Model-based Camera Calibration Using Analysis by Synthesis Techniques

Peter Eisert

Computer Vision & Graphics Group
Image Processing Department
Heinrich-Hertz-Institute, Berlin, Germany
Email: eisert@hhi.de

Abstract

A new technique for the determination of extrinsic and intrinsic camera parameters is presented. Instead of searching for a limited number of discrete feature points of the calibration test object, the entire image captured with the camera is exploited to robustly determine the unknown parameters. Shape and texture of the test object are described by a 3-D computer graphics model. With this 3-D representation, synthetic images are rendered and matched with the original frames in an analysis by synthesis loop. Therefore, arbitrary test objects with sophisticated patterns can be used to determine the camera settings. The scheme can easily be extended to deal with multiple frames for a higher intrinsic parameter accuracy.

1 Introduction

Many algorithms in computer vision assume that the intrinsic parameters of the camera are known. Therefore, camera calibration is often performed in an initial step to determine focal length, aspect ratio, and radial lens distortion of the camera. Calibration test objects as, e.g., shown in Fig. 1 are captured with the camera whose parameters are to be determined. Since the geometry of the test object is explicitly known, the 2-D correspondences between the projected 3-D points of the computer model and the 2-D points in the captured camera frame can be exploited to determine the parameters of the camera [1]. Due to the sensitivity of the obtained parameters to noise in the point correspondences [2, 3], a significant amount of effort is spent to determine the position of the 2-D feature points with subpixel accuracy [4, 5].

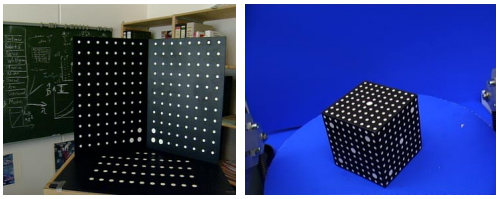

Figure 1: Two different calibration test objects.

In this paper, a model-based calibration approach is presented that does not require the search for discrete feature points. Instead, a 3-D computer graphic model is used to represent shape and texture of the calibration test object. Synthetic views of the body are rendered and the entire image information is exploited to match these views with the camera frame. Hence, the calibration body is not restricted to showing simple discrete features that can be detected easily, but arbitrarily textured objects with a wide range of spatial frequencies can be used. Moreover, no special treatment at the image borders (providing much information about radial distortion) has to be done if the test object is only partly visible. Thus, much more image information compared to the discrete feature point case is exploited leading to highly accurate estimates. The complexity of the approach, however, remains low due to its linearity which also allows the estimation of a common set of intrinsic parameters for a large number of views.

The paper is organized as follows. In Section 2, the camera model used for calibration is described. Then, the algorithm for the model-based estimation of extrinsic camera parameters using an analysis by synthesis loop is presented. This method is extended in Section 4 to jointly estimate intrinsic and extrinsic parameters from single or multiple views. Experimental results on real camera images are finally shown in Section 6 in order to illustrate the accuracy of the presented approach.

2 Camera Model

The camera model describes the image capturing process where a 3-D scene is projected onto a sampled and quantized 2-D image. During that process, several geometric and photometric distortions (for the photometric changes which are not considered in this paper see, e.g., [6]) apply. The projection, signal modifications, and the camera characteristics are modeled and specified by the free intrinsic parameters of the camera model. The real camera is represented in this context as a concatenation of perspective projection, a lens, and a CCD model as shown in Fig. 2.

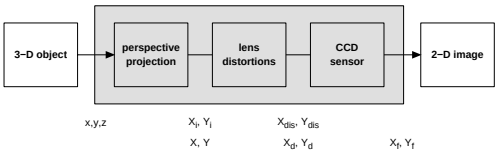

Figure 2: Camera model.

2.1 Perspective Projection

The perspective projection describes the relation between a 3-D object point $\mathbf{x} = [x \quad y \quad z]^T$ in the camera coordinate system and its corresponding 2-D point $\mathbf{X}_i = [X_i \quad Y_i]^T$ in the image plane as shown in Fig. 3. If the origin of the camera coordinate system is located at the focal point this projection is given by

$$\begin{aligned} X_i &= -f \cdot \frac{x}{z} \\ Y_i &= -f \cdot \frac{y}{z}, \end{aligned} \tag{1}$$

with f being the focal length.

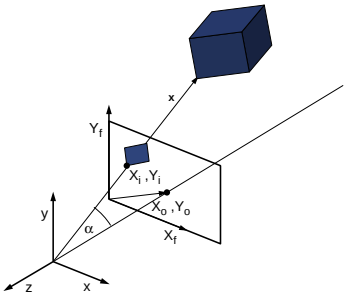

Figure 3: Perspective projection and coordinate systems.

2.2 Lens Distortion

The perspective projection describes the ideal pin-hole camera. In real cameras, the system of lenses introduces geometric and photometric distortions. The exact modeling of these effects is quite complex [7] and less suitable for the incorporation into computer vision schemes. Therefore, a simpler model is used here that considers only dominant effects. The dominant geometrical distortion is the radial lens distortion [1, 8] that can account for barrel or pincushion distortions. With this model, the relation between the undistorted image point $\mathbf{X}_i$ and the corresponding distorted point $\mathbf{X}_{dist}$ is described by a radial displacement according to [1]

$$\begin{aligned} X_i &= X_{dist}(1 + \sum_{i=1} \kappa_i r^{2i}) \\ Y_i &= Y_{dist}(1 + \sum_{i=1} \kappa_i r^{2i}), \end{aligned} \tag{2}$$

with κ_i being the parameters of the displacement function. In practical realizations, only the first or the first two parameters of this series are utilized. The radius r is calculated in this work with normalized coordinates according to

$$r = \sqrt{\left(\frac{X_{dist}}{f}\right)^2 + \left(\frac{Y_{dist}}{f}\right)^2}. \tag{3}$$

2.3 Sampling and Signal Post-Processing

At the CCD-sensor consisting of $N_x \times N_y$ light sensitive sensor elements, the projected and distorted image is spatially quantized and the samples are stored in the frame buffer. The position of each sensor element is related to the frame buffer coordinates $\mathbf{X}_f = [X_f \quad Y_f]^T$ according to

$$\begin{aligned} X_f &= X_0 + s_x \cdot X_{dist} \\ Y_f &= Y_0 + s_y \cdot Y_{dist}, \end{aligned} \tag{4}$$

with s_x and s_y being horizontal and vertical scaling factors, respectively. X_0 and Y_0 denote the location of the optical axis in frame buffer coordinates.

Similarly to the frame buffer coordinates $\mathbf{X}_f$, $\mathbf{X}_i$ and $\mathbf{X}_{dist}$ can be scaled appropriately to relate them with pixel coordinates of the CCD chip. The corresponding values $\mathbf{X} = [X \quad Y]^T$ and $\mathbf{X}_d = [X_d \quad Y_d]^T$ are computed according to

$$\begin{aligned} X &= s_x \cdot X_i \\ Y &= s_y \cdot Y_i \end{aligned} \tag{5}$$

and

$$\begin{aligned} X_d &= s_x \cdot X_{dist} \\ Y_d &= s_y \cdot Y_{dist}. \end{aligned} \tag{6}$$

2.4 Combined Camera Model

Perspective projection, CCD chip deviation, and scaling in order to consider the pixel geometry are combined, leading to the following simple camera model

$$\begin{aligned} X &= -f_x \cdot \frac{x}{z} \\ Y &= -f_y \cdot \frac{y}{z}, \end{aligned} \tag{7}$$

with f_x and f_y being the focal length scaled by s_x and s_y, respectively. The internal parameters used here for describing the camera geometry are f_x, f_y, and κ_i. From these values, the aspect ratio a and height angle φ_{height} can be computed according to

$$\begin{aligned} a &= \frac{N_x}{N_y} \cdot \frac{f_y}{f_x} \\ \varphi_{height} &= 2 \arctan(\frac{2 f_y}{N_y}). \end{aligned} \tag{8}$$

Geometric lens distortions are not included in the model defined by (7). They are taken into account by compensating the distortions in the original frames and then applying the algorithms to the corrected frames. The compensation for radial lens distortion as described in Section 2.2 can also be performed using discrete pixel coordinates

$$\begin{aligned} X &= X_d(1+\sum_{i=1}\kappa_i r^{2i}) \\ Y &= Y_d(1+\sum_{i=1}\kappa_i r^{2i}), \end{aligned} \tag{9}$$

with

$$r = \sqrt{\left(\frac{X_d}{f_x}\right)^2 + \left(\frac{Y_d}{f_y}\right)^2}. \tag{10}$$

The use of normalized coordinates for the computation of r simplifies the estimation of κ in a hierarchical scheme since the value becomes independent of the image resolution.

3 Estimation of Extrinsic Parameters

In a first step, the intrinsic camera parameters as, e.g., focal length and aspect ratio are assumed to be known and only the extrinsic camera parameters are estimated from the image I that is captured for calibration purposes. These 6 parameters (3 for rotation and 3 for translation) specify the position of the test object relative to the camera origin. Given the 3-D object points $\mathbf{x}_0$ in a local object coordinate system representing the shape of the calibration body, the corresponding points in the camera coordinate system $\mathbf{x}$ can be calculated according to

$$\mathbf{x} = \mathbf{R}\cdot\mathbf{x}_0 + \mathbf{t}, \tag{11}$$

with $\mathbf{R}$ being a rotation matrix and $\mathbf{t}$ a translation vector representing the position of the test object relative to the camera origin. Both $\mathbf{R}$ and $\mathbf{t}$ are not known and have to determined from the captured image I. If estimates $\hat{\mathbf{R}}$ of $\mathbf{R}$ and $\hat{\mathbf{t}}$ of $\mathbf{t}$ are already available the estimated object point locations are given by

$$\hat{\mathbf{x}} = \hat{\mathbf{R}}\cdot\mathbf{x}_0 + \hat{\mathbf{t}}. \tag{12}$$

Since all these values are known, an estimate $\hat{I}$ of the camera image I can be created by rendering the 3-D model of the test object at the estimated position $\hat{\mathbf{x}}$. The correct but unknown values $\mathbf{x}$ are related with the known estimates by

$$\mathbf{x} = \Delta\mathbf{R}(\hat{\mathbf{x}} - \mathbf{x}_c) + \mathbf{x}_c + \Delta\mathbf{t}, \tag{13}$$

with $\mathbf{x}_c = [x_c \;\; y_c \;\; z_c]^T = \hat{\mathbf{t}}$ being the center of the object, $\Delta\mathbf{R} = \mathbf{R}\cdot\hat{\mathbf{R}}^T$ the unknown correction for the rotation matrix, and $\Delta\mathbf{t} = \mathbf{t} - \hat{\mathbf{t}}$ the correction for the translation vector, respectively. For the determination of the extrinsic camera parameters, the six degrees of freedom of $\Delta\mathbf{R}$ and $\Delta\mathbf{t}$ have to be estimated.

Since the relative rotation described by $\Delta\mathbf{R}$ can be expected to be very small, a linearized version of the rotation matrix is utilized to obtain a model that is linear in the unknown extrinsic parameters

$$\Delta\mathbf{R} \approx \begin{bmatrix} 1 & -\Delta R_z & \Delta R_y \\ \Delta R_z & 1 & -\Delta R_x \\ -\Delta R_y & \Delta R_x & 1 \end{bmatrix}. \tag{14}$$

The three parameters ΔR_x, ΔR_y, and ΔR_z of the matrix $\Delta\mathbf{R}$ are the three Euler angles of the rotation matrix.

For the camera calibration, the real 3-D coordinates $\mathbf{x}$ are not known but their projection $\mathbf{X} = [X \;\; Y]^T$ into the 2-D image I. Therefore, we have to relate these values by inserting (13) into the camera model (7) leading to

$$\begin{aligned} X &= -f_x\frac{\hat{x} + (\hat{z}-z_c)\Delta R_y - (\hat{y}-y_c)\Delta R_z + \Delta t_x}{\hat{z} + (\hat{y}-y_c)\Delta R_x - (\hat{x}-x_c)\Delta R_y + \Delta t_z} \\ Y &= -f_y\frac{\hat{y} - (\hat{z}-z_c)\Delta R_x + (\hat{x}-x_c)\Delta R_z + \Delta t_y}{\hat{z} + (\hat{y}-y_c)\Delta R_x - (\hat{x}-x_c)\Delta R_y + \Delta t_z}. \end{aligned} \tag{15}$$

Substitution of the 3-D coordinates by

$$\hat{x} = -\frac{\hat{z}\hat{X}}{f_x} \quad \hat{y} = -\frac{\hat{z}\hat{Y}}{f_y} \tag{16}$$

and first order Taylor series expansion lead to the desired constraint for the 2-D displacement field with the displacement vectors $\mathbf{d} = [d_x \;\; d_y]^T$ given by

$$\begin{aligned} d_x \quad &= X - \hat{X} \approx \\ &f_x\left[-\Delta R_y(1-\frac{z_c}{\hat{z}}) - \frac{\Delta R_z}{f_y}Y_n - \frac{\Delta t_x}{\hat{z}}\right] + \\ &+\hat{X}\left[\frac{\Delta R_x}{f_y}Y_n - \frac{\Delta R_y}{f_x}X_n - \frac{\Delta t_z}{\hat{z}}\right] \\ d_y \quad &= Y - \hat{Y} \approx \\ &f_y\left[\Delta R_x(1-\frac{z_c}{\hat{z}}) + \frac{\Delta R_z}{f_x}Y_n - \frac{\Delta t_y}{\hat{z}}\right] + \\ &+\hat{Y}\left[\frac{\Delta R_x}{f_y}Y_n - \frac{\Delta R_y}{f_x}X_n - \frac{\Delta t_z}{\hat{z}}\right], \end{aligned} \tag{17}$$

with $Y_n = \hat{Y} + f_y\frac{y_c}{\hat{z}}$ and $X_n = \hat{X} + f_x\frac{x_c}{\hat{z}}$. In this equation, $\hat{z}$ is the depth at pixel position $[\hat{X} \;\; \hat{Y}]^T$ corresponding to image $\hat{I}$. This depth map is obtained by rendering the 3-D object model into a z-buffer.

3.1 Gradient-Based Parameter Estimation

The six unknown extrinsic parameters ΔR_x, ΔR_y, ΔR_z, Δt_x, Δt_y, and Δt_z describe the changes in position and orientation between the two frames $\hat{I}$ and I. These values are estimated using information from spatial and temporal intensity gradients of the images. Assuming that the brightness of an object point is the same in the two frames I and $\hat{I}$, the corresponding pixel intensities in the images

can be related by the 2-D displacement vector $\mathbf{d}$ according to

$$\begin{aligned} \hat{I}(\mathbf{X}) &= I(\mathbf{X}+\mathbf{d}) \\ I(\mathbf{X}) &= \hat{I}(\mathbf{X}-\mathbf{d}). \end{aligned} \tag{18}$$

First order Taylor series expansion of $I(\mathbf{X}+\mathbf{d})$ and $\hat{I}(\mathbf{X}-\mathbf{d})$ around $\mathbf{d} = [0 \quad 0]^T$ and averaging of the two equations of (18) result in the optical flow constraint equation [9]

$$\frac{1}{2}\left(\frac{\partial \hat{I}}{\partial X}+\frac{\partial I}{\partial X}\right)d_x + \frac{1}{2}\left(\frac{\partial \hat{I}}{\partial Y}+\frac{\partial I}{\partial Y}\right)d_y = \frac{\partial \bar{I}}{\partial X}d_x + \frac{\partial \bar{I}}{\partial Y}d_y = \hat{I} - I. \tag{19}$$

Despite the first order intensity approximation, it can be shown that this constraint is even valid for quadratic functions of the intensity due to the use of averaged spatial gradients $\frac{\partial \bar{I}}{\partial X}$ and $\frac{\partial \bar{I}}{\partial Y}$. The gradients are here approximated from 8 samples originating from the two images as suggested in [9].

The optical flow constraint (19) can be set up for any pixel in the image. However, since the two unknown components d_x and d_y of the displacement vector vary from pixel to pixel, the flow constraint is under-determined and cannot be solved directly without additional information. This information is taken directly from the explicit 2-D motion model (17) that describes the displacement vector $\mathbf{d}$ for all image points as a function of the 6 unknown camera parameters. Inserting (17) into the optical flow constraint (19) leads to a set of linear equations for the 6 unknowns

$$\begin{aligned} &a_0\Delta R_x + a_1\Delta R_y + a_2\Delta R_z + a_3\Delta t_x + \\ &+a_4\Delta t_y + a_5\Delta t_z = \hat{I} - I \\ &\mathbf{A}_{ext}\mathbf{p}_{ext} = \mathbf{b}, \end{aligned} \tag{20}$$

with a_0 to a_5 given as

$$\begin{aligned} a_0 &= \frac{\partial \bar{I}}{\partial X}\frac{\hat{X}}{f_y}Y_n + \frac{\partial \bar{I}}{\partial Y}\left(f_y(1-\frac{z_c}{\hat{z}}) + \frac{\hat{Y}}{f_y}Y_n\right) \\ a_1 &= -\frac{\partial \bar{I}}{\partial X}\left(f_x(1-\frac{z_c}{\hat{z}}) + \frac{\hat{X}}{f_x}X_n\right) - \frac{\partial \bar{I}}{\partial Y}\frac{\hat{Y}}{f_x}X_n \\ a_2 &= -\frac{\partial \bar{I}}{\partial X}\frac{f_x}{f_y}Y_n + \frac{\partial \bar{I}}{\partial Y}\frac{f_y}{f_x}X_n \\ a_3 &= -\frac{\partial \bar{I}}{\partial X}\frac{f_x}{\hat{z}}; \qquad a_4 = -\frac{\partial \bar{I}}{\partial Y}\frac{f_y}{\hat{z}} \\ a_5 &= -\frac{\partial \bar{I}}{\partial X}\frac{\hat{X}}{\hat{z}} - \frac{\partial \bar{I}}{\partial Y}\frac{\hat{Y}}{\hat{z}}. \end{aligned} \tag{21}$$

At least six equations are necessary for the algorithm to determine the extrinsic parameters. For robustness, we set up (20) for each pixel corresponding to the object and solve the resulting over-determined system of linear equations in the least-squares sense.

3.2 Hierarchical Analysis-Synthesis Loop

The inherent linearization of the intensity in the optical flow constraint and the approximations used for obtaining a linear solution do not allow dealing with large displacement vectors between two views. To overcome this limitation, a hierarchical scheme in an analysis by synthesis loop [10] is used for the motion estimation as illustrated in Fig. 4. First, an initial approximation $\hat{I}$ of the original

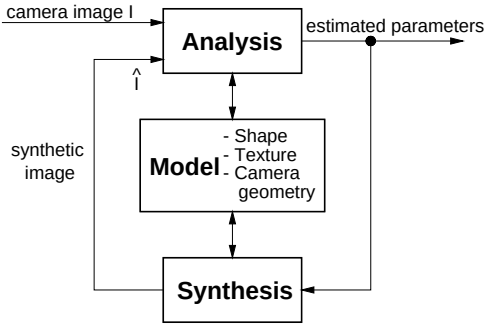

Figure 4: Analysis-synthesis loop of the motion estimation framework.

camera frame I is rendered and both images are low-pass filtered and sub-sampled. As a result, the linear intensity assumption is valid over a wider range and an approximation of the extrinsic camera parameters is estimated from these images. With the estimated parameter set, a motion-compensated image $\hat{I}$ is generated by simply moving the 3-D model of the calibration test object and rendering it at the new position. Due to the motion-compensation, the differences between the synthetic and the camera image decrease. Then, the procedure is repeated at higher resolutions, each time yielding a more accurate motion parameter set.

4 Estimation of Intrinsic Parameters

In the previous section, it is assumed that the internal camera parameters are already known. In this section, the model-based estimator is extended to jointly estimate both extrinsic and intrinsic camera parameters. It turns out that the same linear set of equations as in (20) is obtained with some additional unknowns which account for the unknown internal camera parameters like focal length, aspect ratio, and radial lens distortion parameters. The hierarchical structure with analysis-synthesis loop and the combination of explicit 3-D information with the optical flow constraint remains identical.

The derivation of the displacement field follows the one of Section 3. For the sake of simplicity, the radial lens distortion in the camera frame is compensated for instead of rendering distorted synthetic frames. This way, the motion parameter estimation is performed on undistorted synthetic and camera frames (see Fig. 5) and is therefore not influenced by the lenses. If necessary, synthetic frames

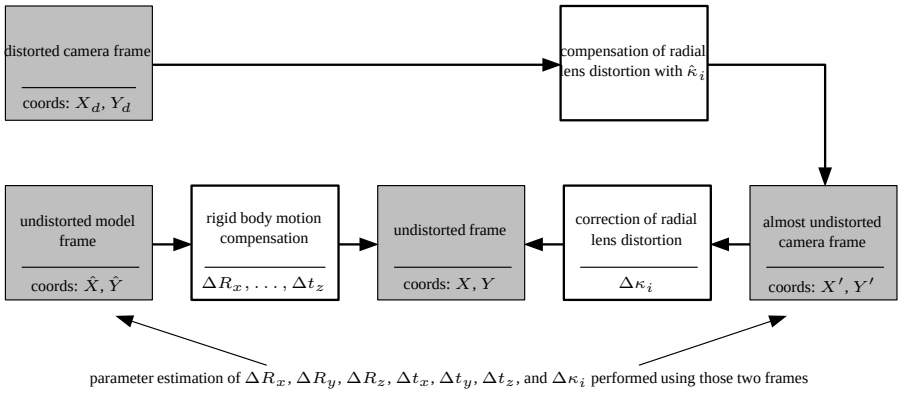

Figure 5: Estimation of radial lens distortion and rigid body motion.

showing the same lens distortion as the camera image can be generated easily by distorting the frame afterwards according to the estimated lens parameters.

Assuming that an initial estimate of the radial lens distortion parameters $\hat{\kappa}_i$ is given, the distorted camera frames with coordinates $\mathbf{X}_d$ can be compensated for, i.e., the distortion described by $\hat{\kappa}_i$. An almost undistorted version of the camera frame with coordinates $\mathbf{X}'$ is created that is used together with the undistorted model frame with coordinates $\hat{\mathbf{X}}$ to estimate rigid body motion and radial lens distortion changes. The compensation of the distorted camera frame I is performed using the radial lens distortion model (9)

$$\mathbf{X}' = \mathbf{X}_d(1 + \sum_{i=1} \hat{\kappa}_i r^{2i}), \tag{22}$$

with the radius r computed according to (10). Since the $\hat{\kappa}_i$ are only estimates, they have to be corrected by the unknown values $\Delta\kappa_i$ to obtain the correct undistorted coordinates $\mathbf{X}$

$$\mathbf{X} = \mathbf{X}_d(1 + \sum_{i=1} (\hat{\kappa}_i + \Delta\kappa_i) r^{2i}). \tag{23}$$

In the error free case, the coordinates $\mathbf{X}$ can also be obtained by motion-compensating the coordinates $\hat{\mathbf{X}}$ of the undistorted model frame with the unknown parameters ΔR_x, ΔR_y, ΔR_z, Δt_x, Δt_y, and Δt_z as depicted in Fig. 5. Eliminating $\mathbf{X}_d$ from (22) and (23) leads to

$$\mathbf{X}' = \mathbf{X} \cdot \frac{1 + \sum_{i=1} \hat{\kappa}_i r^{2i}}{1 + \sum\limits_{i=1} (\hat{\kappa}_i + \Delta\kappa_i) r^{2i}}, \tag{24}$$

where $\mathbf{X}$ can be substituted by the extrinsic parameter model (15). Besides the radial lens distortions, parameters describing the projection onto the image plane are estimated. To decouple changes in focal length and aspect ratio which remains fixed if, e.g., the zoom is altered, f_y is substituted by

$$f_y = f_x \cdot q. \tag{25}$$

Instead of estimating f_x and f_y, f_x and q are determined with f_x specifying the focal length and q being related to the aspect ratio. Since both parameters f_x and q are allowed to change, they are described by the estimates $\hat{f}_x$ and $\hat{q}$ which are corrected by the unknown values Δf and Δq according to

$$\begin{aligned} f_x &= \hat{f}_x(1 + \Delta f) \\ f_y &= f_x \hat{q}(1 + \Delta q). \end{aligned} \tag{26}$$

Inserting (26) together with the replacement for the 3-D coordinates (16) into (24) leads, after first order Taylor series expansion, to the following displacement

$$\begin{aligned} & d_x = X' - \hat{X} \approx \\ & \hat{f}_x \left[-\Delta R_y (1 - \frac{z_c}{\hat{z}}) - \frac{\Delta R_z}{\hat{f}_x \hat{q}} Y_n - \frac{\Delta t_x}{\hat{z}} \right] + \\ & + \hat{X} \left[\frac{\Delta R_x}{\hat{f}_x \hat{q}} Y_n - \frac{\Delta R_y}{f_x} X_n - \frac{\Delta t_z}{\hat{z}} \right] + \\ & + \hat{X} \Delta f - \frac{\hat{X}}{1 + \sum_{i=1} \hat{\kappa}_i r^{2i}} \sum_{i=1} r^{2i} \Delta\kappa_i, \\ & d_y = Y' - \hat{Y} \approx \\ & \hat{f}_x \hat{q} \left[\Delta R_x (1 - \frac{z_c}{\hat{z}}) + \frac{\Delta R_z}{\hat{f}_x} X_n - \frac{\Delta t_y}{\hat{z}} \right] + \\ & + \hat{Y} \left[\frac{\Delta R_x}{\hat{f}_x \hat{q}} Y_n - \frac{\Delta R_y}{\hat{f}_x} X_n - \frac{\Delta t_z}{\hat{z}} \right] + \\ & + \hat{Y} \Delta f + \hat{Y} \Delta q - \frac{\hat{Y}}{1 + \sum_{i=1} \hat{\kappa}_i r^{2i}} \sum_{i=1} r^{2i} \Delta\kappa_i. \end{aligned} \tag{27}$$

This displacement field is quite similar to the one for the extrinsic parameter estimation (17) but contains the addi-

tional unknowns Δf, Δq, and $\Delta\kappa_i$. As a result, by combining (28) with the optical flow constraint (19), a linear set of equations is obtained that can be set up by appending the following terms to the left side of (20)

$$\begin{aligned} a_6\Delta f + a_7\Delta q + \sum_{i=1} a_{8,i}\Delta\kappa_i &= \hat{I} - I \\ \mathbf{A}_{int}\mathbf{p}_{int} &= \mathbf{b}, \end{aligned} \quad (28)$$

with a_6 to a_8 given as

$$\begin{aligned} a_6 &= \tfrac{\partial\bar{I}}{\partial X}\hat{X} + \tfrac{\partial\bar{I}}{\partial Y}\hat{Y} \\ a_7 &= \tfrac{\partial\bar{I}}{\partial Y} \\ a_{8,i} &= \tfrac{r^{2i}}{1+\sum_{j=1}\hat{\kappa}_j r^{2j}}\left(\tfrac{\partial\bar{I}}{\partial X}\hat{X} + \tfrac{\partial\bar{I}}{\partial Y}\hat{Y}\right). \end{aligned} \quad (29)$$

Dependent on the various camera parameters that are being estimated for a particular application, one or more columns can be added without changing the rest of the matrix. The resulting system of equations now looks like

$$[\mathbf{A}_{ext} \;\; \mathbf{A}_{int}]\,[\mathbf{p}_{ext} \;\; \mathbf{p}_{int}]^T = \mathbf{b} \quad (30)$$

with $\mathbf{A}_{int}$ containing the coefficients for the internal parameters according to (28). Given the over-determined system of linear equations, the parameters are estimated in the same way as for the extrinsic parameters in the hierarchical analysis-synthesis framework described in Section 3.2.

5 Multiple Frame Estimation

The use of a linear solution method for the estimation allows the simultaneous exploitation of multiple frames to improve the accuracy of the internal and external camera parameter estimates without disproportionate computational complexity. In some applications, the internal camera parameters have to be consistent with all views even if they are affected by noise. For such cases, the simultaneous estimation of all external parameters together with a common set of internal camera parameters is highly beneficial.

The joint estimation is achieved by simply extending the solution vector of the linear system of equations by the external parameters of the additional frames. Instead of having $6+3$ unknowns (if only one radial lens distortion parameter is used), $N \cdot 6 + 3$ parameters

$$[\mathbf{p}_{ext,1} \;\; \mathbf{p}_{ext,2} \;\; \cdots \;\; \mathbf{p}_{ext,N} \;\; \Delta f \;\; \Delta q \;\; \Delta\kappa] \quad (31)$$

are estimated with N being the number of frames used. The number of equations also increases, since information from all N frames is exploited. The matrix $\mathbf{A}_{cal}$ of the resulting system of equations is given as

$$\mathbf{A}_{cal} = \begin{bmatrix} \mathbf{A}_{ext,1} & \dots & 0 & \mathbf{A}_{int,1} \\ 0 & \dots & 0 & \mathbf{A}_{int,2} \\ \vdots & & \vdots & \vdots \\ 0 & \dots & \mathbf{A}_{ext,N} & \mathbf{A}_{int,N} \end{bmatrix} \quad (32)$$

with $\mathbf{A}_{ext,i}$ and $\mathbf{A}_{int,i}$ containing the information from frame i for the extrinsic and intrinsic parameters, respectively, as specified by (20) and (28).

6 Experimental Results

To quantify the accuracy of the algorithm, synthetic images are rendered in a first experiment using the calibration test object shown on the left hand side of Fig. 1. The experimental setup is specified in Table 1. Different sets of internal and external camera parameters are chosen to synthesize the frames. To simulate manufacturing errors of the calibration body, the 2-D positions of the white points on the test object planes are distorted by adding zero mean normally distributed noise with standard deviation σ to the point locations before rendering the 3-D model. Given the synthetic frames corresponding to different camera settings, internal and external camera parameters (a, φ_{height}, κ_1, R_x, R_y, R_z, t_x, t_y, and t_z) are estimated and compared with the correct values. In Fig. 6, the absolute errors of all estimated parameters

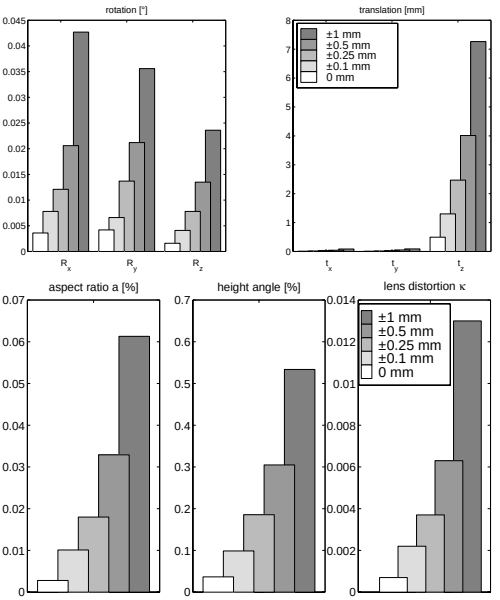

Figure 6: Average absolute error of the estimated internal and external camera parameters for test objects of varying accuracy. Aspect ratio a and height angle φ_{height} are specified relative to their mean $\bar{a} = \frac{4}{3}$ and $\bar{\varphi}_{height} = 0.5$, respectively.

averaged over the 25 frames are shown for different standard deviations ($\sigma = 0$ mm, 0.1 mm, 0.25 mm, 0.5 mm, and 1.0 mm) specifying the displacements of the distorted

number of frames N	25
image size N_x, N_y	CIF resolution (352×288)
object center distance z_c from camera	1.35 m
aspect ratio a	$\lvert a - 4/3 \rvert \leq 0.05$
height angle φ_{height}	$\lvert \varphi_{height} - 0.5 \rvert \leq 0.05$
radial distortion κ_1	$-0.1 \leq \Delta\kappa_1 \leq 0.1$
rotation around x-axis	$-5^\circ \leq \Delta R_x \leq 5^\circ$
rotation around y-axis	$-5^\circ \leq \Delta R_y \leq 5^\circ$
rotation around z-axis	$-5^\circ \leq \Delta R_z \leq 5^\circ$
translation in x-direction	-1 cm $\leq \Delta t_x \leq 1$ cm
translation in y-direction	-1 cm $\leq \Delta t_y \leq 1$ cm
translation in z-direction	-2 cm $\leq \Delta t_z \leq 2$ cm

Table 1: Experimental setup for the synthetic test sequence.

test object point positions. Even if the point positions on the test object are slightly distorted, the parameters can be determined very accurately. Only focal length related to the height angle φ_{height} by (8) and the distance to the test object t_z exhibit larger errors, since the parameters are highly correlated and their errors can compensate each other with only small distortions in the image plane.

6.1 Multiple Frame Estimation

For the verification of the multiple frame calibration algorithm, 36 frames of the calibration object are recorded with a studio camera in CIF resolution from different viewing positions without changing zoom or focus. The intrinsic camera parameters are thus constant. As shown in Fig. 7, the object is not fully visible in all frames.

From these images, both internal and external camera parameters are estimated. The internal camera parameters being determined in this experiment are the aspect ratio a, the height angle φ_{height} and one radial lens distortion parameter κ_1. Since the correct values are not known, the standard deviations of the common parameters are used as an indicator for the accuracy. Fig. 8 shows how the deviations in the internal parameters decrease if more views of the object are used simultaneously. The first bar in this figure refers to the case when each frame is estimated independently of the others whereas the following bars correspond to the case of multiple frame calibration. Using all 36 frames simultaneously leads to an aspect ratio of $a = 1.34$, a height angle $\varphi_{height} = 0.66$, and lens distortion given by $\kappa_1 = 0.18$. This experiment indicates that all available views of a test object should be jointly exploited since not only consistent internal camera parameters are obtained but also their accuracy increases.

6.2 Irregular Test Objects

In contrast to traditional methods where usually regular point grids are used as texture for the calibration test objects, the analysis by synthesis approach allows dealing with more sophisticated surfaces since no feature points have to be searched and no point correspondences have to be established. This can be used to find an initial approximation for the camera parameters. The left hand side of

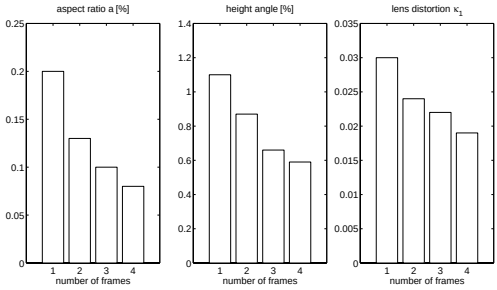

Figure 8: Standard deviation of the internal camera parameters for different numbers of frames used simultaneously in the estimation.

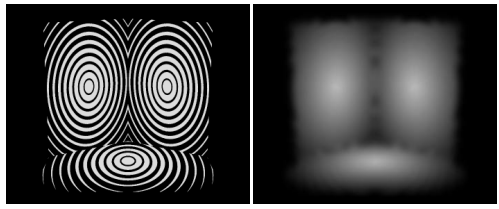

Figure 9: Original and low-pass filtered image of a calibration test object.

Fig. 9 shows a test object with circular structures where the width of the rings is modulated with a low frequency function. When filtering and down-sampling the images in the hierarchical calibration scheme, the low-pass com-

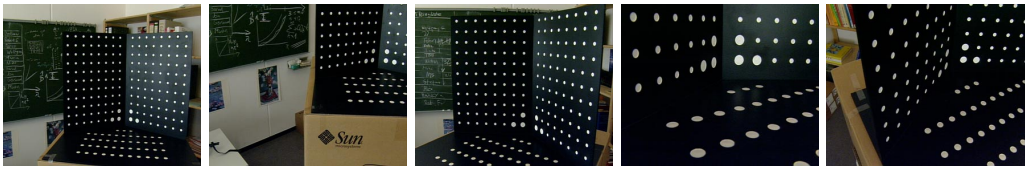

Figure 7: Frames of the test object from different viewing positions.

ponents are preserved and allow the determination of the test object position in spite of image intensity linearizations in the algorithm. In first experiments, deviations of more than 75 pixels from the initial guess are correctly estimated for images in CIF resolution.

7 Conclusions

In this paper, a new method for the determination of intrinsic and extrinsic camera parameters is presented. The model-based algorithm uses a 3-D model of the test objects and estimates the parameters by matching synthesized frames with the camera image in an analysis by synthesis loop. An error prone search for feature points and the establishment of their correspondences can thus be avoided. Also, calibration test bodies with sophisticated textures can be used without additional effort. Experiments on synthetic and real data show the high accuracy of the method which can even be increased by exploiting multiple frames in a joint estimation framework.

References

[1] R. Y. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf cameras and lenses", *IEEE Journal of Robotics and Automation*, vol. RA-3, no. 4, pp. 323–344, Aug. 1987.

[2] I. J. Cox, J. B. Kruskal, and D. A. Wallach, "Predicting and estimating the accuracy of a subpixel registration algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 8, pp. 721–734, Aug. 1990.

[3] P. Brand and R. Mohr, "Accuracy in image measure", in *SPIE Vol. 2350 Videometrics III*, Boston, MA, USA, Nov. 1994, pp. 218–228.

[4] A. Huertas and G. Medioni, "Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 5, pp. 651–664, 1986.

[5] M. R. Shortis, T. A. Clarke, and T. Short, "A comparison of some techniques for the subpixel location of discrete target images", in *SPIE Vol. 2350 Videometrics III*, Boston, MA, USA, Nov. 1994, pp. 239–249.

[6] P. Eisert, *Very Low Bit-Rate Video Coding Using 3-D Models*, PhD thesis, University of Erlangen, Shaker Verlag, Aachen, Germany, 2000.

[7] C. Kolb, D. Mitchell, and P. Hanrahan, "A realistic camera model for computer graphics", in *Proc. Computer Graphics (SIGGRAPH)*, Los Angeles, CA, USA, Aug. 1995, pp. 317–324.

[8] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 965–980, Oct. 1992.

[9] B. K. P. Horn and B. G. Schunck, "Determining optical flow", *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, Aug. 1981.

[10] R. Koch, "Dynamic 3-D scene analysis through synthesis feedback control", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 556–568, June 1993.