Chapter 5

A Fixed Perspective-Projection Camera Model

To build an adjustable camera model for an imaging property we are interested in requires a fixed model of that property. In this chapter and the next we describe the approach we use to model 3D to 2D perspective-projection. In the following sections we explain the formulation for the fixed model along with the algorithms and techniques for calibrating it.

5.1 Perspective-projection camera models

Perspective-projection camera models map the coordinates of points in 3D object space into 2D image coordinates. Alternately, an inverted perspective-projection model can be used to determine the corresponding line-of-sight through the 3D object space based on a 2D image coordinate. Perspective-projection models are used in viewpoint planning and for measurement applications, such as stereo.

Many different models have been used for perspective projection [2], [12], [58], [22], [20], [14], [5]. We use the model described by Tsai [53].

Tsai's camera model consists of 11 parameters: six extrinsic, "exterior-orientation" parameters $(R_x, R_y, R_z, T_x, T_y, T_z)$ that describe the position and orientation of the camera's coordinate frame with respect to the world-coordinate frame, and five intrinsic, "interior-orientation" parameters $(f, C_x, C_y, s_x, \kappa_1)$ that describe the camera's image-formation process. All 11 camera parameters are constants estimated from calibration data taken from a single camera view (i.e. the exterior and interior orientation of the camera is fixed). Whenever the camera is moved in the world-coordinate system its exterior orientation must be recomputed while its interior orientation remains unchanged.

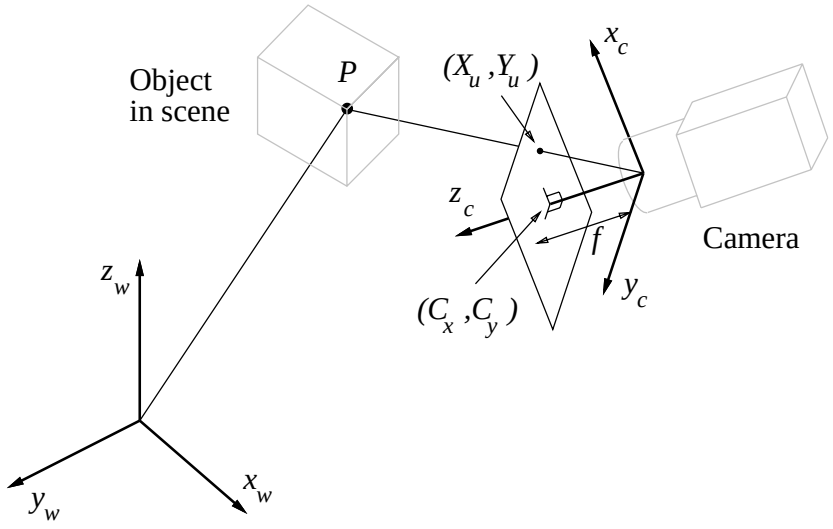

Figure 5.1: Fixed perspective-projection camera model geometry

5.2 Formulation of Tsai's model

In Tsai's model, illustrated in Fig. 5.1, the origin of the camera-centered coordinate system (x_c, y_c, z_c) coincides with the front nodal point of the camera, the z_c axis coincides with the camera's optical axis. The image plane is assumed to be parallel to the (x_c, y_c) plane and at a distance f from the origin, where f is the effective focal length of the pinhole camera.

The relationship between the position of a point P within the world coordinates (x_w, y_w, z_w) and the point's image in the camera's frame buffer (X_f, Y_f) is defined by a sequence of coordinate transformations. The first transformation is a rigid body rotation and translation from the world-coordinate system (x_w, y_w, z_w) to the camera-centered coordinate system (x_c, y_c, z_c). This is described by

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{5.1}$$

where

$$R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \tag{5.2}$$

is the 3×3 rotation matrix describing the orientation of the camera in the world-coordinate system. R can also be expressed as

$$R = \mathrm{Rot}(R_x)\mathrm{Rot}(R_y)\mathrm{Rot}(R_z) \tag{5.3}$$

the product of three rotations around the x, y, and z axes of the world-coordinate system.

The second transformation is a perspective projection (using an ideal pinhole-camera model) of the point in the camera coordinates to the position of its image in undistorted sensor-plane

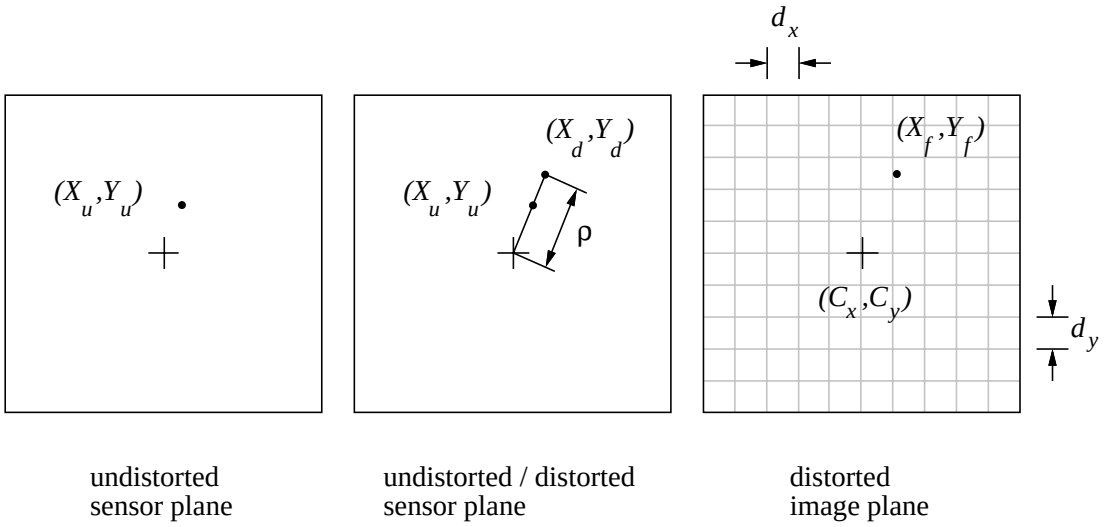

Figure 5.2: Transformation from undistorted sensor to distorted frame coordinates

coordinates, (X_u, Y_u). This transformation is described by

$$X_u = f\frac{x_c}{z_c} \tag{5.4}$$

and

$$Y_u = f\frac{y_c}{z_c} \tag{5.5}$$

where f is the effective focal length of the pinhole camera.

The third transformation (illustrated in Fig. 5.2) is from the undistorted (ideal) position of the point's image in the sensor plane to the true position of the point's image, (X_d, Y_d), which results from geometric lens distortion. This is described by

$$X_u = X_d(1 + \kappa_1\rho^2), \tag{5.6}$$

$$Y_u = Y_d(1 + \kappa_1\rho^2) \tag{5.7}$$

and

$$\rho = \sqrt{X_d^2 + Y_d^2} \tag{5.8}$$

where κ_1 is the coefficient of radial lens distortion. While a more complex model describing both radial and tangential geometric lens distortion could have been used, the accuracy provided by this model is sufficient to demonstrate the development of the adjustable camera model.

The final transformation is between the true position of the point's image on the sensor plane and its coordinates in the camera's frame buffer, (X_f, Y_f). This is described by

$$X_f = d_x^{-1}X_d s_x + C_x \tag{5.9}$$

and

$$Y_f = d_y^{-1}Y_d + C_y \tag{5.10}$$

where C_x and C_y are the coordinates (in pixels) of the intersection of the z_c axis and the camera's sensor plane; d_x and d_y are the effective center-to-center distances between the camera's sensor elements in the x_c and y_c directions; and s_x is a scaling factor to compensate for any uncertainty in the ratio between the number of sensor elements on the CCD and the number of pixels in the camera's frame buffer in the x direction.

5.3 Performance metrics

One of the first questions we have about any camera model is how accurately it captures the imaging behavior. This information is necessary both for measuring progress during model calibration and estimating the performance or accuracy of any application the model is used in.

Given the measured coordinates of a point in the object space (x_w, y_w, z_w) and the measured position of the point's image in the frame buffer (X_f, Y_f) we can define an error metric for the model anywhere along the model's chain of coordinate transformations. One obvious error metric is the difference between the position of a point's image we measure and the position the camera model predicts. If we use the difference in positions following the last coordinate transformation (i.e. after the lens distortion effects have been added to the point's projection through the camera model) we can define the distorted image plane error (DIPE) as

$$\text{DIPE} = \sqrt{(X_f - X'_f)^2 + (Y_f - Y'_f)^2}$$

where (X_f, Y_f) is the measured position of the point's image and (X'_f, Y'_f) is the position of the point's 3D coordinates (x_w, y_w, z_w) projected through the camera model.

In many applications it is desirable to operate in a virtual, undistorted image plane in the camera. In fact, Tsai's fixed camera model is designed to allow converting directly from distorted sensor coordinates (X_d, Y_d) into undistorted sensor coordinates (X_u, Y_u), while going in the opposite direction requires significantly more computation. We define the undistorted image plane error (UIPE) as

$$\text{UIPE} = \sqrt{(\Delta X_{f_u})^2 + (\Delta Y_{f_u})^2} \tag{5.11}$$

where

$$\begin{aligned} \Delta X_{f_u} &= d_x^{-1}(X_{u_2} - X_{u_1})s_x, \\ \Delta Y_{f_u} &= d_y^{-1}(Y_{u_2} - Y_{u_1}). \end{aligned}$$

(X_{u_2}, Y_{u_2}) are calculated from the measured position of the point's image (X_f, Y_f) using equations (5.6), (5.7), and (5.8), while (X_{u_1}, Y_{u_1}) are calculated from the 3D coordinates of the point (x_w, y_w, z_w) using (5.1), (5.4), and (5.5). The algorithms that we use to calibrate the camera model (and that we will describe later on) minimize the sum-of-squared error in the undistorted image plane for the calibration data.

In inverse perspective-projection problems it is often helpful to know what level of accuracy the camera model has in the object space. By projecting an image point (X_f, Y_f) back through the camera model we can calculate the closest distance of approach between the image point's line-of-sight and the point in 3D object space (x_w, y_w, z_w) that was supposed to have cast the image. This object-space error (OSE) can be calculated as

$$\text{OSE} = \sqrt{(x_c - X_u t)^2 + (y_c - Y_u t)^2 + (z_c - f t)^2}$$

where

$$t = \frac{x_c X_u + y_c Y_u + z_c f}{X_u^2 + Y_u^2 + f^2},$$

f is the camera's effective focal length, (x_c, y_c, z_c) are the 3D coordinates (x_w, y_w, z_w) rotated and translated into the camera's coordinate frame, and (X_u, Y_u) are (X_f, Y_f) transformed into undistorted sensor coordinates.

The above three error metrics are measurements of how well the camera model captures the perspective projection imaging property of the camera system. Another possible error metric is how well the camera model performs in a particular type of application. Tsai [53] gives a theoretical upper bound for a 3D measurement error in the case where two calibrated cameras have been used in a stereo pair. Das [18] describes a set of equations that can be used to convert uncertainty in the image plane (UIPE) into uncertainty in the range error for a general stereo configuration.

5.4 Calibration data

Our estimation of the unknown parameters in the fixed camera model is based on calibration data consisting of 3D object space coordinates and corresponding 2D image coordinates. For the experiments described in this thesis we used a planar calibration target mounted on a translation stage (see Fig. 5.3). The normal of the calibration target is exactly parallel to the stage's direction of travel. The calibration target itself contains 1/8-inch-diameter, black reference points precisely spaced out on a regular, 1-inch grid.

For any set of images of the calibration target the relative 3D coordinates (x_w, y_w, z_w) of the reference points are known from their grid position in the target plane and from the position of the target plane along the translation stage. The (X_f, Y_f) positions of the dots in the image plane are measured to sub-pixel, accuracy using the procedure described in Appendix C.

The accuracy of the parameter estimation for the fixed camera model depends in part on the distribution of data points across both the 3D object space and the 2D image plane. To be able to accurately estimate the f and T_z parameters the calibration data must have some variation in depth in the camera coordinate frame's z axis. As a general rule the calibration data should cover a range from the closest to the farthest extent of the volume that the camera model is to be used for. Also, to accurately measure the radial lens-distortion coefficient (κ_1) and the image center (C_x, C_y) the data's radial positions should vary across the camera's field of view.

5.5 Model calibration

In the calibration of our fixed camera models we assume that the six exterior orientation parameters $(R_x, R_y, R_z, T_x, T_y, T_z)$ and five interior orientation parameters $(f, \kappa_1, C_x, C_y, s_x)$ are all unknown and must estimated from the calibration data. We calibrate the fixed camera model in two steps. First we use a quick algorithm to obtain approximate estimates for nine the camera's 11 parameters and then we use iterative, non-linear optimization to refine all 11 parameters.

5.5.1 Initial parameter estimation

For our first calibration step we use Tsai's [53] non-coplanar calibration algorithm to determine values for the T_x, T_y, T_z, f, κ_1, and s_x parameters and for the 3×3 rotation matrix R. Tsai's algorithm begins by setting the radial lens distortion coefficient (κ_1) to zero and then calculating R, T_x, T_y, and s_x directly from the calibration data. Next, with κ_1 still zero an initial estimate is calculated for f and T_z. Finally, f, T_z, and κ_1 are refined using an iterative, non-linear optimization.

In Tsai's published algorithm the non-linear optimization of the camera model minimizes the sum of the error

$$\text{error} = (Y_u - Y'_u)^2$$

where Y_u is the y component of the measured position of the point's image (X_f, Y_f) transformed into sensor coordinates and Y'_u is the y component of point's world coordinates projected through the camera model into the sensor plane. Presumably this was done for speed and/or to reduce the influence of timing jitter noise that can be present in the x components of images taken using video cameras[4]. In any event this approach makes almost no use of the calibration information in the x direction. In our camera systems timing jitter is either not present or can be averaged out of the image. Thus for our implementation of Tsai's algorithm, and for the full non-linear optimization approach described next, we make full use of the calibration data and minimize the sum-of-squared error in both the x and y directions, i.e.

$$\text{error} = (X_u - X'_u)^2 + (Y_u - Y'_u)^2 \tag{5.12}$$

In Tsai's algorithm C_x and C_y are assumed to be known a priori. Three methods for measuring C_x and C_y are are outlined by Lenz and Tsai in [30]. As we have demonstrated in Chapter 3, the validity of these types of measurement is questionable. However, since the output of Tsai's algorithm is only serving as an initial estimate for the second step of our camera calibration a rough estimate for C_x and C_y is sufficient for our purposes. We use the center of an autocollimated laser (described in Section 3.3.2) to give us a rough estimate.

For closed circuit TV (CCTV) cameras the one-to-one correspondence between the CCD's rows and the camera's frame buffer rows allows the value of d_y to be obtained directly from the CCD's specification sheets. The value for d_x depends on the relative frequencies of the clock used to shift data from the CCD onto the video signal and the clock used to sample the

video signal at the frame buffer. Several methods exist for accurately measuring d_x [30][3]. In Tsai's camera model the s_x parameter is used to compensate for any error in d_x. Since we include s_x in the camera model's non-linear optimization only a rough estimate of d_x is required.

For digital output cameras, such as the Photometrics camera that we use in some of the following experiments, we can obtain exact values for d_x and d_y directly from the specifications for the camera's CCD.

5.5.2 Full non-linear optimization

In the second step of calibrating the fixed camera model we take the output of Tsai's non-coplanar algorithm and use iterative non-linear optimization to refine the values of all 11 variable model parameters $(R_x, R_y, R_z, T_x, T_y, T_z, f, \kappa_1, C_x, C_y, s_x)$. To perform the non-linear optimization we use the IMSL routine DUNLSF which employs a modified Levenberg-Marquardt algorithm and a finite-difference Jacobian[1] to minimize the error function (5.12) for all points in the calibration data.

Since we cannot easily enforce the orthonormality constraint on the rotation matrix R in (5.1) during the non-linear optimization we use (5.3) and optimize R_x, R_y, and R_z instead. The rotation angles R_x, R_y, and R_z can be determined from R using the equations

$$R_z = \arctan\left(\frac{r_4}{r_1}\right) \tag{5.13}$$

$$R_y = \arctan\left(\frac{r_7}{r_1\cos(R_z) + r_4\sin(R_z)}\right) \tag{5.14}$$

$$R_x = \arctan\left(\frac{r_3\sin(R_z) - r_6\cos(R_z)}{r_5\cos(R_z) - r_2\sin(R_z)}\right) \tag{5.15}$$

While the R provided by Tsai's algorithm is not guaranteed to be orthonormal and thus the calculated rotation angles are not necessarily that good, their accuracy is sufficient for them to serve a starting point for the non-linear optimization.

5.5.3 Public archive for code

An implementation of the above algorithms, written in the C programming language, can be found in the Vision List Archive SHAREWARE subdirectory available via anonymous ftp from FTP.TELEOS.COM. The code requires two non-linear optimization routines that can be found in either of the IMSL or the NAG commercial software packages.

[1]The complexity of the fixed camera model makes the direct calculation of the Jacobian computationally prohibitive.

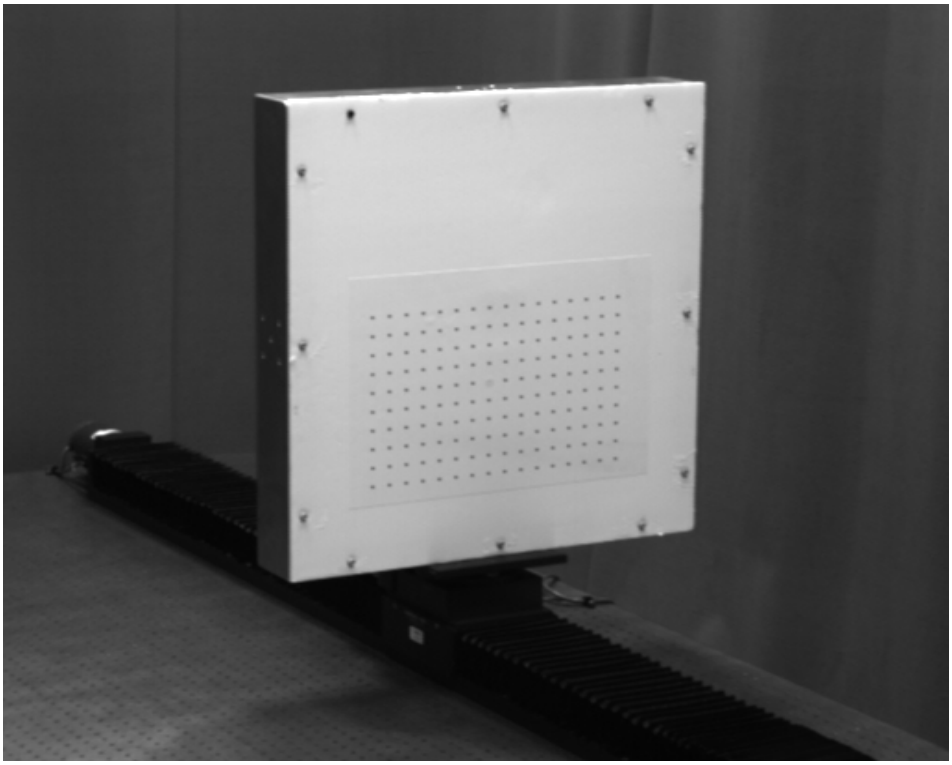

Figure 5.3: Calibration target and translation stage

5.6 Recalibrating exterior orientation

Whenever the camera is moved to a new pose the camera's exterior orientation must be recomputed from calibration data taken at the new position. The camera's interior orientation remains unchanged.

To compute just the exterior orientation we use a modification of Tsai's algorithm. We start by using the previously calibrated intrinsic parameters C_x, C_y, and s_x to transform (X_f, Y_f) from the new calibration data into undistorted sensor coordinates (X_u, Y_u). These coordinates are used in the first stage of Tsai's algorithm to determine R (and subsequently R_x, R_y, and R_z). Given (X_u, Y_u) and R we then estimate T_x, T_y, and T_z using the following approach. Rewriting (5.1), (5.2), (5.4), and (5.5) we obtain two equations,

$$\begin{aligned} X_{u_i} &= f\frac{x_{w_i}r_1 + y_{w_i}r_2 + z_{w_i}r_3 + T_x}{x_{w_i}r_7 + y_{w_i}r_8 + z_{w_i}r_9 + T_z} \\ &= f\frac{x_{k_i} + T_x}{z_{k_i} + T_z} \end{aligned} \tag{5.16}$$

and

$$\begin{aligned} Y_{u_i} &= f\frac{x_{w_i}r_4 + y_{w_i}r_5 + z_{w_i}r_6 + T_y}{x_{w_i}r_7 + y_{w_i}r_8 + z_{w_i}r_9 + T_z} \\ &= f\frac{y_{k_i} + T_y}{z_{k_i} + T_z}, \end{aligned} \tag{5.17}$$

for each point in the new calibration data. Using (5.16) and (5.17) and the new calibration data we then form an over-determined set of linear equations

$$\left[\begin{array}{ccc} f & 0 & -X_{u_i} \\ \vdots & \vdots & \vdots \\ \hline 0 & f & -Y_{u_i} \\ \vdots & \vdots & \vdots \end{array}\right] \left[\begin{array}{c} T_x \\ T_y \\ T_z \end{array}\right] = \left[\begin{array}{c} X_{u_i}z_{k_i} - fx_{k_i} \\ \vdots \\ \hline Y_{u_i}z_{k_i} - fy_{k_i} \\ \vdots \end{array}\right]$$

which can be solved to obtain estimates for T_x, T_y, and T_z.

Finally, R_x, R_y, R_z, T_x, T_y, and T_z are refined using the iterative non-linear optimization routine described in Section 5.5.2.

Parameter	Value	Units
f	60.013	mm
C_x	267.198	pixels
C_y	255.040	pixels
κ_1	-0.000103	$1/\text{mm}^2$
s_x	1.079	
R_x	-0.084	degrees
R_y	0.589	degrees
R_z	0.182	degrees
T_x	-521.238	mm
T_y	-527.935	mm
T_z	1581.238	mm
mean UIPE	0.064	pixels
standard deviation UIPE	0.033	pixels
maximum UIPE	0.182	pixels
mean OSE	0.042	mm
standard deviation OSE	0.024	mm
maximum OSE	0.135	mm

Table 5.1: Example of a calibrated fixed camera model

5.7 Calibration example

To demonstrate the calibration of a fixed camera model we calibrated the Cosmicar/Panasonic camera system for the lens setting ($m_f = 2000$, $m_z = 1000$, $m_a = 1500$). The calibration data for the model came from two images of the calibration target taken with sensor-to-target ranges of 1.5 m (Fig. 5.4) and 2.5 m (Fig. 5.5). The absolute position of the origin for the world-coordinate system was arbitrarily assigned to be in the target plane at 1.5 m range, approximately 520 mm up and 520 mm to the left of the center of the camera's field of view. The two images provided 186 data points.

Table 5.1 shows the calibrated fixed camera model after the final non-linear optimization step. The small values for the mean UIPE and maximum UIPE indicate that the calibrated camera model does a good job of capturing the lens's 3D to 2D imaging behavior.

Figures 5.6 and 5.7 show plots of the actual error between the measured positions of the reference points from the calibration data and the positions predicted by the calibrated model. In the plots the measured positions are marked with small square boxes. The vectors from each box point to the position of the point's image predicted by the camera model. In both figures the error vectors are magnified 100×. The non-random distribution of the direction vectors indicates limitations in the dimensional accuracy of the target and translation stage used to obtain the calibration data.

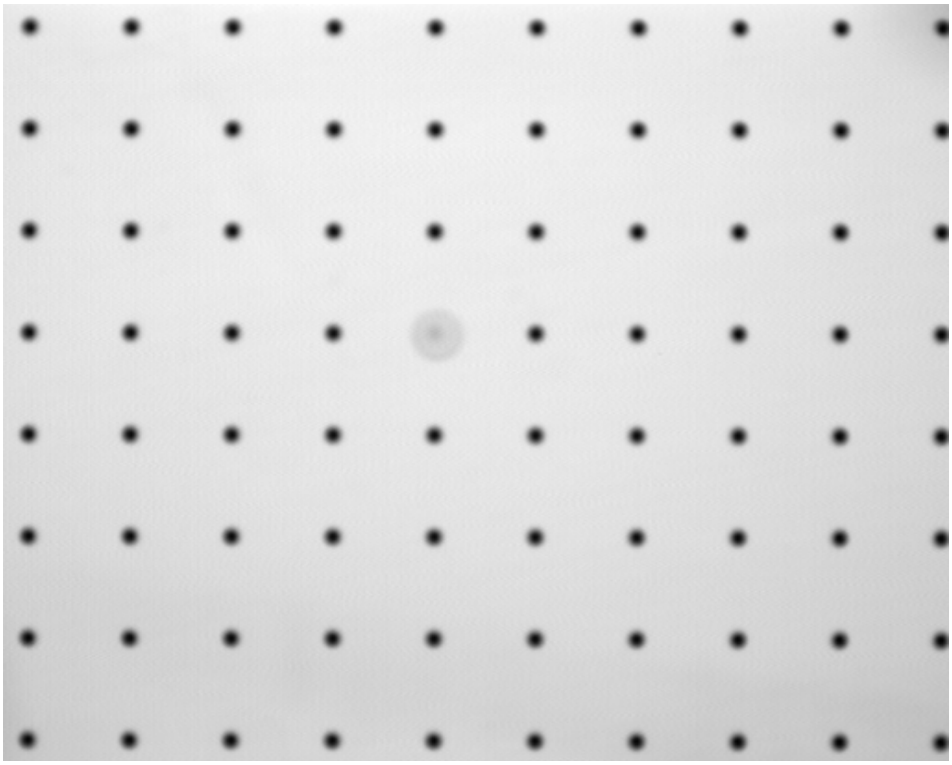

Figure 5.4: Calibration target at 1.5 m

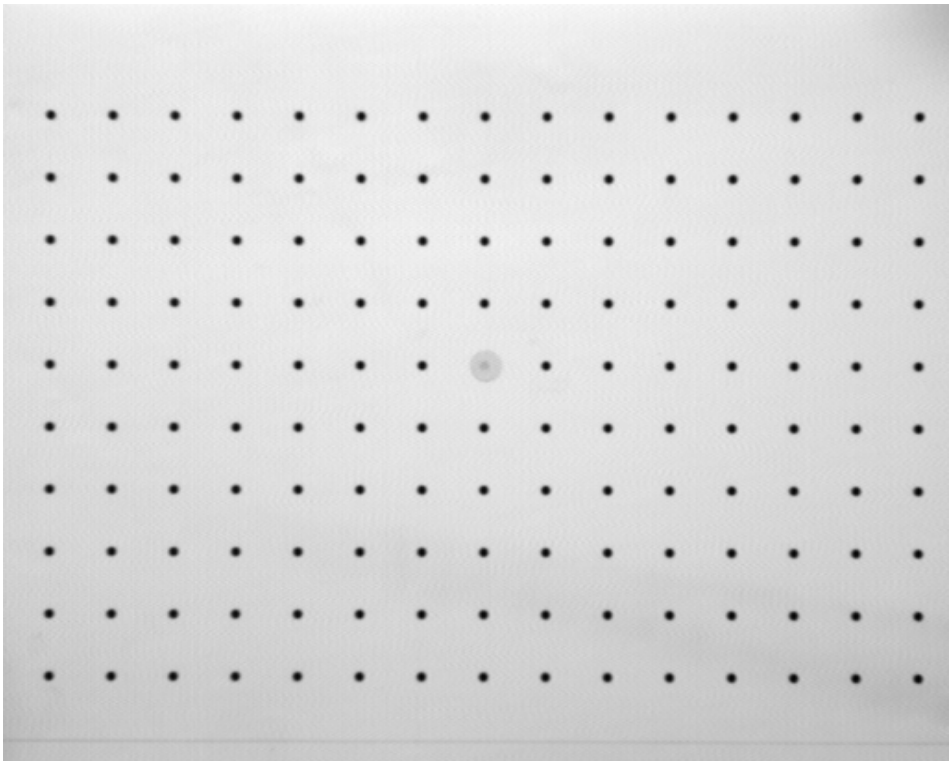

Figure 5.5: Calibration target at 2.5 m

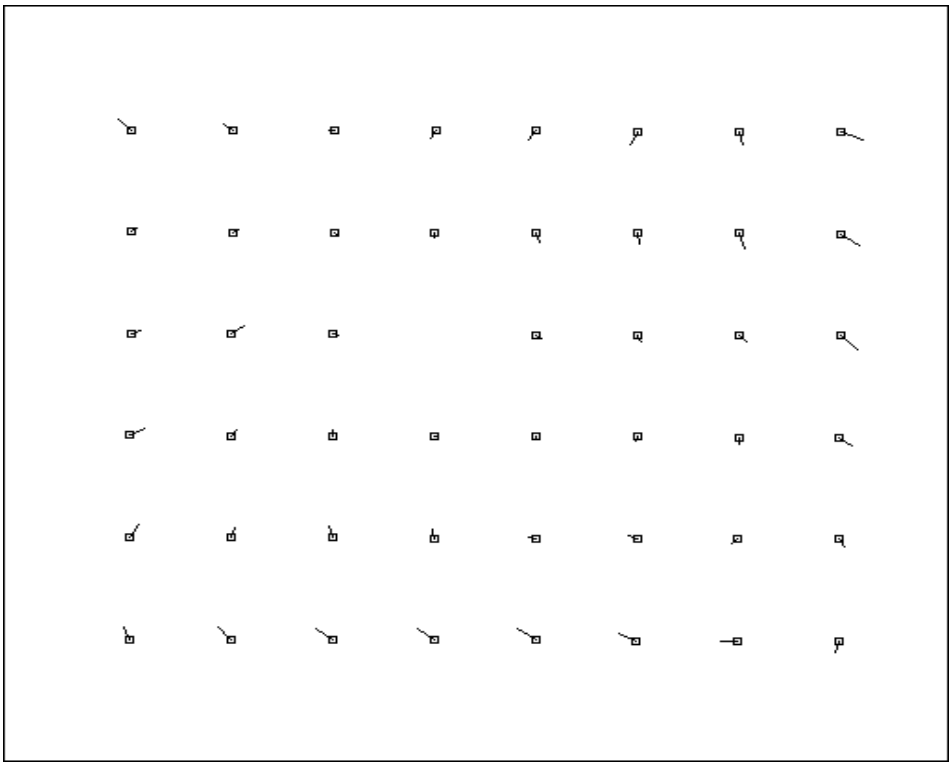

Figure 5.6: Residual error for calibration target at 1.5 m (magnified 100×)

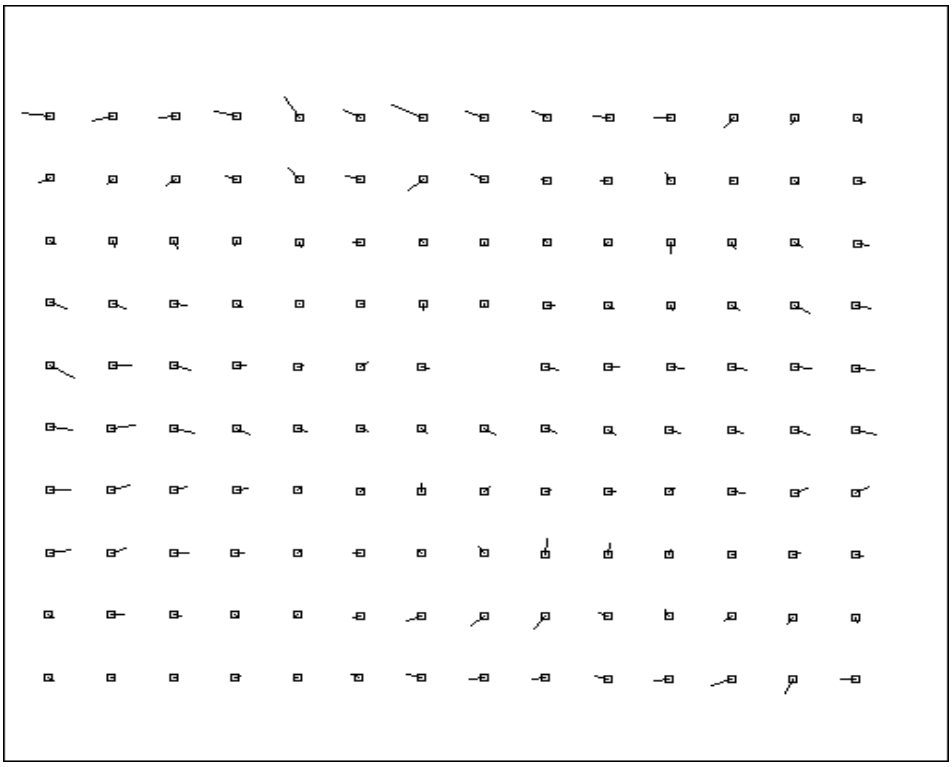

Figure 5.7: Residual error for calibration target at 2.5 m (magnified 100×)

5.8 Exterior orientation recalibration example

To test our ability to reacquire the camera's extrinsic parameter models after the camera has been moved we took one calibration dataset, changed the camera's exterior orientation (pose), and then took a second dataset. Using the first dataset we obtained a fully calibrated fixed camera model. The intrinsic parameters from this model are then used in a partial calibration with the second dataset to estimate the camera's new extrinsic parameters. The second set of data was then used to obtain a fully calibrated fixed camera model. The test is illustrated graphically in Fig. 5.10.

For the first dataset we actually used the data from Section 5.7. Before taking the second dataset the camera was shifted -100 mm in the camera x_w coordinate, -100 mm in the camera y_w coordinate and -100 mm in the camera z_w coordinate, and then rotated until the center of the field of view was located roughly in the middle of the two target plane positions. The new dataset was obtained from two images of the calibration target taken at ranges of 1.5 m and 2.5 m (Figs. 5.8 and 5.9). The world-coordinate frame for the new dataset was the same as that used for the first dataset. The second dataset of data contains 203 data points.

Column 2 of Table 5.2 shows the results for the full calibration on the data from the first pose. Column 4 shows the results for the full calibration for the second pose. The 20% increase in the mean UIPE for the second set of data is due in part to the lack of flatness of the calibration target. With the first set of data the target is viewed head on and target flatness is less critical. For the second set the viewing angle is more oblique.

Column 3 of Table 5.2 shows the calibration results when the intrinsic parameters from the first set of data are used in a partial calibration to obtain the extrinsic parameters from the second set of data. Comparing the results for the second set of data (columns 3 and 4) we see that the exterior orientation and mean UIPE are virtually the same for the partial and full calibration. This demonstrates that the algorithm in Section 5.6 can accurately reacquire the camera's exterior orientation after the camera has been moved. It also demonstrates that the intrinsic model can be carried between poses.

5.9 Summary

In this chapter we presented the fixed camera model that will be the basis of the adjustable camera models developed in the next chapter. We have discussed the formulation of the fixed model along with the algorithms and techniques required to calibrate the it for a fixed-parameter camera system. We have also presented performance metrics and validation techniques for the fixed camera model.

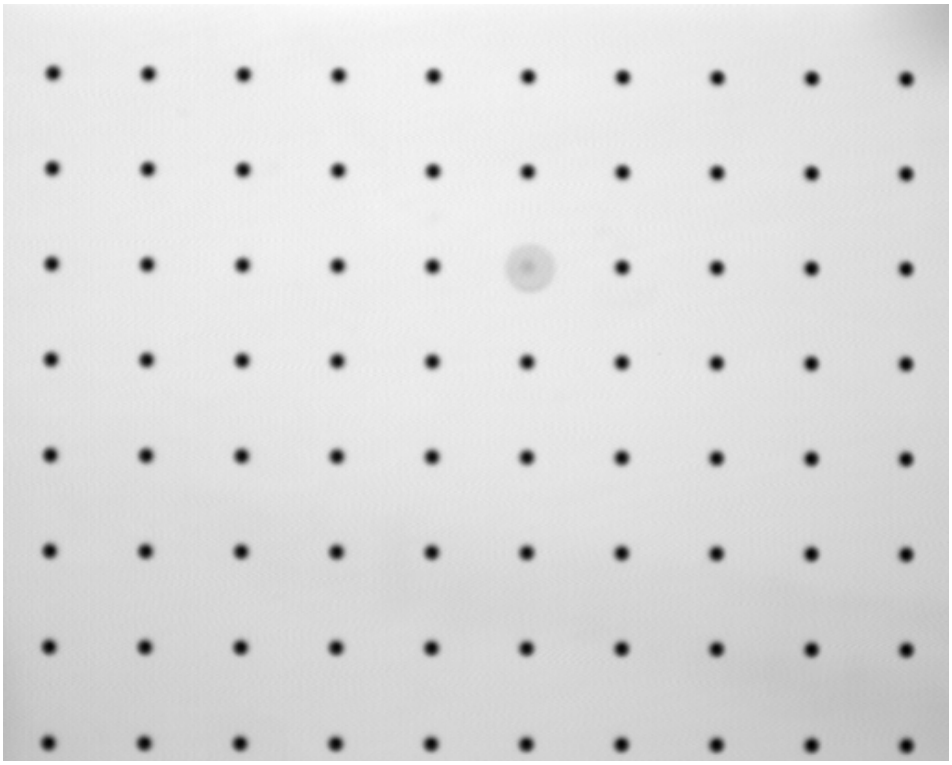

Figure 5.8: Calibration target at 1.5 m with camera shifted

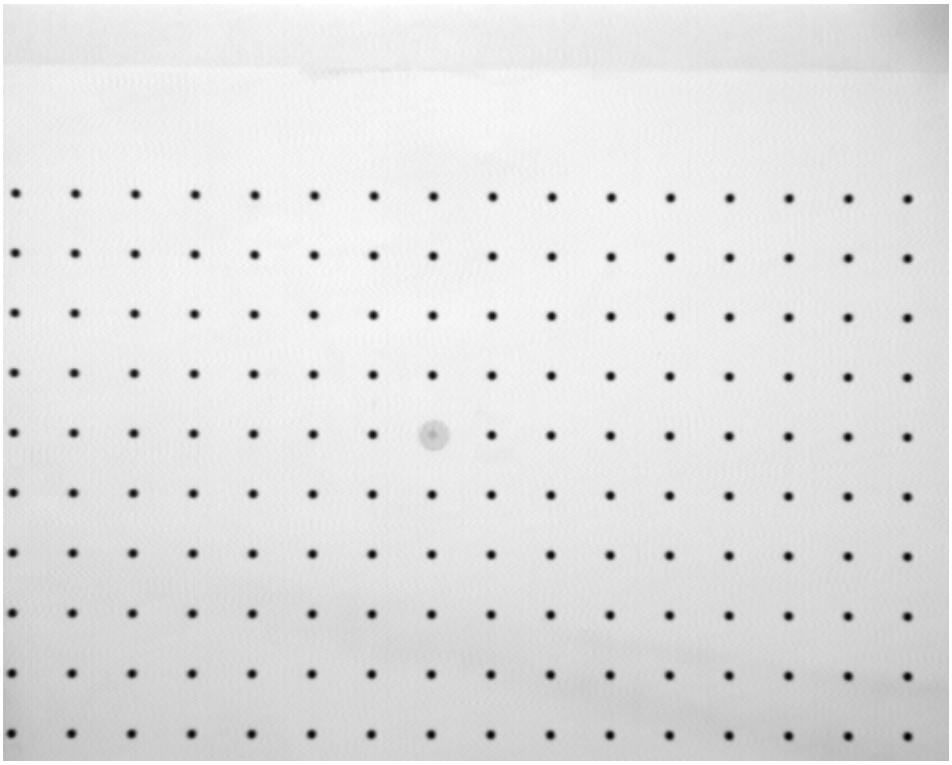

Figure 5.9: Calibration target at 2.5 m with camera shifted

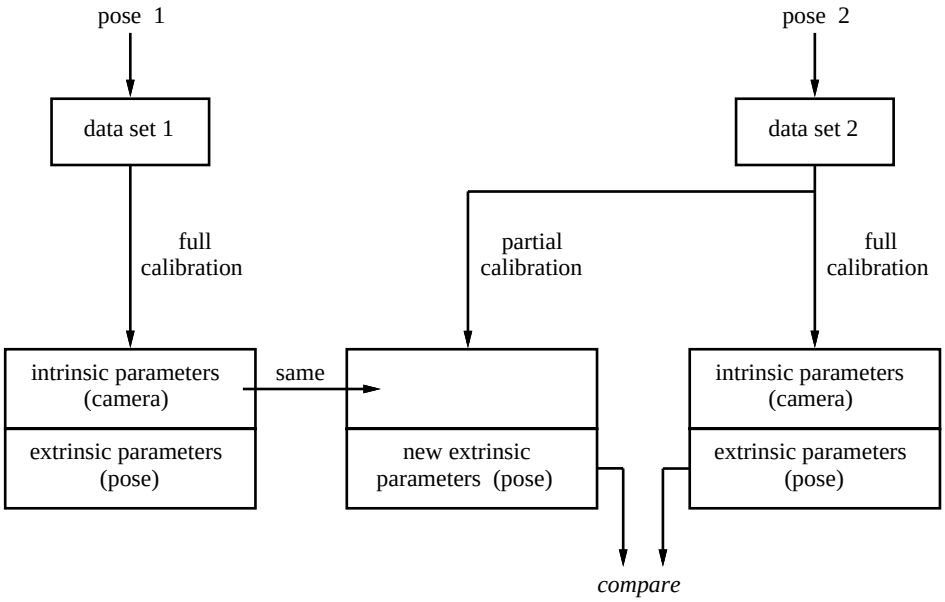

Figure 5.10: Experiment in carrying intrinsic parameters to new pose

Parameter	Pose 1 full calibration	Pose 2 partial calibration	Pose 2 full calibration	Units
f	60.013	60.013	60.058	mm
C_x	267.198	267.198	267.296	pixels
C_y	255.040	255.040	254.626	pixels
κ_1	-0.000103	-0.000103	-0.000096	$1/\text{mm}^2$
s_x	1.079	1.079	1.078	
R_x	-0.084	-2.832	-2.838	degrees
R_y	0.589	-2.042	-2.044	degrees
R_z	0.182	0.303	0.303	degrees
T_x	-521.238	-497.003	-497.045	mm
T_y	-527.935	-547.358	-547.186	mm
T_z	1581.238	1689.919	1690.989	mm
mean UIPE	0.064	0.077	0.076	pixels
standard deviation UIPE	0.033	0.045	0.040	pixels
maximum UIPE	0.182	0.223	0.202	pixels
mean OSE	0.042	0.050	0.049	mm
standard deviation OSE	0.024	0.031	0.029	mm
maximum OSE	0.135	0.153	0.157	mm

Table 5.2: Calibration results from carrying intrinsic parameters to new pose