GEOMETRIC CAMERA CALIBRATION

INTRODUCTION

Geometric camera calibration is the process of determining geometric properties of a camera. Here, the camera is considered as a ray-based sensing device, and the camera geometry defines how the observed rays of light are mapped onto the image. The purpose of calibration is to discover the mapping between the rays and image points. Hence, a calibrated camera can be used as a direction sensor for which both the forward-projection and back-projection are known, that is, one may compute the image point corresponding to a given projection ray and vice versa.

The geometric calibration of a camera is usually performed by imaging a calibration object whose geometric properties are known. The calibration object often consists of one to three planes that contain visible control points in known positions. The calibration is achieved by fitting a camera model to the observations, which are the measured positions of the control points in the calibration images. The camera model contains two kinds of parameters: The external parameters relate the camera orientation and position to the object coordinate frame, and the internal parameters determine the projection from the camera coordinate frame onto image coordinates. Typically, both the external and the internal camera parameters are estimated in the calibration process, which usually involves nonlinear optimization and minimizes a suitable cost function over the camera parameters. The sum of squared distances between the measured and modeled control point projections is used frequently as the cost function because it gives the maximum likelihood parameter estimates assuming isotropic and independent normally distributed measurement errors.

Calibration by nonlinear optimization requires a good initial guess for the camera parameters. Hence, various methods have been proposed for the direct estimation of the parameters. Most methods deal with conventional perspective cameras, but recently effort has been made in developing models and calibration methods for more general cameras. In fact, the choice of a suitable camera model is an important issue in camera calibration. For example, the pinhole camera model, which is based on the ideal perspective projection model and often used for conventional cameras, is not a suitable model for omnidirectional cameras that have a very large field of view. Hence, a recent trend has occurred towards generic calibration techniques that would allow the calibration of various types of cameras.

In this article, we will provide an overview into geometric camera calibration and its current state. However, because the literature for camera calibration is vast and ever-evolving, it is not possible to cover all aspects in detail. Nevertheless, we hope that this article serves as an introduction to the literature in which more details can be found. The article is structured as follows. First, in the Background section, we describe some historical background for camera calibration. Thereafter, we review different camera models with an emphasis on central cameras. After describing camera models we discuss methods for camera calibration. The focus is on our previous works (1, 2). Finally, in the calibration examples section, we present some calibration examples with real cameras.

BACKGROUND

Geometric camera calibration is a prerequisite for image-based metric three-dimensional measurements, and it has a long history in photogrammetry and computer vision. One of the first references is by Conrady (3), who derived an analytical expression for the geometric distortion in a decentered lens system. Conrady’s model for decentering distortion was used by Brown (4), who proposed a plumb line method for calibrating radial and decentering distortion. Later the approach used by Brown was commonly adopted in photogrammetric camera calibration (5).

In photogrammetry, the emphasis has traditionally been in the rigorous geometric modeling of the camera and optics. On the other hand, in computer vision it is considered important that the calibration procedure is automatic and fast. For example, the well-known calibration method developed by Tsai (6) was designed to be an automatic and efficient calibration technique for machine vision metrology. This method uses a simpler camera model than Brown (4) and avoids the full-scale nonlinear search by using simplifying approximations. However, because of the increased processing power of personal computers, the nonlinear optimization is not as time-consuming now as it was before. Hence, when the calibration accuracy is important the camera parameters are usually refined by a full-scale nonlinear optimization.

Besides increasing the theoretical understanding, the advances in geometric computer vision have also affected the practice of image-based 3-D reconstruction during the last two decades (7,8). For example, although the traditional photogrammetric approach assumes a precalibrated camera, an alternative approach is to compute a projective reconstruction with an uncalibrated perspective camera. The projective reconstruction is defined up to a 3-D projective transformation, and it can be upgraded to a metric reconstruction by self-calibration (8). In self-calibration, the camera parameters are determined without a calibration object; feature correspondences over multiple images are used instead. However, the conventional calibration is typically more accurate and stable than self-calibration. In fact, self-calibration methods are beyond the scope of this article.

CAMERA MODELS

In this section, we describe several camera models that have appeared in the literature. We concentrate on central
cameras, that is, cameras with a single effective viewpoint. Single viewpoint means that all the rays of light that arrive onto the image travel through a single point in space.

Perspective Cameras

The pinhole camera model is the most common camera model, and it is a fair approximation for most conventional cameras that obey the perspective model. Typically these conventional cameras have a small field of view, up to 60°.

The pinhole camera model is widely used and simple; essentially, it is just a perspective projection followed by an affine transformation in the image plane (8).

The pinhole camera geometry is illustrated in Fig. 1. In a pinhole camera, the projection rays meet at a single point that is the camera center \( C \) and its distance from the image plane is the focal length \( f \). By similar triangles, it may be seen in Fig. 1 that the point \( (X_c, Y_c, Z_c)^T \) in the camera coordinate frame is projected to the point \( (fX_c/Z_c, fY_c/Z_c)^T \) in the image coordinate frame. In terms of homogeneous coordinates, this perspective projection can be represented by a \( 3 \times 4 \) projection matrix,

\[
\begin{pmatrix}
X_c \\
y \\
1
\end{pmatrix} 
\approx 
\begin{bmatrix}
f & 0 & 0 & 0 \\
0 & f & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{pmatrix}
x \\
y \\
1
\end{pmatrix}
\]

where \( \approx \) denotes equality up to scale.

However, instead of the image coordinates \( (x, y)^T \), the pixel coordinates \( (u, v)^T \) are usually used, and they are obtained by the affine transformation

\[
\begin{pmatrix}
u \\
v
\end{pmatrix} = \begin{pmatrix}
m_u & -m_u\cot\alpha & u_0 \\
0 & m_v & v_0 \\
m_u f & -m_u f \cot\alpha & u_0 \\
0 & m_v f & v_0
\end{pmatrix} \begin{pmatrix}
x \\
y \\
1
\end{pmatrix} = \begin{pmatrix}
R \\
t
\end{pmatrix} X
\]

where \( (u_0, v_0)^T \) is the principal point, \( \alpha \) is the angle between \( u \) and \( v \) axis, and \( m_u \) and \( m_v \) give the number of pixels per unit distance in \( u \) and \( v \) directions, respectively. The angle \( \alpha \) is \( \pi/2 \) in the conventional case of orthogonal pixel coordinate axes.

In practice, the 3-D point is expressed in some world coordinate system that is different from the camera coordinate system. The motion between these coordinate systems is given by a rotation \( R \) and translation \( t \). Hence, in homogeneous coordinates, the mapping of the 3-D point \( X \) to its image \( m \) is given by

\[
m \approx m_u f \frac{\sin\alpha}{} \begin{pmatrix}
u_0 \\
v_0 \\
u_0 \\
v_0
\end{pmatrix} \begin{pmatrix}
R \\
t
\end{pmatrix} X
\]

where we have introduced the parameters \( \gamma = \frac{m_u}{m_v \sin\alpha} \) and

\[
\begin{pmatrix}
u \\
v
\end{pmatrix} = \begin{pmatrix}
m_u f & \gamma f & u_0 \\
m_u f & \gamma f & v_0 \\
0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
R \\
t
\end{pmatrix} X
\]

\[\text{Figure 1. Pinhole camera model. Here, } C \text{ is the camera center and the origin of the camera coordinate frame. The principal point } p \text{ is the origin of the normalized image coordinate system } (x, y) \text{. The pixel image coordinate system is } (u, v).\]
GEOMETRIC CAMERA CALIBRATION

Cameras, the camera hand submatrix which is called the camera projection matrix and contains the five internal parameters of a pinhole camera.

It follows from Eq. (3) that a general pinhole camera may be represented by a homogeneous $3 \times 4$ matrix

$$
P = K \begin{bmatrix} R & t \end{bmatrix}
$$

which is called the camera projection matrix. If the left hand submatrix $KR$ is nonsingular, as it is for perspective cameras, the camera $P$ is called a finite projective camera.

A camera represented by an arbitrary homogeneous $3 \times 4$ matrix of rank 3 is called a general projective camera. This class covers the affine cameras, which have a projection matrix whose last row is $(0, 0, 0, 1)$ up to scale. A common example of an affine camera is the orthographic camera in which the scene points are orthogonally projected onto the image plane.

**Lens Distortion.** The pinhole camera is an idealized mathematical model for real cameras that may often deviate from the ideal perspective imaging model. Hence, the basic pinhole model is often accompanied with lens distortion models for more accurate calibration of real lens systems. The most important type of geometric distortion is the radial distortion that causes an inward or outward displacement of a given image point from its ideal location. Decentering of lens elements causes additional distortion, which also has tangential components.

A commonly used model for lens distortion accounts for radial and decentering distortion (4, 5). According to this model, the corrected image coordinates $x', y'$ are obtained by

$$
x' = x + \frac{\xi (\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \cdots)}{1 + (\rho_1 r^2 + 2\rho_2 x^2 + \rho_3 (x^2 + y^2))(1 + \rho_4 r^2 + \cdots)}
$$

$$
y' = y + \frac{\eta (\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \cdots)}{1 + (\rho_1 r^2 + 2\rho_2 x^2 + \rho_3 (x^2 + y^2))(1 + \rho_4 r^2 + \cdots)}
$$

where $x$ and $y$ are the measured coordinates, and

$$
\xi = x - x_p
$$

$$
\eta = y - y_p
$$

$$
r = \sqrt{(x - x_p)^2 + (y - y_p)^2}
$$

Here, the center of distortion $(x_p, y_p)$ is a free parameter in addition to the radial distortion coefficients $\kappa_i$ and decentering distortion coefficients $\rho_j$. In the traditional photogrammetric approach, the values for the distortion parameters are computed by least-squares adjustment by requiring that images of straight lines are straight after the correction (4). However, the problem with this approach is that not only the distortion coefficients but also the other camera parameters are initially unknown. For example, the formulation above requires that the scales in both coordinate directions are equal, which is not the case with pixel coordinates unless the pixels are square.

In Ref. 1, the distortion model of Eq. (6) was adapted and combined with the pinhole model to make a complete and accurate model for real cameras. This camera model has the form

$$
m = P(X) = PX + C(PX)
$$

where $P$ denotes the nonlinear camera projection and $P$ is the camera projection matrix of a pinhole camera. The nonlinear part $C$ is the distortion model derived from Eq. (6). In Ref. 1, two parameters were used for both the radial and decentering distortion (i.e., the parameters $\kappa_1, \kappa_2$ and $\rho_1, \rho_2$), and it was assumed that the center of distortion coincides with the principal point of the pinhole camera.

**Central Omnidirectional Cameras.** Although the pinhole model accompanied with lens distortion models is a fair approximation for most conventional cameras, it is not a suitable model for omnidirectional cameras whose field of view is over 180°. The reason is that, when the angle between the incoming light ray and the optical axis of the camera approaches 90°, the perspective projection maps the ray infinitely far in the image, and it is not possible to remove this singularity with the distortion model described above. Hence, more flexible models are needed, and below we discuss different models for central omnidirectional cameras.

**Catadioptric Cameras.** In a catadioptric omnidirectional camera, the wide field of view is achieved by placing a mirror in front of the camera lens. In a central catadioptric camera, the shape and configuration of the mirror are such that the complete catadioptric system has a single effective viewpoint. It has been shown that the mirror surfaces that produce a single viewpoint are surfaces of revolution whose two-dimensional profile is a conic section. Practically useful mirror surfaces used in real central catadioptric cameras are planar, hyperbolic, elliptical, and parabolic. However, a planar mirror does not change the field of view of the camera.

The central catadioptric configurations with hyperbolic, elliptical, and parabolic mirrors are illustrated in Fig. 2. To satisfy the single viewpoint constraint, the parabolic mirror is combined with an orthographic camera, whereas the other mirrors are combined with a perspective camera. In each case, the effective viewpoint of the catadioptric system is the focal point of the mirror denoted by $F$ in Fig. 2.
Single-viewpoint catadioptric image formation is well studied (9, 10), and it has been shown that a central catadioptric projection, including the cases shown in Fig. 2, is equivalent to a two-step mapping via the unit sphere (10, 11). As described in Refs. 11 and 12, the unifying model for central catadioptric cameras may be represented by a composed function $\mathcal{H} \circ \mathcal{F}$ so that

$$m = (\mathcal{H} \circ \mathcal{F})(\Phi)$$

where $\Phi = (\theta, \phi)^T$ defines the direction of the incoming light ray, which is mapped to the image point $m = (u, v, 1)^T$. Here $\mathcal{F}$ first projects the object point onto a virtual image plane, and then the planar projective transformation $\mathcal{H}$ maps the virtual image point to the observed image point $m$. The two-step mapping $\mathcal{F}$ is illustrated in Fig. 3(a), in which the object point $X$ is first projected to $q = (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)$ on the unit sphere, whose center $O$ is the effective viewpoint of the camera. Thereafter, the point $q$ is perspectively projected to $x$ from another point $Q$ so that the line determined by $O$ and $Q$ is perpendicular to the image plane. The distance $l = |OQ|$ is a parameter of the catadioptric camera.

Mathematically the function $\mathcal{F}$ has the form

$$x = \mathcal{F}(\Phi) = r(\theta) \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix}$$

where the function $r$ is the radial projection that does not depend on $\phi$ because of radial symmetry. The precise form of $r$ as a function of $\theta$ is determined by the parameter $l$, that is,

$$r = \frac{(l + 1)\sin \theta}{l + \cos \theta}$$

This follows from the fact that the corresponding sides of similar triangles must have the same ratio, thus $\frac{r}{\sin \phi} = \frac{l + 1}{l + \cos \theta}$, as Fig. 3(a) illustrates.

In a central catadioptric system with a hyperbolic or elliptical mirror, the camera axis does not have to be aligned with the mirror symmetry axis. The camera can be rotated with respect to the mirror as long as the camera center is at the focal point of the mirror. Hence, in the general case, the mapping $\mathcal{H}$ from the virtual image plane to the real image plane is a planar projective transformation (12). However, often the axes of the camera and mirror are close to collinear so that the mapping $\mathcal{H}$ can be approximated with an affine transformation $A$ (11). That is,

$$m = A(x) = K \begin{pmatrix} x \\ 1 \end{pmatrix}$$

where the upper triangular matrix $K$ is defined in Eq. (4) and contains five parameters. (Here the affine transformation $A$ has only five degrees of freedom because we may always fix the camera coordinate frame so that the $x$-axis is parallel to the $\mu$-axis.)

**Fish-Eye Lenses.** Fish-eye cameras achieve a large field of view by using only lenses, whereas the catadioptric cameras use both mirrors and lenses. Fish-eye lenses are designed to cover the whole hemispherical field in front of the camera, and the angle of view is very large, possibly over 180°. Because it is impossible to project the hemispherical field of view on a finite image plane by a perspective projection, the fish-eye lenses are designed to obey some other projection model (13).
The perspective projection of a pinhole camera can be represented by the formula

\[ r = \tan \theta \]  \hspace{1cm} \text{(i. perspective projection)} \tag{12}

where \( \theta \) is the angle between the principal axis and the incoming ray, and \( r \) is the distance between the image point and the principal point measured on a virtual image plane that is placed at a unit distance from the pinhole. Fish-eye lenses instead are usually designed to obey one of the following projections:

\[ r = 2 \tan(\theta/2) \]  \hspace{1cm} \text{(ii. stereographic projection)} \tag{13}

\[ r = \theta \]  \hspace{1cm} \text{(iii. equidistance projection)} \tag{14}

\[ r = 2 \sin(\theta/2) \]  \hspace{1cm} \text{(iv. equisolid angle projection)} \tag{15}

\[ r = \sin(\theta) \]  \hspace{1cm} \text{(v. orthogonal projection)} \tag{16}

where the equidistance projection is perhaps the most common model. The behavior of the different projection models is illustrated in Fig. 3(b).

Although the central catadioptric cameras and fish-eye cameras have a different physical construction, they are not too different from the viewpoint of mathematical modeling. In fact, the radial projection curves defined by Eq. (10) are similar to those shown in Fig. 3(b). In particular, when \( l = 0 \) Eq. (10) defines the perspective projection, \( l = 1 \) gives the stereographic projection (because \( \tan \frac{\theta}{2} = \frac{\sin \theta}{1 + \cos \theta} \)), and on the limit \( l \to \infty \) we obtain the orthogonal projection. Hence, the problem of modeling radially symmetric central cameras is essentially reduced to modeling radial projection functions such as those in Fig. 3(b).

**Generic Model for Central Cameras.** Here we describe a generic camera model, which was proposed in Ref. 2 and is suitable for central omnidirectional cameras as well as for conventional cameras. As discussed above, the radially symmetric central cameras may be represented by Eq. (10) where the function \( F \) is given by Eq. (9). The radial projection function \( r \) in \( F \) is an essential part of the model. If \( r \) is fixed to have the form of Eq. (12), then the model is reduced to the pinhole model. However, modeling of omnidirectional cameras requires a more flexible model and here we consider the model

\[ r = k_1 \theta + k_2 \theta^3 + k_3 \theta^5 + k_4 \theta^7 + k_5 \theta^9 + \cdots \] \hspace{1cm} \text{\textit{(17)}}

which allows good approximation of all the projections in Fig. 3(b). In Ref. 2 it was shown that the first five terms, up to the ninth power of \( \theta \), give enough degrees of freedom for accurate approximation of different projection curves. Hence, the generic camera model used here contains five parameters in the radial projection function \( r \).
However, real lenses may deviate from precise radial symmetry and, therefore, the radially symmetric model above was supplemented with an asymmetric part in Ref. 2. Hence, instead of Eq. (10) the camera model proposed in Ref. 2 has the form

\[
m = (A \circ D \circ F) \times (\Phi)
\]  

(18)

where \(D\) is the asymmetric distortion function so that \(D \circ F\) gives the distorted image point \(x_d\), which is then transformed to pixel coordinates by the affine transformation \(A\). In detail,

\[
x_d = (D \circ F)(\Phi) = r(\theta)u_r(\phi) + \Delta_r(\theta, \phi)u_r(\phi) + \Delta_t(\theta, \phi)u_\phi(\phi)
\]  

(19)

where \(u_r(\phi)\) and \(u_\phi(\phi)\) are the unit vectors in the radial and tangential directions and

\[
\Delta_r(\theta, \phi) = (g_1 \theta + g_2 \theta^3 + g_3 \theta) (i_1 \cos \phi + i_2 \sin \phi + i_3 \cos 2\phi + i_4 \sin 2\phi)
\]  

(20)

\[
\Delta_t(\theta, \phi) = (h_1 \theta + h_2 \theta^3 + h_3 \theta) (j_1 \cos \phi + j_2 \sin \phi + j_3 \cos 2\phi + j_4 \sin 2\phi)
\]  

(21)

Here both the radial and tangential distortion terms contain seven parameters.

The asymmetric part in Eq. (19) models the imperfections in the optical system in a somewhat similar manner as the distortion model by Brown (4). However, instead of rigorous modeling of optical distortions, here the aim is to provide a flexible mathematical distortion model that is just fitted to agree with the observations. This approach is often practical because several possible sources of imperfections may exist in the optical system, and it is difficult to model all of them in detail.

The camera model defined above is denoted by \(M_{24}\) in the following because the number of parameters is 24: \(F\) and \(A\) have both 5 parameters and \(D\) has 14 parameters. However, often it is assumed that the pixel coordinate system is orthogonal [i.e., \(s = 0\) in Eq. (4)], so that the number of parameters in \(A\) is only four. This model is denoted by \(M_{25}\). In addition, sometimes it may be useful to leave out the asymmetric part to avoid overfitting. The corresponding radially symmetric models are here denoted by \(M_9\) and \(M_{16}\). The model \(M_{26}\) contains only two terms in Eq. (17) whereas \(M_9\) contains five.

**Other Distortion Models.** In addition to the camera models described in the previous sections, also several other models have appeared in the literature. For example, the so-called division model for radial distortion is defined by

\[
r = \frac{r_d}{1 - cr_d^2}
\]  

(22)

where \(r_d\) is the measured distance between the image point and the distortion center and \(r\) is the ideal undistorted distance (14, 15). A positive value of the distortion coefficient \(c\) corresponds to the typical case of barrel distortion (14). However, the division model is not suitable for cameras whose field of view exceeds 180°. Hence, other models must be used in this case and, for instance, the two-parametric projection model

\[
r = \frac{a - \sqrt{a^2 - 4b\theta^2}}{2b\theta}
\]  

(23)

has been used for fish-eye lenses (16). Furthermore, a parameter-free method for determining the radial distortion was proposed in Ref. 17.

**Noncentral Cameras**

Most real cameras are strictly speaking noncentral. For example, in the case of parabolic mirror in Fig. 2, it is difficult to align the mirror axis and the axis of the camera precisely. Likewise, in the hyperbolic and elliptic configurations, the precise positioning of the optical center of the perspective camera in the focal point of the mirror is practically infeasible. In addition, if the shape of the mirror is not a conic section or the real cameras are not truly orthographic or perspective the configuration is noncentral. However, in practice the camera is usually negligibly small compared with the viewed region so that it is effectively point-like. Hence, the central camera models are widely used and tenable in most situations so that also here, in this article, we concentrate on central cameras.

Still, some works suggest the single viewpoint constraint is relaxed and a noncentral camera model is used. For example, a completely generic camera calibration approach was discussed in Refs. 18 and 19, in which a nonparametric camera model was used. In this model, each pixel of the camera is associated with a ray in 3-D and the task of calibration is to determine the coordinates of these rays in some local coordinate system. In addition, work has been published about designing mirrors for noncentral catadioptric systems that are compliant with predefined requirements (20).

Finally, as a generalization of central cameras, we would like to mention the axial cameras in which all the projection rays go through a single line in space (19). For example, a catadioptric camera that consists of a mirror and a central camera is an axial camera if the mirror is any surface of revolution and the camera center lies on the mirror axis of revolution. A central camera is a special case of an axial camera. The equiangular (21, 22) and equiareal (23) catadioptric cameras are other classes of axial cameras. In equiareal cameras, the projection is area preserving, whereas the equiangular mirrors are designed so that the radial distance measured from the center of symmetry
in the image is linearly proportional to the angle between the incoming ray and the optical axis.

CALIBRATION METHODS

Camera calibration is the process of determining the parameters of the camera model. Here we consider conventional calibration techniques that use images of a calibration object that contains control points in known positions. The choice of a suitable calibration algorithm depends on the camera model, and below we describe methods for calibrating both perspective and omnidirectional central cameras.

Although the details of the calibration procedure may differ depending on the camera, the final step of the procedure is usually the refinement of camera parameters by nonlinear optimization regardless of the camera model. The cost function normally used in the minimization is the sum of squared distances between the measured and modeled control point projections, that is

$$\sum_{j=1}^{N} \sum_{i=1}^{N} d_j^2(m_j^i, m_i^j)^2$$

(24)

where $m_j^i$ contains the measured image coordinates of the control point $i$ in the view $j$, the binary variable $\delta_i$ indicates whether the control point $i$ is observed in the view $j$ and

$$m_j^i = P_j(X_i)$$

(25)

is the projection of the control point $X_i$ in the view $j$. Here, $P_j$ denotes the camera projection in the view $j$ and it is determined by the external and internal camera parameters. The justification for minimizing Eq. (24) is that it gives the maximum likelihood solution for the camera parameters when the image measurement errors obey a zero-mean isotropic Gaussian distribution. However, the successful minimization of Eq. (24) with standard local optimization methods requires a good initial guess for the parameters. Methods for computing such an initial guess are discussed below.

Perspective Cameras

In the case of a perspective camera, the camera projection $P$ is represented by a $3 \times 4$ matrix $P$ as described in the section on perspective cameras. In general, the projection matrix $P$ can be determined from a single view of a non-coplanar calibration object using the Direct Linear Transform (DLT) method, which is described below. Then, given $P$, the parameters $K$ and $\mathbf{R}$ in Eq. (5) are obtained by decomposing the left $3 \times 3$ submatrix of $P$ using the QR-decomposition whereafter also $\mathbf{t}$ can be computed (8).

On the other hand, if the calibration object is planar and the internal parameters in $K$ are all unknown, several views are needed. In this case, the constant camera calibration matrix $K$ can be determined first using the approach described below in the section on planar calibration object. Thereafter, the view dependent parameters $R_j$ and $t_j$ can be computed and used for initializing the nonlinear optimization. If the perspective camera model is accompanied with a lens distortion model, then the distortion parameters in Eq. (7) may be initialized by setting them to zero (24).

Noncoplanar Calibration Object. Assuming that the known space points $X_i$ are projected at the image points $m_i^j$, the unknown projection matrix $P$ can be estimated using the DLT method (8, 25, 26). The projection equation gives

$$m_i^j = PX_i^j$$

(26)

which can be written in the equivalent form

$$m_i^j \times PX_i^j = 0$$

(27)

where the unknown scale in Eq. (26) is eliminated by the cross product. The equations above are linear in the elements of $P$ so they can be written in the form

$$A^iv = 0$$

(28)

where

$$v = (P_{11}P_{12}P_{13}P_{14}P_{21}P_{22}P_{23}P_{24}P_{31}P_{32}P_{33}P_{34})^T$$

(29)

and

$$A^i = \begin{bmatrix} 0^T & -m_3X_i^T & m_2X_i^T \\ m_3X_i^T & 0^T & -m_1X_i^T \\ -m_2X_i^T & m_1X_i^T & 0^T \end{bmatrix}$$

(30)

Thus, each point correspondence provides three equations, but only two of them are linearly independent. Hence, given $M \geq 6$ point correspondences, we get an overdetermined set of equations $Av = 0$, where the matrix $A$ is obtained by stacking the matrices $A^i, i = 1, \ldots, M$. In practice, because of the measurement errors no exact solution to these equations exist, but the solution $v$ that minimizes $||Av||$ can be computed using the singular value decomposition of $A$ (8). However, if the points $X_i$ are coplanar, then ambiguous solutions exist for $v$, and hence the DLT method is not applicable in such case.

The DLT method for solving $P$ is a linear method that minimizes the algebraic error $||Av||$ instead of the geometric error in Eq. (24). This method implies that, in the presence of noise, the estimation result depends on the coordinate frames in which the points are expressed. In practice, it has been observed that a good idea is to normalize the coordinates in both $m_i^j$ and $X_i$ so that they have zero mean and unit variance. This kind of normalization may significantly improve the estimation result in the presence of noise (8).

Planar Calibration Object. In the case of a planar calibration object, the camera calibration matrix $K$ can be solved by using several views. This approach was described in Refs. 24 and 27, and it is briefly summarized in the following.
The mapping between a scene plane and its perspective image is a planar homography. Because one may assume that the calibration plane is the plane $Z = 0$, the homography is defined by

$$m \simeq K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

(31)

where the $3 \times 3$ homography matrix

$$H = K [r_1 \quad r_2 \quad t]$$

(32)

where the columns of the rotation matrix $R$ are denoted by $r^i$. The outline of the calibration method is to first determine the homographies for each view and then use Eq. (32) to derive constraints for the determination of $K$. The constraints for $K$ are described in more detail below. Methods for determining a homography from point correspondences are described, for example, in Ref. 8.

Denoting the columns of $H$ by $h^j$ and using the fact that $r^1$ and $r^2$ are orthonormal one obtains from Eq. (32) that

$$h^j^T K^{-1} K^{-1} h^2 = 0$$

(33)

$$h^j^T K^{-1} K^{-1} h^1 = h^2^T K^{-1} K^{-1} h^2$$

(34)

Thus, each homography provides two constraints that may be written as linear equations on the elements of the homogeneous symmetric matrix $\omega = K^{-1} K^{-1}$. Hence, the system of equations, derived from Eqs. (33) and (34) above, is of the form $A \omega = 0$, where the vector of unknowns $\omega = (\omega_{11}, \omega_{12}, \omega_{13}, \omega_{22}, \omega_{23}, \omega_{33})^T$ consists of the elements of $\omega$. Matrix $A$ has $2N$ rows, where $N$ is the number of views. Given three or more views, the solution vector $v$ is the right singular vector of $A$ that corresponds to the smallest singular value. When $\omega$ is solved (up to scale), one may compute the upper triangular matrix $K$ by Cholesky-factorization. Thereafter, given $H$ and $K$, the external camera parameters can be retrieved from Eq. (32). Finally, the obtained estimates should be refined by minimizing the error of Eq. (24) in all views.

**Omnidirectional Cameras**

In this section, we describe a method that uses a planar calibration pattern to calibrate the parameters of the camera model introduced in the section entitled “Generic Model for Central Cameras” (2). Planar calibration patterns are very common because they are easy to create. In fact, often also the noncoplanar calibration objects contain planar patterns because they usually consist of two or three different planes.

The calibration procedure consists of four steps that are described below. We assume that $M$ control points are observed in $N$ views so that, for each view $j$, a rotation matrix $R_j$ and a translation vector $t_j$, exist, which describe the orientation and position of the camera with respect to the calibration object. In addition, we assume that the object coordinate frame is chosen so that the plane $Z = 0$ contains the calibration pattern and the coordinates of the control point $i$ are denoted by $X^i = (X^i, Y^i, 0)^T$. The corresponding homogeneous coordinates in the calibration plane are denoted by $x^i_j = (X^i, Y^i, 1)^T$ and the observed image coordinates in the view $j$ are $m^j = (u^j, v^j, 1)^T$.

**Step 1: Initialization of internal parameters.**

In the first three steps of the calibration procedure, we use the camera model $M_6$, which contains only six nonzero internal parameters i.e., the parameters $(k_1, k_2, f, \gamma, u_0, v_0)$. These parameters are initialized using a priori knowledge about the camera. For example, the principal point $(u_0, v_0)$ is usually located close to the image center, $\gamma$ has a value close to 1, and $f$ is the focal length in pixels. The initial values for $k_1$ and $k_2$ can be obtained by fitting the model $r = k_1 \theta + k_2 \theta^2$ to the desired projection curve in Fig. 3(b).

**Step 2: Back-projection and computation of homographies.**

Given the internal parameters, we may back-project the observed points $m^j$ onto the unit sphere centered at the camera origin. For each $m^j$, the back-projection gives the direction $\Phi^j = (\phi^j_1, \phi^j_2)^T$ and the points on the unit sphere are defined by $q^j = (\sin \phi^j_1 \sin \phi^j_2, \cos \phi^j_1 \sin \phi^j_2, \cos \phi^j_2)^T$. Because the mapping between the points on the calibration plane and on the unit sphere is a central projection, a planar homography $H_j$ exists so that $q^j \sim H_j x^j$. For each view $j$, the homography $H_j$ is estimated from the correspondences $(q^j, x^j)$. In detail, the initial estimate for $H_j$ is computed by the linear algorithm (8), and it is then refined by minimizing $\sum \sin^2 \phi^j_2$, where $\phi^j_2$ is the angle between the unit vectors $q^j$ and $H_j x^j/||H_j x^j||$.

**Step 3: Initialization of external parameters.**

The initial values for the external camera parameters are extracted from the homographies $H_j$. It holds that

$$q^j \simeq \begin{bmatrix} R_j & t_j \end{bmatrix} \begin{bmatrix} X^j \\ Y^j \\ 0 \end{bmatrix} = \begin{bmatrix} r^j & r^j & t_j \end{bmatrix} \begin{bmatrix} X^j \\ Y^j \\ 1 \end{bmatrix}$$

which implies $H_j \simeq [r^j \; r^j \; t_j]$. Hence,

$$r^j = \lambda_j h^j_1, \quad r^j = \lambda_j h^j_2, \quad r^j = r^j \times r^j, \quad t_j = \lambda_j h^j_3$$

where $\lambda_j = \pm ||h^j||^{-1}$. The sign of $\lambda_j$ can be determined by requiring that the camera is always on the front side of the calibration plane. However, the obtained rotation matrices may not be orthogonal because of estimation errors. Hence, the singular value decomposition is used to compute the closest orthogonal matrices in the sense of Frobenius norm, which are then used for initializing each $R_j$. 

Step 4: Minimization of projection error.

If a camera model with more than six parameters is used, then the additional camera parameters are initialized to zero at this stage. As we have the estimates for the internal and external camera parameters, we may compute the imaging function \( P_i \) for each camera, in which a control point is projected to \( m_i = P_i(X) \). Finally, all the camera parameters are refined by minimizing Eq. (24) using nonlinear optimization, such as the Levenberg-Marquardt algorithm.

Precise Calibration with Circular Control Points

To achieve an accurate calibration, we have used a calibration plane with circular control points because the centroids of the projected circles can be detected with a sub-pixel level of accuracy (28). However, in this case the problem is that the centroid of the projected circle is not the image of the center of the original circle. Therefore, because \( m_i \) in Eq. (24) is the measured centroid, we should not project the centers as points \( m_i \) because it may introduce bias in the estimates. Of course, it is not an issue if the control points are really pointlike, such as the corners of a checkerboard pattern.

In the case of a perspective camera, the centroids of the projected circles can be solved analytically given the camera parameters and the circles on the calibration plane (1). However, in the case of more generic camera models, the projection is more complicated and the centroids of the projected circles must be solved numerically (2).

CALIBRATION EXAMPLES

In this section, we illustrate camera calibration with real examples involving different kinds of cameras.

Conventional Cameras with Moderate Lens Distortion

The first calibrated camera was a Canon S1 IS digital camera with a zoom lens whose focal length range is 5.8–58.0 mm, which corresponds to a range of 38–380 mm in the 35-mm film format. The calibration was performed with the zoom fixed to 11.2 mm. Hence, the diagonal field of view was about 30°, which is a relatively narrow angle. The other camera was a Sony DFW-X710 digital video camera equipped with a Cosmicar H4/16 wide-angle lens. The focal length of this wide-angle lens is 4.2 mm, and it produces a diagonal field of view of about 80°.

Both cameras were calibrated by using a planar calibration pattern that contains white circles on black background. The pattern was displayed on a digital plasma display (Samsung PPM50M6HS) whose size is 1204 x 724 mm \(^2\). A digital flat-screen display provides a reasonably planar object and because of its self-illuminating property, it is easy to avoid specular reflections that might otherwise hamper the accurate localization of the control points. Some examples of the calibration images are shown in Fig. 4. The image in Fig. 4(a) was taken with the narrow-angle lens and the image in Fig. 4(b) with the wide-angle lens. The resolution of the images is 2048 x 1536 pixels and 1024 x 768 pixels, respectively. The lens distortion is clearly visible in the wide-angle image.

The number of calibration images was six for both cameras, and each image contained 220 control points. The images were chosen so that the whole image area was covered by the control points. In addition to the set of calibration images we took another set of images that likewise contained six images in which the control points were distributed onto the whole image area. These images were used as a test set to validate the results of calibration. In all cases, the control points were localized from the images by computing their grayscale centroids (28).

The cameras were calibrated using four different camera models. The first model, denoted by \( M_1 \), was the skew-zero pinhole model accompanied with four distortion parameters. This model was used in Ref. 1, and it is described in the section on lens distortion. The other three models were the models \( M_6, M_9 \), and \( M_{23} \) defined in the section entitled “Generic Model for Central Cameras.” All calibrations were performed by minimizing the sum of squared projection errors as described in the section on calibration methods. The computations were carried out by using the publicly available implementations of the calibration procedures proposed in Refs. 1 and 2. \(^1\)

The calibration results are shown in Table 1 in which the first four rows give the figures for the narrow-angle and wide-angle lenses introduced above. The first and third row in Table 1 contain the RMS calibration errors, that is, the root-mean-squared distances between the measured and modeled control point projections in the calibration images. The second and fourth row show the RMS projection errors in the test images. In the test case, the values of the internal camera parameters were those estimated from the calibration images, and only the external camera parameters were optimized by minimizing the sum of the squared projection errors.

The results illustrate that the most flexible model \( M_{23} \) generally performs the best. However, the difference between the models \( M_6, M_9 \), and \( M_{23} \) is not large for the narrow-angle and wide-angle lens. The model \( M_9 \) performs well with the narrow-angle lens, but it seems that the other models are better in modeling the severe radial distortion of the wide-angle lens. The relatively low values of the test set error indicate that the risk of overfitting is small. This risk could be decreased even more by using more calibration images. The RMS error is somewhat larger for the narrow-angle lens than for the wide-angle lens, but it may be caused by the fact that the image resolution is higher in the narrow-angle case. Hence, the pixel units are not directly comparable for the different cameras.

\(^1\)http://www.ee.oulu.fi/mvg/page/downloads.
Omnidirectional Cameras

We calibrated a fish-eye lens camera and two different catadioptric cameras. The fish-eye lens that was used in the experiments is the ORIFL190-3 lens manufactured by Omnitech Robotics (Englewood, CA). This lens has a 190° field of view, and it was attached to a PointGrey Dragonfly (PointGrey, Vancouver, BC, Canada) color video camera, which has a resolution of 1024\(\times\)768 pixels. The catadioptric cameras were constructed by placing two different mirrors in front of the Canon S1 IS camera (Canon, Tokyo, Japan), which has a resolution of 2048\(\times\)1536 pixels. The first mirror was a hyperbolic mirror from Eizoh (Tokyo, Japan) and the other mirror was an equiangular mirror from Kaidan (Feasterville, PA). The field of view provided by the hyperbolic mirror is such that when the mirror is placed above the camera so that the optical axis is vertical, the camera observes about 30° above and 50° below the horizon (the region directly below the mirror is obscured by the camera). The equiangular mirror by Kaidan provides a slightly larger view of field because it observes about 50° above and below the horizon. In the azimuthal direction the viewing angle is 360° for both mirrors.

Because the field of view of all three omnidirectional cameras exceeds a hemisphere, the calibration was not performed with the model \(M_p\), which is based on the perspective projection model. Hence, we only report the results obtained with the central camera models \(M_6\), \(M_9\), and \(M_{23}\). The calibration experiments were conducted in a similar way as for the conventional cameras above, and the same calibration object was used. However, here the number of images was 12 both in the calibration set and the test set. The number of images was increased to have a better coverage for the wider field of view.

The results are illustrated in Table 1, where it can be observed that the model \(M_{23}\) again shows best performance. The radially symmetric model \(M_9\) performs almost equally well with the fish-eye camera and the equiangular camera. However, for the hyperbolic camera, the additional degrees of freedom in \(M_{23}\) clearly improve the calibration accuracy. This improvement might be an indication that the optical axis of the camera is not aligned precisely with the mirror axis. Nevertheless, the asymmetric central camera model \(M_{23}\) provides a good approximation for this catadioptric camera. Likewise, here the central model

### Table 1. The RMS projection errors in pixels

<table>
<thead>
<tr>
<th></th>
<th>(M_6)</th>
<th>(M_9)</th>
<th>(M_{23})</th>
</tr>
</thead>
<tbody>
<tr>
<td>narrow-angle lens</td>
<td>0.293</td>
<td>0.339</td>
<td>0.325</td>
</tr>
<tr>
<td>test set error</td>
<td>0.249</td>
<td>0.309</td>
<td>0.259</td>
</tr>
<tr>
<td>wide-angle lens</td>
<td>0.908</td>
<td>0.078</td>
<td>0.077</td>
</tr>
<tr>
<td>test set error</td>
<td>0.823</td>
<td>0.089</td>
<td>0.088</td>
</tr>
<tr>
<td>fish-eye lens</td>
<td>-</td>
<td>0.359</td>
<td>0.233</td>
</tr>
<tr>
<td>test set error</td>
<td>-</td>
<td>0.437</td>
<td>0.168</td>
</tr>
<tr>
<td>hyperbolic mirror</td>
<td>-</td>
<td>4.178</td>
<td>1.225</td>
</tr>
<tr>
<td>test set error</td>
<td>-</td>
<td>3.708</td>
<td>1.094</td>
</tr>
<tr>
<td>equiangular mirror</td>
<td>-</td>
<td>2.716</td>
<td>0.992</td>
</tr>
<tr>
<td>test set error</td>
<td>-</td>
<td>3.129</td>
<td>1.065</td>
</tr>
</tbody>
</table>

---

Figure 4. Images of the calibration pattern taken with different types of cameras, a) narrow-angle lens, b) wide-angle lens, c) fish-eye lens, d) hyperbolic mirror combined with a narrow-angle lens.
seems to be tenable also for the equiangular catadioptric camera, which is strictly speaking noncentral. Note that the sensor resolution of the fish-eye camera was different than that of the catadioptric cameras.

CONCLUSION

Geometric camera calibration is a prerequisite for image-based, accurate 3-D measurements, and it is therefore a fundamental task in computer vision and photogrammetry. In this article, we presented a review of calibration techniques and camera models that commonly occur in applications. We concentrated on the traditional calibration approach in which the camera parameters are estimated by using a calibration object whose geometry is known. The emphasis was on central camera models, which are the most common in applications and provide a reasonable approximation for a wide range of cameras. The process of camera calibration was additionally demonstrated with practical examples in which several different kinds of real cameras were calibrated.

Camera calibration is a wide-ranging topic and much research exists that was not possible to be covered here. For example, recently research efforts have focused toward completely generic camera calibration techniques that could be used for all kinds of cameras, including the noncentral ones. In addition, camera self-calibration is an active research area that was not discussed in the scope of this article. However, camera calibration using a calibration object and a parametric camera model, as discussed here, is the most viable approach when a high level of accuracy is required.

BIBLIOGRAPHY


JUHO KANNALA
JANNE HEIKKILA
SAMT S. BRANDT
University of Oulu
Oulu, Finland