbf{p}_2^{3T} - \mathbf{p}_2^{2T} \\ \vdots \\ u_n\mathbf{p}_n^{3T} - \mathbf{p}_n^{1T} \\ v_n\mathbf{p}_n^{3T} - \mathbf{p}_n^{2T} \end{bmatrix} \quad (3.59)$$

Equation (3.58) can be solved by Least-Squares method. The solutions are unit singular vector which corresponding to smallest singular value of matrix $\mathbf{A}$.

3.4.3 Bundle Adjustment

Due to the presentation of noise in image measurement, exact solution of equation 3.14 is not simply obtained. In this case the Maximum Likelihood (ML) solution has been used to obtain the best solution of equation 3.14 by assuming that the measurement noise is Gaussian. Projection matrix $\hat{\mathbf{P}}_i$ and 3-D points $\hat{\mathbf{X}}_j$, which project exactly to image points $\hat{\mathbf{x}}_{ij}$ as $\hat{\mathbf{x}}_{ij} = \hat{\mathbf{P}}_i\hat{\mathbf{X}}_j$, wished to estimate and also minimize the image distance between the reprojected point and detected (measured) image points $\hat{\mathbf{x}}_{ij}$ for every view in which the 3-D point appears, i.e.

$$\min_{\hat{\mathbf{P}}_i \hat{\mathbf{X}}_j} \sum_{ij} d\left(\hat{\mathbf{P}}_i\hat{\mathbf{X}}_j, \mathbf{x}_{ij}\right)^2 \quad (3.60)$$

where

$d(\mathbf{x},\mathbf{y})$ is the geometric image distance between the homogeneous points $\mathbf{x}$ and $\mathbf{y}$.

This estimation involving minimizing the reprojection error is known as bundle adjustment – it involves adjusting the bundle of rays between each camera center and the set of 3-D points (and equivalently between each 3-D point and the set of camera canters).

Bundle adjustment should generally be used as a final step of any reconstruction algorithm. This method has the advantages of being tolerant of missing data while providing a true ML estimate. At the same time it allows assignment of individual covariances to each measurement and may also be extended to include estimates of priors and constraints on camera parameters or point positions. In short, it would seem to be an ideal algorithm, except for the fact that: (i) It requires a good initialization to be provided, and (ii) it can become an extremely large minimization problem because of the number of parameters involved.

3.5 Circular Hough Transform (CHT)

The Hough transform is a general technique for identifying the locations and orientations of certain types of features in a digital image. Developed by Hough (1962) and patented by IBM, the transform consists of parameterize a description of a feature at any given location in the original image's space. A mesh in the space defined by these parameters is then generated, and at each mesh point a value is accumulated, indicating how well an object generated by the parameters defined at that point fits the given image. Mesh points that accumulate relatively larger values describe features that may be projected back onto the image, fitting to some degree the features actually present in the image.

CHT is uses Hough transform to find circles in image. The circle can be parameterized as:

$$(x-a)^2+(x-b)^2=r^2 \tag{3.61}$$

This model has three parameters: two parameters for the centre of the circle and one parameter for the radius of the circle. If the gradient angle for the edges is available, then this provides a constraint that reduces the number of degrees of freedom and hence the required size of the parameter space. The direction of the vector from the centre of the circle to each edge point is determined by the gradient angle, leaving the value of the radius as the only unknown parameter.

Thus, the parametric equations for a circle in polar coordinates are:

$$x=a+r\cos\theta \tag{3.62}$$

$$y=b+r\sin\theta \tag{3.63}$$

Solving for the parameters of the circle

$$a=x-r\cos\theta \tag{3.64}$$

$$b=y-r\sin\theta \tag{3.65}$$

Now given the gradient angle θ at an edge point (x, y), $\cos\theta$ and $\sin\theta$ can be computed. Note that these quantities may already be available as a by-product of edge detection. The radius can be eliminated from the pair of equations above to yield.

$$b = a\tan\theta - x\tan\theta + y \tag{3.66}$$

Given

$M(a,b)$ is accumulator array corresponding to (a,b).

$G(x,y)$ is gradient magnitude at an edge point (x,y).

$\theta(x,y)$ is angle θ at an edge point (x,y).

Then the Hough Transform algorithm for circle fitting can be described as follows.

i) Quantize the parameter space for the parameters a and b.
ii) Zero the accumulator array $M(a,b)$.
iii) Compute the gradient magnitude $G(x,y)$ and angle $\theta(x,y)$.
iv) For each edge point in $G(x,y)$, increment all points in the accumulator array $M(a,b)$ along the line defined by (3.66)
v) Local maxima in the accumulator array correspond to centers of circles in the image.