Simplified Camera Projection Models 2

Abstract The chapter focuses on the approximation of full perspective projection model. We first present a review on affine camera model, including orthographic projection, weak-perspective projection, and paraperspective projection. Then, under the assumption that the camera is far away from the object with small lateral rotations, we prove that the imaging process can be modeled by quasi-perspective projection. The model is proved to be more accurate than affine model using both geometrical error analysis and experimental studies.

Everything should be made as simple as possible, but not simpler.

Albert Einstein (1879–1955)

2.1 Introduction

The modeling of imaging process is an important issue for many computer vision applications, such as structure from motion, object recognition, pose estimation, etc. Geometrically, a camera maps data from 3D space to a 2D image space. The general camera model used in computer vision is modeled by perspective projection. This is an ideal and accurate model for a wide range of existing cameras. However, the resulting equations from perspective projection are complicated and often nonlinear due to the unknown scaling factor [7]. To simplify computations, researchers have proposed many approximations to the full perspective projection.

The most common approximation includes weak-perspective projection, orthographic projection, and paraperspective projection [1]. These approximations are generalized as affine camera model [5, 9]. Faugeras [2] introduced the properties of projective cameras. Hartley and Zisserman [3] presented a comprehensive survey and in-depth analysis on different camera models. Affine camera is a zero-order (for weak-perspective) or a first-order (for paraperspective) approximation of full perspective projection. It is valid when the depth variation of the object is small compared to the distance from camera to the object. Kanatani *et al.* [4] analyzed a general form of symmetric affine camera model to mimic

G. Wang, Q.M.J. Wu, *Guide to Three Dimensional Structure and Motion Factorization,* Advances in Pattern Recognition,
DOI 10.1007/978-0-85729-046-5_2,

perspective projection and provided the minimal requirements for orthographic, weak perspective, and para-perspective simplification. The model contains two free variables that can be determined through self-calibration.

Affine assumption is widely adopted for the study of structure from motion due to its simplicity. In this chapter, we try to make a trade-off between simplicity of the affine model and accuracy of the full perspective projection model. Assuming that the camera is far away from the object with small lateral rotations, which is similar to affine assumption and is easily satisfied in practice. We propose a quasi-perspective projection model and present an error analysis of different projection models [10]. The model is proved to be more accurate than affine approximation. In the subsequent chapters of this book, we will provide some two-view properties of the model [11] and its application to structure and motion factorization [12].

The remaining part of the chapter is organized as follows. The affine projection model is reviewed in Sect. 2.2. The proposed quasi-perspective model and error analysis are elaborated in Sect. 2.3. Some experimental evaluations on synthetic data are given in Sect. 2.4.

2.2 Affine Projection Model

Under perspective projection, a 3D point $\mathbf{X}_j$ is projected onto an image point $\mathbf{x}_{ij}$ in frame i according to equation

$$\lambda_{ij}\mathbf{x}_{ij} = \mathbf{P}_i\mathbf{X}_j = \mathbf{K}_i[\mathbf{R}_i, \mathbf{T}_i]\mathbf{X}_j \tag{2.1}$$

where λ_{ij} is a non-zero scale factor, commonly denoted as the projective depth; the image point $\mathbf{x}_{ij}$ and space point $\mathbf{X}_j$ are expressed in homogeneous form; $\mathbf{P}_i$ is the projection matrix of the i-th frame; $\mathbf{R}_i$ and $\mathbf{T}_i$ are the corresponding rotation matrix and translation vector of the camera with respect to the world system; $\mathbf{K}_i$ is the camera calibration matrix of the form

$$\mathbf{K}_i = \begin{bmatrix} f_i & \varsigma_i & u_{0i} \\ 0 & \kappa_i f_i & v_{0i} \\ 0 & 0 & 1 \end{bmatrix} \tag{2.2}$$

For some precise industrial CCD cameras, we assume zero skew $\varsigma_i = 0$, known principal point $u_{0i} = v_{0i} = 0$, and unit aspect ratio $\kappa_i = 1$. Then the camera is simplified to have only one intrinsic parameter f_i.

When the distance of an object from a camera is much greater than the depth variation of the object, we may assume affine camera model. Under affine assumption, the last row of the projection matrix is of the form $\mathbf{P}_{3i}^T \simeq [0, 0, 0, 1]$, where '$\simeq$' denotes equality up to scale. Thus a general affine projection matrix for the ith view can be written as

$$\mathbf{P}_{Ai} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_i & \bar{\mathbf{T}}_i \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{2.3}$$

where, $\mathbf{A}_i \in \mathbb{R}^{2\times 3}$ is composed by the upper-left 2×3 submatrix of $\mathbf{P}_i$, $\bar{\mathbf{T}}_i$ is a translation vector. Then, the projection process (2.1) can be simplified by removing the scale factor λ_{ij}.

$$\bar{\mathbf{x}}_{ij} = \mathbf{A}_i \bar{\mathbf{X}}_j + \bar{\mathbf{T}}_i \tag{2.4}$$

Under affine projection, the mapping from space to the image is linear. One attractive attribute of affine camera model is that the mapping is independent of the translation term if relative coordinates are employed in both space and image coordinate frames.

Suppose $\bar{\mathbf{X}}_r$ is a reference point in space and $\bar{\mathbf{x}}_{ir}$ is its image in the ith frame. Then, we have $\bar{\mathbf{x}}_{ir} = \mathbf{A}_i \bar{\mathbf{X}}_r + \bar{\mathbf{T}}_i$. Let us denote

$$\bar{\mathbf{x}}'_{ij} = \bar{\mathbf{x}}_{ij} - \bar{\mathbf{x}}_{ir}, \qquad \bar{\mathbf{X}}'_j = \bar{\mathbf{X}}_j - \bar{\mathbf{X}}_r$$

as the relative image and space coordinates. We can immediately obtain a simplified affine projection equation in terms of relative coordinates.

$$\bar{\mathbf{x}}'_{ij} = \mathbf{A}_i \bar{\mathbf{X}}'_j \tag{2.5}$$

Actually, the translation term $\bar{\mathbf{T}}_i$ is exactly the image of world origin. It is easy to verify that the centroid of a set of space points is projected to the centroid of their images. In practice, we can simply choose the centroid as the reference point, then the translation term vanishes if all the image points in each frame are registered to the corresponding centroid. The affine matrix $\mathbf{A}_i$ has six independent variables which encapsulate both intrinsic and extrinsic parameters of the affine camera. According to RQ decomposition [3], matrix $\mathbf{A}_i$ can be uniquely decomposed into the following form.

$$\mathbf{A}_i = \mathbf{K}_{Ai}\mathbf{R}_{Ai} = \begin{bmatrix} \alpha_{1i} & \zeta_i \\ & \alpha_{2i} \end{bmatrix} \begin{bmatrix} \mathbf{r}_{1i}^T \\ \mathbf{r}_{2i}^T \end{bmatrix} \tag{2.6}$$

where $\mathbf{K}_{Ai}$ is the intrinsic calibration matrix. In accordance to the camera matrix of perspective projection, α_{1i} and α_{2i} are the scaling factors of the two image axes and α_{1i}/α_{2i} is defined as the aspect ratio, ζ_i is the skew factor of the affine camera. For most CCD cameras, we usually assume unit aspect ratio $\alpha_{1i} = \alpha_{2i} = \alpha_i$, and zero skew $\zeta_i = 0$. $\mathbf{R}_{Ai}$ is the rotation matrix, $\mathbf{r}_{1i}^T$ and $\mathbf{r}_{2i}^T$ are the first two rows of the rotation matrix with the constraint

$$\mathbf{r}_{1i}^T \mathbf{r}_{2i} = 0, \qquad \|\mathbf{r}_{1i}\|^2 = \|\mathbf{r}_{2i}\|^2 = 1 \tag{2.7}$$

while the third row of the rotation matrix can always be recovered as $\mathbf{r}_{3i} = \mathbf{r}_{1i} \times \mathbf{r}_{2i}$. From the above analysis, we can easily see that the affine matrix $\mathbf{A}_i$ has six degrees of freedom. Under affine assumption, the camera projection is usually modeled by three special cases, i.e. orthographic projection, weak perspective projection, and para-perspective projection, as shown in Fig. 2.1.

Orthographic projection is the most simple approximation. In this case, it is assumed $\alpha_{1i} = \alpha_{1i} = 1$ and $\zeta_i = 0$. Thus the projection can be modelled as

$$\mathbf{K}_{ortho} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.8}$$

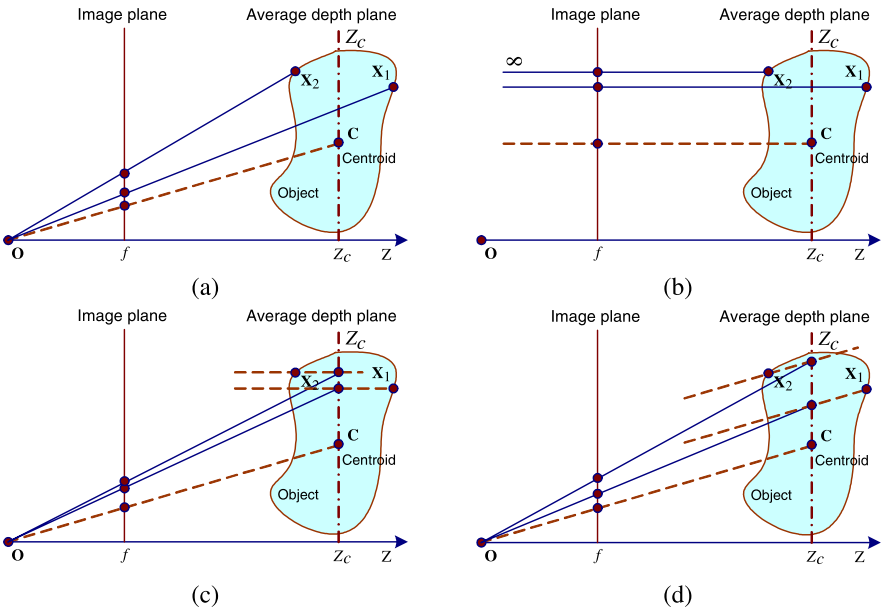

Fig. 2.1 The imaging process of different projection models. (**a**) Perspective projection; (**b**) Orthographic projection; (**c**) Weak-perspective projection; (**d**) Para-perspective projection. $\mathbf{O}$ is the optical center, $Z = f$ is the image plane. $\mathbf{C}$ is the centroid of the object, $Z = Z_c$ is the average depth plane, $\mathbf{X}_1$ and $\mathbf{X}_2$ are two space points on the object

where the subscript index i is omitted. In a weak-perspective projection, the space point is first projected to the average depth plane via orthographic projection, then projected to the image by perspective projection. Thus, the scaling factor is included as $\alpha_{1i} = \alpha_{2i} = \alpha_i$ and $\zeta_i = 0$, which is equivalent to a scaled orthography.

$$\mathbf{K}_{weak} = \alpha \mathbf{K}_{ortho} = \alpha \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.9}$$

Thus, the weak-perspective projection can model the scaling effect caused by depth changes between images. It is suitable for objects with small depth variations. Para-perspective is a more generalized affine model which is a step closer to perspective projection. As shown in Fig. 2.1, the main difference between para-perspective and weak-perspective projection is that the space point is first projected to the average depth plane along the line passing through optical center and the centroid of the object. Thus it not only models the scaling of weak perspective, but also the apparent result of an object moving towards the edge of the image. Please refer to [6, 8] for more details on para-perspective projection. It can be verified that weak-perspective is a zero-order approximation of full perspective projection, while paraperspective is a first-order approximation.

2.3 Quasi-Perspective Projection Model

In this section, we will propose a new quasi-perspective projection model to fill the gap between simplicity of affine camera and accuracy of perspective projection.

2.3.1 Quasi-Perspective Projection

Under perspective projection, the image formation process is shown in Fig. 2.2. In order to ensure that large overlapping part of the object is reconstructed, the camera usually undergoes really small movements across adjacent views, especially for images of a video sequence.

Suppose $O_w - X_w Y_w Z_w$ is a world coordinate system selected on the object to be reconstructed. $O_i - X_i Y_i Z_i$ is the camera coordinate system with O_i being the optical center of the camera. Without loss of generality, we assume that there is a reference camera system $O_r - X_r Y_r Z_r$. Since the world system can be set freely, we align it with the reference frame as illustrated in Fig. 2.2. Therefore, the rotation $\mathbf{R}_i$ of frame i with respect to the reference frame is the same as the rotation of the camera to the world system.

Definition 2.1 (Axial and lateral rotation) The orientation of a camera is usually described by roll-pitch-yaw angles. For the i-th frame, we define the pitch, yaw, and roll as the rotations α_i, β_i, and γ_i of the camera with respect to the X_w, Y_w, and Z_w axes of the world system. As shown in Fig. 2.2, the optical axis of the cameras usually point towards the object. For convenience of discussion, we define γ_i as the axial rotation angle, and define α_i and β_i as lateral rotation angles.

Proposition 2.1 *Suppose the camera undergoes small lateral rotation with respect to the reference frame, then the variation of projective depth λ_{ij} is mainly proportional to the*

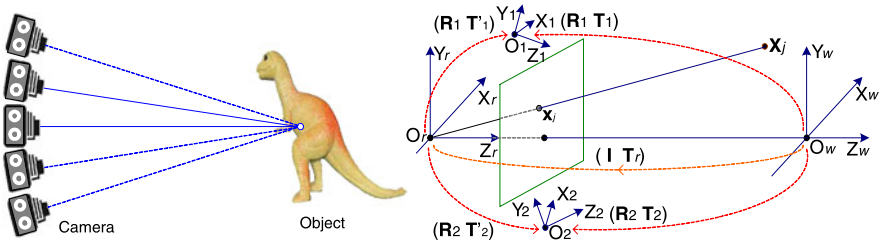

Fig. 2.2 Imaging process of camera take image sequence around an object. (**a**) Camera setup with respect to the object; (**b**) The relationship of world coordinate system and camera systems at different viewpoint

2

depth of the space point. The projective depths of a point at different views have similar trend of variation.

Proof Suppose the rotation matrix and translation vector of the ith frame with respect to the world system are

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{r}_{1i}^T \\ \mathbf{r}_{2i}^T \\ \mathbf{r}_{3i}^T \end{bmatrix}, \qquad \mathbf{T}_i = \begin{bmatrix} t_{xi} \\ t_{yi} \\ t_{zi} \end{bmatrix} \tag{2.10}$$

Then, the projection matrix can be written as

$$\begin{aligned} \mathbf{P}_i &= \mathbf{K}_i[\mathbf{R}_i, \mathbf{T}_i] \\ &= \begin{bmatrix} f_i \mathbf{r}_{1i}^T + \varsigma_i \mathbf{r}_{2i}^T + u_{0i}\mathbf{r}_{3i}^T & f_i t_{xi} + \varsigma_i t_{yi} + u_{0i} t_{zi} \\ \kappa_i f_i \mathbf{r}_{2i}^T + v_{0i}\mathbf{r}_{3i}^T & \kappa_i f_i t_{yi} + v_{0i} t_{zi} \\ \mathbf{r}_{3i}^T & t_{zi} \end{bmatrix} \end{aligned} \tag{2.11}$$

The rotation matrix can be decomposed into rotations around the three axes of the world frame.

$$\begin{aligned} \mathbf{R}_i &= \mathbf{R}(\gamma_i)\mathbf{R}(\beta_i)\mathbf{R}(\alpha_i) \\ &= \begin{bmatrix} \cos\gamma_i, & -\sin\gamma_i, & 0 \\ \sin\gamma_i, & \cos\gamma_i, & 0 \\ 0, & 0, & 1 \end{bmatrix} \begin{bmatrix} \cos\beta_i, & 0, & \sin\beta_i \\ 0, & 1, & 0 \\ -\sin\beta_i, & 0, & \cos\beta_i \end{bmatrix} \begin{bmatrix} 1, & 0, & 0 \\ 0, & \cos\alpha_i, & -\sin\alpha_i \\ 0, & \sin\alpha_i, & \cos\alpha_i \end{bmatrix} \\ &= \begin{bmatrix} \cos\gamma_i\cos\beta_i, & \cos\gamma_i\sin\beta_i\sin\alpha_i - \sin\gamma_i\cos\alpha_i, & \cos\gamma_i\sin\beta_i\cos\alpha_i + \sin\gamma_i\sin\alpha_i \\ \sin\gamma_i\cos\beta_i, & \sin\gamma_i\sin\beta_i\sin\alpha_i + \cos\gamma_i\cos\alpha_i, & \sin\gamma_i\sin\beta_i\cos\alpha_i - \cos\gamma_i\sin\alpha_i \\ -\sin\beta_i, & \cos\beta_i\sin\alpha_i, & \cos\beta_i\cos\alpha_i \end{bmatrix} \end{aligned} \tag{2.12}$$

Inserting (2.11) and (2.12) into (2.1), we have

$$\lambda_{ij} = [\mathbf{r}_{3i}^T, t_{zi}]\mathbf{X}_j = -(\sin\beta_i)x_j + (\cos\beta_i\sin\alpha_i)y_j + (\cos\beta_i\cos\alpha_i)z_j + t_{zi} \tag{2.13}$$

From Fig. 2.2, we know that the rotation angles $\alpha_i, \beta_i, \gamma_i$ of the camera to the world system are the same as those to the reference frame. Under small lateral rotations, i.e., small angles of α_i and β_i, we have

$$\sin\beta_i \ll \cos\beta_i\cos\alpha_i, \qquad \cos\beta_i\sin\alpha_i \ll \cos\beta_i\cos\alpha_i \tag{2.14}$$

Thus, (2.13) can be approximated by

$$\lambda_{ij} \approx (\cos\beta_i\cos\alpha_i)z_j + t_{zi} \tag{2.15}$$

All features $\{x_{ij} | j = 1, \ldots, n\}$ in the ith frame correspond to the same rotation angles $\alpha_i, \beta_i, \gamma_i$, and translation t_{zi}. It is evident from (2.15) that the projective depths of a point

in all frames have similar trend of variation, which are proportional to the value of z_j. Actually, the projective depths have no relation with the axial rotation γ_i. □

Proposition 2.2 *Under small lateral rotations and a further assumption that the distance from the camera to an object is significantly greater than the object depth, i.e., $t_{zi} \gg z_j$, the ratio of $\{\lambda_{ij}|i = 1, \ldots, m\}$ corresponding to any two different frames can be approximated by a constant.*

Proof Let us take the reference frame as an example, the ratio of the projective depths of any frame i to those of the reference frame can be written as

$$\begin{aligned}\mu_i = \frac{\lambda_{rj}}{\lambda_{ij}} &\approx \frac{(\cos\beta_r \cos\alpha_r) z_j + t_{zr}}{(\cos\beta_i \cos\alpha_i) z_j + t_{zi}} \\ &= \frac{\cos\beta_r \cos\alpha_r (z_j/t_{zi}) + t_{zr}/t_{zi}}{\cos\beta_i \cos\alpha_i (z_j/t_{zi}) + 1}\end{aligned} \tag{2.16}$$

where $\cos\beta_i \cos\alpha_i \leq 1$. Under the assumption that $t_{zi} \gg z_j$, the ratio can be approximated by

$$\mu_i = \frac{\lambda_{rj}}{\lambda_{ij}} \approx \frac{t_{zr}}{t_{zi}} \tag{2.17}$$

All features in a frame have the same translation term. Therefore, we can see from (2.17) that the projective depth ratios of two frames for all features have the same approximation μ_i. □

According to Proposition 2.2, we have $\lambda_{ij} = \frac{1}{\mu_i}\lambda_{rj}$. Thus the perspective projection equation (2.1) can be approximated by

$$\frac{1}{\mu_i}\lambda_{rj}\mathbf{x}_{ij} = \mathbf{P}_i\mathbf{X}_j \tag{2.18}$$

Let us denote λ_{rj} as $\frac{1}{\ell_j}$, and reformulate (2.18) to

$$\mathbf{x}_{ij} = \mathbf{P}_{qi}\mathbf{X}_{qj} \tag{2.19}$$

where

$$\mathbf{P}_{qi} = \mu_i\mathbf{P}_i, \mathbf{X}_{qj} = \ell_j\mathbf{X}_j \tag{2.20}$$

We call (2.19) quasi-perspective projection model. Compared with general perspective projection, the quasi-perspective model assumes that projective depths between different frames are defined up to a constant μ_i. Thus, the projective depths are implicitly embedded in the scalars of the homogeneous structure $\mathbf{X}_{qj}$ and the projection matrix $\mathbf{P}_{qi}$, and the difficult problem of estimating the unknown depths is avoided. The model is more general than affine projection model (2.4), where all projective depths are simply assumed to be equal to $\lambda_{ij} = 1$.

2

2.3.2 Error Analysis of Different Models

In the following section, we will give a heuristic analysis on imaging errors of quasi-perspective and affine camera models with respect to the general perspective projection. For simplicity, the subscript 'i' of the frame number is omitted hereafter.

Suppose the intrinsic parameters of the cameras are known, and all images are normalized by the cameras as $\mathbf{K}_i^{-1}\mathbf{x}_{ij} \rightarrow \mathbf{x}_{ij}$. Then, the projection matrices under different projection model can be written as

$$\mathbf{P} = \begin{bmatrix} \mathbf{r}_1^T & t_x \\ \mathbf{r}_2^T & t_y \\ \mathbf{r}_3^T & t_z \end{bmatrix}, \quad \mathbf{r}_3^T = [-\sin\beta, \cos\beta\sin\alpha, \cos\beta\cos\alpha] \tag{2.21}$$

$$\mathbf{P}_q = \begin{bmatrix} \mathbf{r}_1^T & t_x \\ \mathbf{r}_2^T & t_y \\ \mathbf{r}_{3q}^T & t_z \end{bmatrix}, \quad \mathbf{r}_{3q}^T = [0, 0, \cos\beta\cos\alpha] \tag{2.22}$$

$$\mathbf{P}_a = \begin{bmatrix} \mathbf{r}_1^T & t_x \\ \mathbf{r}_2^T & t_y \\ \mathbf{0}^T & t_z \end{bmatrix}, \quad \mathbf{0}^T = [0, 0, 0] \tag{2.23}$$

where $\mathbf{P}$ is the projection matrix of perspective projection, $\mathbf{P}_q$ is that of quasi-perspective assumption, and $\mathbf{P}_a$ is that of affine projection. It is clear that the main difference between the projection matrices lies only in the last row. For a space point $\bar{\mathbf{X}} = [x, y, z]^T$, its projection under different camera models is given by

$$\mathbf{m} = \mathbf{P}\begin{bmatrix} \bar{\mathbf{X}} \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ \mathbf{r}_3^T\bar{\mathbf{X}} + t_z \end{bmatrix} \tag{2.24}$$

$$\mathbf{m}_q = \mathbf{P}_q\begin{bmatrix} \bar{\mathbf{X}} \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ \mathbf{r}_{3q}^T\bar{\mathbf{X}} + t_z \end{bmatrix} \tag{2.25}$$

$$\mathbf{m}_a = \mathbf{P}_a\begin{bmatrix} \bar{\mathbf{X}} \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ t_z \end{bmatrix} \tag{2.26}$$

where

$$u = \mathbf{r}_1^T\bar{\mathbf{X}} + t_x, \, v = \mathbf{r}_2^T\bar{\mathbf{X}} + t_y \tag{2.27}$$

$$\mathbf{r}_3^T\bar{\mathbf{X}} = -(\sin\beta)x + (\cos\beta\sin\alpha)y + (\cos\beta\cos\alpha)z \tag{2.28}$$

$$\mathbf{r}_{3q}^T\bar{\mathbf{X}} = (\cos\beta\cos\alpha)z \tag{2.29}$$

and the nonhomogeneous image points can be denoted as

$$\bar{\mathbf{m}} = \frac{1}{\mathbf{r}_3^T \bar{\mathbf{X}} + t_z} \begin{bmatrix} u \\ v \end{bmatrix} \tag{2.30}$$

$$\bar{\mathbf{m}}_q = \frac{1}{\mathbf{r}_{3q}^T \bar{\mathbf{X}} + t_z} \begin{bmatrix} u \\ v \end{bmatrix} \tag{2.31}$$

$$\bar{\mathbf{m}}_a = \frac{1}{t_z} \begin{bmatrix} u \\ v \end{bmatrix} \tag{2.32}$$

The point $\bar{\mathbf{m}}$ is an ideal image by perspective projection. Let us define $\mathbf{e}_q = |\bar{\mathbf{m}}_q - \bar{\mathbf{m}}|$ as the error of quasi-perspective, and $\mathbf{e}_a = |\bar{\mathbf{m}}_a - \bar{\mathbf{m}}|$ as the error of affine, where '$|\cdot|$' stands for the norm of a vector. Then, we have

$$\begin{aligned} \mathbf{e}_q &= |\bar{\mathbf{m}}_q - \bar{\mathbf{m}}| \\ &= \left| \frac{\mathbf{r}_3^T \bar{\mathbf{X}} + t_z}{\mathbf{r}_{3q}^T \bar{\mathbf{X}} + t_z} \bar{\mathbf{m}} - \bar{\mathbf{m}} \right| = \det\left(\frac{(\mathbf{r}_3^T - \mathbf{r}_{3q}^T)\bar{\mathbf{X}}}{\mathbf{r}_{3q}^T \bar{\mathbf{X}} + t_z} \right) |\bar{\mathbf{m}}| \\ &= \det\left(\frac{-(\sin\beta)x + (\cos\beta \sin\alpha)y}{(\cos\beta\cos\alpha)z + t_z} \right) |\bar{\mathbf{m}}| \end{aligned} \tag{2.33}$$

$$\begin{aligned} \mathbf{e}_a &= |\bar{\mathbf{m}}_a - \bar{\mathbf{m}}| \\ &= \left| \frac{\mathbf{r}_3^T \bar{\mathbf{X}} + t_z}{t_z} \bar{\mathbf{m}} - \bar{\mathbf{m}} \right| = \det\left(\frac{\mathbf{r}_3^T \bar{\mathbf{X}}}{t_z} \right) |\bar{\mathbf{m}}| \\ &= \det\left(\frac{-(\sin\beta)x + (\cos\beta \sin\alpha)y + (\cos\beta\cos\alpha)z}{t_z} \right) |\bar{\mathbf{m}}| \end{aligned} \tag{2.34}$$

Based on the above equations, it is rational to state the following results for different projection models.

1. The axial rotation angle γ around Z-axis has no influence on the images of $\bar{\mathbf{m}}$, $\bar{\mathbf{m}}_q$ and $\bar{\mathbf{m}}_a$.
2. When the distance of a camera to an object is much larger than the object depth, both $\bar{\mathbf{m}}_q$ and $\bar{\mathbf{m}}_a$ are close to $\bar{\mathbf{m}}$.
3. When the camera system is aligned with the world system, i.e., $\alpha = \beta = 0$, we have $\mathbf{r}_{3q}^T = \mathbf{r}_3^T = [0, 0, 1]$ and $\mathbf{e}_q = 0$. Thus $\bar{\mathbf{m}}_q = \bar{\mathbf{m}}$, and the quasi-perspective assumption is equivalent to perspective projection.
4. When the rotation angles α and β are small, we have $\mathbf{e}_q < \mathbf{e}_a$, i.e., the quasi-perspective assumption is more accurate than affine assumption.
5. When the space point lies on the plane through the world origin and perpendicular to the principal axis, i.e., the direction of $\mathbf{r}_3^T$, we have $\alpha = \beta = 0$ and $z = 0$. It is easy to verify that $\bar{\mathbf{m}} = \bar{\mathbf{m}}_q = \bar{\mathbf{m}}_a$.

2

2.4 Experimental Evaluations

During simulation, we randomly generated 200 points within a cube of $20 \times 20 \times 20$ in space as shown in Fig. 2.3(a), only the first 50 points are displayed for simplicity. The depth variation in Z-direction of the space points is shown in Fig. 2.3(b). We simulated 10 images from these points by perspective projection. The image size is set at 800×800. The camera parameters are set as follows: focal lengths are set randomly between 900 and 1100, the principal point is set at the image center, and the skew is zero. The rotation angles are set randomly between $\pm 5°$. The X and Y positions of the cameras are set randomly between ± 15, while the Z position is spaced evenly from 200 to 220. The true projective depths λ_{ij} associated with these points across 10 different views are shown in Fig. 2.3(c), where the values are given after normalization so that they have unit mean.

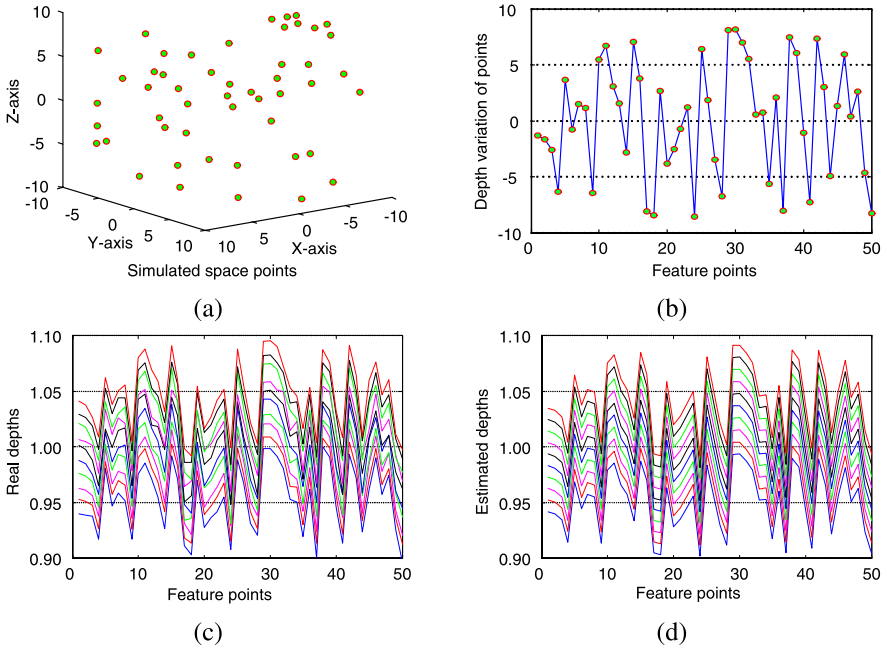

Fig. 2.3 Evaluation on projective depth approximation of the first 50 points. (**a**) Coordinates and distribution of the synthetic space points; (**b**) The depth variation of the space points; (**c**) The real projective depths of the imaged points after normalization; (**d**) The approximated projective depths under quasi-perspective assumption

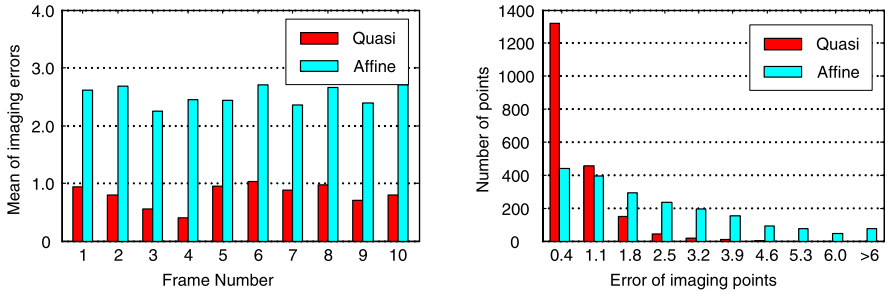

Fig. 2.4 Evaluation of the imaging errors by different camera models. (**a**) The mean error of the generated images by the quasi-perspective and affine projection models; (**b**) The histogram distribution of the errors by different projection models

2.4.1 Imaging Errors

Using the simulated data, we estimate λ_{1j} and μ_i from (2.15) and (2.16), and construct the estimated projective depths from $\hat{\lambda}_{ij} = \frac{\lambda_{1j}}{\mu_i}$. The normalized result is shown in Fig. 2.3(d). We can see from experiment that the recovered projective depths are very close to the ground truths. They are generally in proportion to the depth variation of space points in the Z-direction. If we adopt affine camera model, it is equivalent to setting all projective depths to $\lambda_{ij} = 1$. The error is obviously much bigger than that of the quasi-perspective assumption.

According to projection equations (2.30) to (2.34), different images will be obtained if we adopt different camera models. We generated three sets of images using the simulated space points via general perspective projection model, affine camera model, and quasi-perspective projection model. We compared the errors of quasi-perspective projection model (2.33) and affine assumption (2.34). The mean error of different models in each frame is shown in Fig. 2.4(a), the histogram distribution of the errors for all 200 points across 10 frames is shown in Fig. 2.4(b). Results indicate that the error of quasi-perspective assumption is much smaller than that under affine assumption.

2.4.2 Influence of Imaging Conditions

The proposed quasi-perspective model is based on the assumption of small camera movement. We investigated the influence of different imaging conditions to the model. Initially, we fix the camera position as given in the first test and vary the amplitude of rotation angles from $\pm 5°$ to $\pm 50°$ in steps of $5°$. At each step, we check the relative error of the recovered

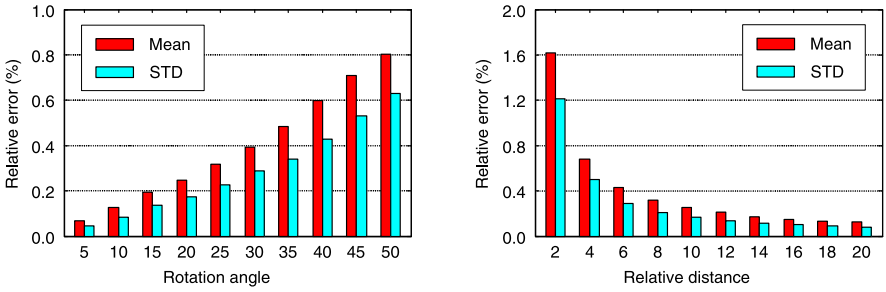

Fig. 2.5 Evaluation on quasi-perspective projection under different imaging conditions. (**a**) The relative error of the estimated depths under different rotation angles; (**b**) The relative error with respect to different relative distances

projective depths, which is defined as

$$e_{ij} = \frac{|\lambda_{ij} - \hat{\lambda}_{ij}|}{\lambda_{ij}} \times 100\,(\%) \tag{2.35}$$

where $\hat{\lambda}_{ij}$ is the estimated projective depth. We carried out 100 independent tests at each step so as to obtain a statistically meaningful result. The mean and standard deviation of e_{ij} are shown in Fig. 2.5(a).

Then, we fix the rotation angles at $\pm 5^\circ$ and vary the relative distance of a camera to an object (i.e. the ratio between the distance of a camera to an object center and the object depth) from 2 to 20 in steps of 2. The mean and standard deviation of e_{ij} at each step for 100 tests are shown in Fig. 2.5(b). Result shows that the quasi-perspective projection is a good approximation ($e_{ij} < 0.5\%$) when the rotation angles are less than $\pm 35^\circ$ and the relative distance is larger than 6. Please note that the result is obtained from noise free data.

2.5 Closure Remarks

2.5.1 Conclusion

In this chapter, we proposed a quasi-perspective projection model and analyzed the projection errors of different projection models. The proposed model is a trade-off between affine and perspective projection. It is computationally simple with better accuracy than affine approximation. The proposed model is suitable for structure and motion factorization of a short sequence with small camera motions. It should be noted that the small rotation assumption of the proposed model is not a limiting factor and is usually satisfied in many real world applications. During image acquisition of an object to be reconstructed,

we tend to control the camera movement so as to guarantee large overlapping part, which also facilitates the feature tracking process. Some geometrical properties of the model in one view and two-view [11] will be presented in the next chapter. The application details to structure and motion factorization [12] will be given in Chap. 9.

2.5.2 Review Questions

1. *Affine camera model*. Provide the general form of the projection matrix under affine camera model. Show that the translation term corresponds to the image of world origin, and it can be removed if the image points are registered to the corresponding centroid. Show that an infinite point in space will also be projected to an infinite point in the image. Illustrate the imaging process of orthographic projection and weak-perspective projection.
2. *Quasi-perspective projection*. Elaborate on the two major assumptions of the quasi-perspective projection. Derive the projection matrix under quasi-perspective projection. What is the main difference between affine and quasi-perspective projection? Present the projection error of the two approximations and make a comparison.

References

1. Aloimonos, J.Y.: Perspective approximation. Image Vis. Comput. **8**(3), 177–192 (1990)
2. Faugeras, O.: Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge (1993)
3. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004). ISBN: 0521540518
4. Kanatani, K., Sugaya, Y., Ackermann, H.: Uncalibrated factorization using a variable symmetric affine camera. IEICE Trans. Inf. Syst. **E90-D**(5), 851–858 (2007)
5. Mundy, J.L., Zisserman, A.: Geometric Invariance in Computer Vision. MIT Press, Cambridge (1992)
6. Ohta, Y.I., Maenobu, K., Sakai, T.: Obtaining surface orientation from texels under perspective projection. In: Proc. International Joint Conferences on Artificial Intelligence, pp. 746–751 (1981)
7. Oliensis, J., Hartley, R.: Iterative extensions of the Sturm/Triggs algorithm: Convergence and nonconvergence. IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2217–2233 (2007)
8. Poelman, C., Kanade, T.: A paraperspective factorization method for shape and motion recovery. IEEE Trans. Pattern Anal. Mach. Intell. **19**(3), 206–218 (1997)
9. Shapiro, L.S., Zisserman, A., Brady, M.: 3D motion recovery via affine epipolar geometry. Int. J. Comput. Vis. **16**(2), 147–182 (1995)
10. Wang, G., Wu, J.: Quasi-perspective projection with applications to 3D factorization from uncalibrated image sequences. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
11. Wang, G., Wu, J.: The quasi-perspective model: Geometric properties and 3D reconstruction. Pattern Recogn. **43**(5), 1932–1942 (2010)
12. Wang, G., Wu, J.: Quasi-perspective projection model: Theory and application to structure and motion factorization from uncalibrated image sequences. Int. J. Comput. Vis. **87**(3), 213–234 (2010)