Calibration of a Moving Camera Using a Planar Pattern: Optimal Computation, Reliability Evaluation and Stabilization by Model Selection

Chikara Matsunaga[1] and Kenichi Kanatani[2]

[1] Broadcast Division, FOR-A Co. Ltd.,
2–3–3 Ohsaku, Sakura, Chiba 285-0802 Japan
matsunaga@for-a.co.jp
[2] Department of Computer Science, Gunma University,
Kiryu, Gunma 376-8515 Japan
kanatani@cs.gunma-u.ac.jp

Abstract. We present a scheme for *simultaneous calibration* of a continuously moving and continuously zooming camera: placing an easily distinguishable pattern in the scene, we calibrate the camera from an unoccluded portion of the pattern image in each frame. We describe an optimal method which provides an evaluation of the reliability of the solution. We then propose a technique for avoiding the inherent degeneracy and statistical fluctuations by *model selection* using the *geometric AIC* and the *geometric MDL*.

1 Introduction

Visually presenting 3-D shapes of real objects is one of the main goals of many Internet applications such as network cataloging and virtual museums. Today, generating virtual images by embedding graphics objects in real scenes or real objects in graphics scenes, known as *mixed reality*, is one of the central themes of image and media applications. In order to reconstruct the 3-D shapes of real objects or scenes for such applications, we need to know the 3-D position of the camera that we use and its internal parameters. Thus, camera calibration is a first step in all vision and media applications.

The standard method for it is *pre-calibration*: the camera internal parameters are determined from images of objects or patterns of known 3-D geometry in a controlled environment [1, 18, 29, 34, 36, 37]. Recently, techniques for computing both the camera parameters and the 3-D positions of the camera from an image sequence of the scene about which we have no prior knowledge have intensively been studied [3, 24]. Such a technique, known as *self-calibration*, may be useful in unknown environments such as outdoors. For stable reconstruction, however, it requires a long sequence of images taken from unconstrained camera positions and feature matching among frames. As a result, the amount of computation is too large for real-time applications, and it cannot be applied if the camera motion is constrained or the scene changes as the camera moves unless we are

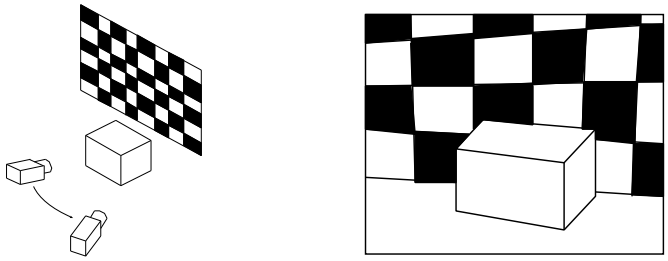


Fig. 1. Simultaneous calibration of a moving camera: we observe an unoccluded part of the image of a planar pattern placed in the scene.

given a priori information about the constraint or the scene change (see, e.g., [6, 9, 28] for self-calibration based on a priori information about the camera motion).

In this paper, we focus on *virtual studio* applications [7, 30]: we take images of moving objects such as persons and superimpose them in a graphics-generated background in real time by computing the 3-D positions and zooming of a moving camera. Since the scene as well as the position and zooming of the camera changes from frame to frame, we cannot pre-calibrate or self-calibrate the camera.

This difficulty can be overcome by placing an easily distinguishable planar pattern with a known geometry in the scene (Fig. 1): we detect an unoccluded portion of the pattern image in each frame, compute the 3-D position and zooming of the camera from it, and remove the pattern image by segmentation. We call this strategy *simultaneous calibration*. It has many elements that do not appear in pre-calibration:

1. While manual interventions can be employed in pre-calibration, simultaneous calibration must be completely automated. In particular, we must automatically identify the 3-D positions of the marker points that are unoccluded in each frame.
2. Since the number of unoccluded marker points is different in each frame, the accuracy of calibration is different from frame to frame. Hence, not only do we need an accurate computational procedure but also a scheme for evaluating the reliability of the computed solution.
3. Since we have no control over the camera position relative to the pattern, degenerate configurations can occur: when the camera optical axis is perpendicular to the pattern, the 3-D position and focal length of the camera are indeterminate because zooming out and moving the camera forward cause the same visual effect.
4. As the object moves in the scene, some unoccluded marker points become occluded while others become occluded. As a result, the computed camera position may not be the same even if the camera is stationary in the scene. This type of statistical fluctuations becomes conspicuous when the camera motion is small.

In this paper, we introduce a statistical model of image noise and describe a procedure for computing an optimal solution that attains the *Cramer-Rao lower bound* (*CRLB*) in the presence of noise. As a result, we can evaluate the reliability of the solution by computing an estimate of the CRLB.

We then show that degeneracy and statistical fluctuations can be avoided by *model selection*. At each frame, we predict the 3-D position and zooming of the camera in multiple ways from the past history. We then evaluate the goodness of each prediction, or *model*, and adopt the best one. In this paper, we use the *geometric AIC* introduced by Kanatani [12, 14] and the *geometric MDL* to be defined shortly as the model selection criterion.

The geometric MDL we use is different from the traditional MDL used in statistics and some vision applications [8, 11, 21, 22, 31]. We compare the performances of the geometric AIC and the geometric MDL by doing numerical simulations and real image experiments.

2 Basic Principle

We fix an XYZ world coordinate system in the scene and place a planar pattern in parallel to the XY plane at a known distance d. We imagine a hypothetical camera with a known focal length f_0 placed at the world origin O in such a way that the optical axis coincides with the Z-axis and the image x- and y-axes are parallel to the X- and Y-axes. The 3-D position of the actual camera is regarded as obtained by rotating the hypothetical camera by $\boldsymbol{R}$ (rotation matrix), translating it by $\boldsymbol{t}$, and changing the focal length into f; we call $\{\boldsymbol{t}, \boldsymbol{R}\}$ the *motion parameters*. We regard the focal length f as a single unknown internal parameter, assuming that other parameters, such as the image skew and the aspect ratio, have already been pre-calibrated so that the imaging geometry can be modeled as a perspective projection.

Suppose N points on the planar pattern with known coordinates (X_α, Y_α, d) are observed at (x_α, y_α) in the image. If we define the 3-D vectors

$$\bar{\boldsymbol{x}}_\alpha = \begin{pmatrix} X_\alpha/d \\ Y_\alpha/d \\ 1 \end{pmatrix}, \qquad \boldsymbol{x}_\alpha = \begin{pmatrix} x_\alpha/f_0 \\ y_\alpha/f_0 \\ 1 \end{pmatrix}, \tag{1}$$

we have the following relationship:

$$\boldsymbol{x}_\alpha = Z[\boldsymbol{H}\bar{\boldsymbol{x}}_\alpha]. \tag{2}$$

Here, $Z[\,\cdot\,]$ denotes normalization to make the third component 1, and $\boldsymbol{H}$ is the matrix in the following form [12]:

$$\boldsymbol{H} = \mathrm{diag}(1, 1, \frac{f_0}{f})\boldsymbol{R}^\top \left(\boldsymbol{I} - \frac{\boldsymbol{t}\boldsymbol{k}^\top}{d} \right). \tag{3}$$

Throughout this paper, $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ denote $(1,0,0)^\top$, $(0,1,0)^\top$, and $(0,0,1)^\top$, respectively, and $\mathrm{diag}(\cdots)$ denotes the diagonal matrix with diagonal elements $\cdots$.

3 Optimal Computation

Eq. (2) defines an image transformation called *homography*. Since the unknown parameters are $\{\boldsymbol{t}, \boldsymbol{R}\}$ and f, the homography has seven degrees of freedom. If the homography is unconstrained with eight degrees of freedom, we can apply our statistically optimal renormalization-based algorithm [15]; its C++ code is available via the Web[3]. Here, however, the homography is constrained. So, we take the bundle-adjustment approach based on Newton iterations.

Let $V[\boldsymbol{x}_\alpha]$ be the covariance matrix of the data vector $\boldsymbol{x}_\alpha$. We assume that it is known only up to scale and write

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha]. \tag{4}$$

We call the unknown magnitude ϵ the *noise level* and the matrix $V_0[\boldsymbol{x}_\alpha]$ the *normalized covariance matrix*. Since the third component of $\boldsymbol{x}$ is 1, $V_0[\boldsymbol{x}_\alpha]$ is a singular matrix of rank 2 with zeros in the third row and the third column. If the noise has no particular dependence on position and orientation, it has the form $\mathrm{diag}(1, 1, 0)$, which we use as the default value.

If the noise is Gaussian, an optimal estimate of $\boldsymbol{H}$ is obtained by *maximum likelihood estimation* [12]: we minimize the average squared *Mahalanobis distance*

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - Z[\boldsymbol{H}\bar{\boldsymbol{x}}_\alpha], V_0[\boldsymbol{x}_\alpha]^-(\boldsymbol{x}_\alpha - Z[\boldsymbol{H}\bar{\boldsymbol{x}}_\alpha])), \tag{5}$$

where and throughout this paper the operation $(\,\cdot\,)^-$ denotes the (Moore-Penrose) generalized inverse and $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. We define the following non-dimensional variables:

$$\phi = \frac{f}{f_0}, \qquad \boldsymbol{\tau} = \frac{\boldsymbol{t}}{d}. \tag{6}$$

The first order perturbation of $\boldsymbol{R}$ is written as $\boldsymbol{R} \to \boldsymbol{R} + \Delta\boldsymbol{\Omega} \times \boldsymbol{R}$, where $\Delta\boldsymbol{\Omega}$ is a 3-D vector and $\Delta\boldsymbol{\Omega} \times \boldsymbol{R}$ is a matrix whose columns are the vector products of $\Delta\boldsymbol{\Omega}$ and each columns of $\boldsymbol{R}$ [12]. We define the gradient ∇J and the Hessian $\nabla^2 J$ with respect to $\{\phi, \boldsymbol{\tau}, \boldsymbol{R}\}$ in such a way that the Taylor expansion of J has the form

$$\begin{aligned} &J(\phi + \Delta\phi, \boldsymbol{\tau} + \Delta\boldsymbol{\tau}, \boldsymbol{R} + \Delta\boldsymbol{\Omega} \times \boldsymbol{R}) \\ &= J(\phi, \boldsymbol{\tau}, \boldsymbol{R}) + (\nabla J, \begin{pmatrix} \Delta\phi \\ \Delta\boldsymbol{\tau} \\ \Delta\boldsymbol{\Omega} \end{pmatrix}) + \frac{1}{2}(\begin{pmatrix} \Delta\phi \\ \Delta\boldsymbol{\tau} \\ \Delta\boldsymbol{\Omega} \end{pmatrix}, \nabla^2 J \begin{pmatrix} \Delta\phi \\ \Delta\boldsymbol{\tau} \\ \Delta\boldsymbol{\Omega} \end{pmatrix}) + \cdots. \end{aligned} \tag{7}$$

The solution that minimizes J is obtained by the following Newton iterations:

1. Give an initial guess of ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$.
2. Compute the gradient ∇J and the Hessian $\nabla^2 J$ (their actual expressions are omitted).

[3] `http://www.ail.cs.gunma-u.ac.jp/kanatani/e`

3. Compute $\Delta\phi$, $\Delta\boldsymbol{\tau}$, and $\Delta\boldsymbol{\Omega}$ by solving the linear equation

$$\left(\nabla^2 J\right)\begin{pmatrix}\Delta\phi\\ \Delta\boldsymbol{\tau}\\ \Delta\boldsymbol{\Omega}\end{pmatrix} = -\nabla J. \tag{8}$$

4. If $|\Delta\phi| < \epsilon_\phi$, $\|\Delta\boldsymbol{\tau}\| < \epsilon_\tau$, and $\|\Delta\boldsymbol{\Omega}\| < \epsilon_{\mathbf{R}}$, return ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ and stop. Otherwise, update ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ in the form

$$\phi \leftarrow \phi + \Delta\phi, \quad \boldsymbol{\tau} \leftarrow \boldsymbol{\tau} + \Delta\boldsymbol{\tau}, \quad \boldsymbol{R} \leftarrow \mathcal{R}(\Delta\boldsymbol{\Omega})\boldsymbol{R}, \tag{9}$$

and go back to Step 2.

The symbol $\mathcal{R}(\Delta\boldsymbol{\Omega})$ denotes the rotation of angle $\|\Delta\boldsymbol{\Omega}\|$ around $\Delta\boldsymbol{\Omega}$; ϵ_ϕ, ϵ_τ, and $\epsilon_{\mathbf{R}}$ are thresholds for convergence.

The initial guess of ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ can be obtained by computing the homography $\boldsymbol{H}$ between $\{\bar{\boldsymbol{x}}_\alpha\}$ and $\{\boldsymbol{x}_\alpha\}$, say, by least squares or by the renormalization-based method [15] without considering the constraint and approximately decomposing it into ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ in the form of eq. (3) (an analytical procedure for this is given in [20]). However, this procedure is necessary only for the initial frame. For the subsequent frames, we can start from the solution in the preceding frame or an appropriate prediction from it, as we will describe shortly.

4 Reliability Evaluation

The squared noise level ϵ^2 can be estimated from the residual $\hat{J}$ (the minimum value of J) in the following form [12]:

$$\hat{\epsilon}^2 = \frac{\hat{J}}{2 - 7/N}. \tag{10}$$

Let $\nabla^2\hat{J}$ be the resulting Hessian. The covariance matrix of $\{\hat{\phi},\ \hat{\boldsymbol{\tau}},\ \hat{\boldsymbol{R}}\}$ is estimated in the following form:

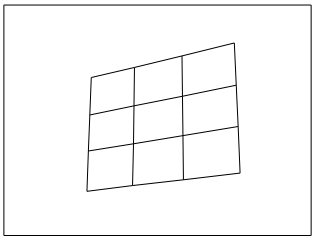
$$V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}] = \frac{2\hat{\epsilon}^2}{N}\left(\nabla^2\hat{J}\right)^{-1}. \tag{11}$$

This gives an estimate of the *Cramer-Rao lower bound* (*CRLB*) on $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ [12].

The (1,1) element of $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ gives the variance $V[\hat{\phi}]$ of ϕ. It follows that if the error distribution is approximated to be Gaussian, the 99.7% confidence interval of f has the form

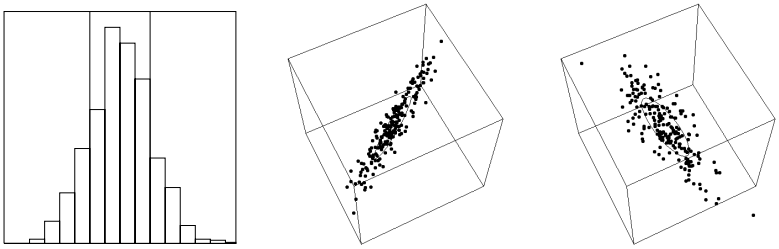
$$\hat{\phi} - 3\sqrt{V[\hat{\phi}]} < \frac{f}{f_0} < \hat{\phi} + 3\sqrt{V[\hat{\phi}]}. \tag{12}$$

The submatrix of $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ defined by its second to fourth rows and columns gives the covariance matrix $V[\hat{\boldsymbol{\tau}}]$ of $\boldsymbol{\tau}$. Let $\Delta\Omega$ and $\boldsymbol{l}$ be, respectively, the angle and axis of the rotation $\hat{\boldsymbol{R}}\bar{\boldsymbol{R}}^\top$ relative to the true rotation $\bar{\boldsymbol{R}}$. Let $\Delta\boldsymbol{\Omega} = \Delta\Omega\boldsymbol{l}$. The submatrix of $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ defined by its fifth to seventh rows and columns gives the covariance matrix $V[\hat{\boldsymbol{R}}]$ of $\Delta\boldsymbol{\Omega}$.



	empirical	CRLB
focal length (pixles)	33.4	34.0
translation (cm)	32.9	32.6
rotation (deg)	0.413	0.414

Fig. 2. Simulated image of a grid pattern (left); the standard deviations of the optimally computed solutions and estimates of their Cramer-Rao lower bounds (right).



(a) (b) (c)

Fig. 3. (a) Histogram of the computed focal length. (b) Error distribution of the computed translation. (c) Error distribution of the computed rotation.

5 Examples of Reliability Evaluation

5.1 Numerical simulation

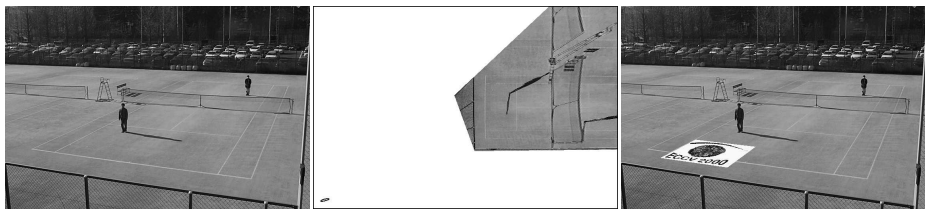
Fig. 2 shows a simulated image of a grid pattern viewed from an angle. We added Gaussian random noise of mean 0 and standard deviation 1 (pixel) to the x and y coordinates of the vertices independently and computed the focal length and the motion parameters 1,000 times, using different noise each time. The standard deviations of the computed solutions and estimates of their CRLBs are listed in Fig. 2.

Fig. 3(a) is the histogram of the computed focal length $\hat{f}$. The vertical lines indicate the estimated CRLB. Fig. 3(b) is a 3-D plot of the distribution of the error vector $\Delta \boldsymbol{t} = \hat{\boldsymbol{t}} - \bar{\boldsymbol{t}}$ of translation. The ellipse indicates the estimated CRLB in each orientation. Fig. 3(c) is a 3-D plot of the error vector $\Delta \boldsymbol{\Omega}$ of rotation depicted similarly.

From these results, we can confirm that the estimated CRLB can be used as a reliability measure of the solution.

5.2 Tennis court scene

Fig. 4(a) is a real image of a tennis court. Since the size of the court is stipulated by an international rule, we can compute the 3-D camera position and the focal length by using this knowledge. The focal length is estimated to be 955 pixels.



(a) (b) (c)

Fig. 4. (a) A real image of a tennis court. (b) The computed camera position viewed from above. (c) A virtual scene generated from (a).

The camera is estimated to be at 627cm above the ground. The standard deviations of the focal length, the translation, and the rotation are evaluated to be 6.99 pixels, 16.14cm, and 0.151 deg, respectively.

Fig. 4(b) shows the top view of the tennis court generated from Fig. 4(a). The estimated camera position is plotted there and encircled by an ellipse, which indicates three times the standard deviation of the estimated position in each orientation (actually it is an ellipsoid viewed from above).

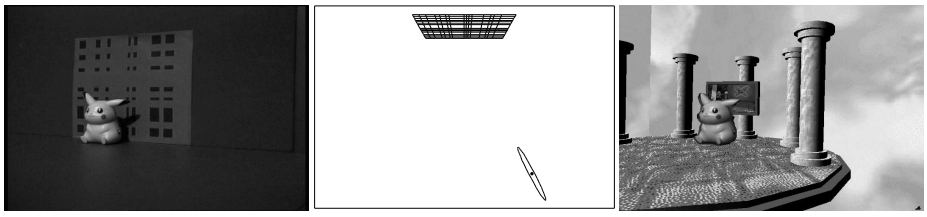
The images of the poles and the persons in Fig. 4(b) can be regarded as their "shadows" on the ground cast by hypothetical light emitted from the camera, so we can compute their heights [5, 10]. The right pole is estimated to be 113cm in height. The person near the camera is estimated to be 171cm tall. This technique can be applied to 3-D analysis of sports broadcasting [25, 28]. Since we know the 3-D structure of the scene, we can generate a virtual view of a new object placed in the scene. Fig. 4(c) is a virtual view of a logo placed on the tennis court.

5.3 Virtual studio

Fig. 5(a) is a real image of a toy, behind which is placed a grid pattern colored light and dark blue. The grid pattern is placed on the floor perpendicularly. The camera optical axis is almost parallel to the floor. Unoccluded grid points in the image were matched to their true positions in the pattern by observing the cross ratio of adjacent points. This pattern is so designed that the cross ratio is different everywhere in such a way that matching can be done in a statistically optimal way in the presence of image noise [17, 19].

After separating the toy image from the background by using a chromakey technique, we computed the 3-D position and focal length of the camera by observing an unoccluded portion of the grid pattern (see [19] for the image processing details). The focal length is estimated to be 576 pixels. The standard deviations of the focal length, the translation, and the rotation are evaluated to be 38.3 pixels, 5.73cm, and 0.812 deg, respectively.

Fig. 5(b) is the top view of the estimated camera position and its uncertainty ellipsoid (three times the standard deviation in each orientation). Fig. 5(c) is a composition of the toy image and a graphics scene generated by VRML.



(a) (b) (c)

Fig. 5. (a) Original image. (b) Estimated camera position and its reliability. (c) A virtual scene generated from (a).

6 Trajectory Stabilization

If the camera optical axis is perpendicular to the planar pattern, the Hessian $\nabla^2 J$ in eq. (8) is a singular matrix, so the solution is indeterminate. This does not occur in practice due to image noise, but the resulting solution is numerically unstable. Also, as pointed out in Introduction, the computed camera position fluctuates when the camera motion is small. We now present a technique for avoiding degeneracy and statistical fluctuations by *model selection*.

6.1 Model selection criteria

The homography $\boldsymbol{H}$ given by eq. (3) is parameterized by $\{\boldsymbol{t}, \boldsymbol{R}\}$ and f, having seven degrees of freedom. If the motion and zooming of the camera are constrained in some way (e.g., the camera is translated without rotation or zooming), the homography $\boldsymbol{H}$ has a smaller degree of freedom, and a smaller number of parameters need to be estimated. In general, parameter estimation becomes stabler as the number of parameters decreases.

It follows that we can stably estimate the parameters or avoid degeneracy if we know the constraint on the camera motion or zooming [6, 9, 28]. In practice, however, we do not know how the camera is moving or zooming. Our strategy here is to assume probable constraints (translation only, etc.), which we call *models*, compare each other, and adopt the best one. A naive idea for this is to compute the residual $\hat{J}$ for each model and choose the one for which it is minimum. However, this does not work: the general model always has the smallest residual, since the residual decreases as the degree of freedom increases.

The best known criterion for balancing the residual and the degree of the freedom of the model is Akaike's *AIC* [2] designed for statistical estimation and used in some vision applications [4]. Kanatani's *geometric AIC* [12, 14] is a variant of Akaike's AIC specifically designed for geometric estimation and has been applied to a variety of vision applications [13, 16, 23, 31, 32, 33, 35]. In the present case, the geometric AIC for minimizing eq. (5) is written as

$$\text{G-AIC} = \hat{J} + 2k\epsilon^2, \tag{13}$$

where k is the degree of freedom of the homography $\boldsymbol{H}$. The square noise level ϵ^2 is estimated from the general model in the form of eq. (10).

Another well known criterion is Rissanen's *MDL* (*minimum description length*) based on the information theoretic code length of the model [26, 27]. It is derived by analyzing the function space of "stochastic models" identified with parameterized probability densities in the asymptotic limit of a large number of observations. Here, the models we want to compare are *geometric constraints*, not parameterized probability densities. Also, we are given only *one* set of data (i.e., one observation) for each frame. Hence, Rissanen's MDL cannot be used in its original form.

The starting point of Rissanen's MDL is the observation that encoding a real number requires an infinite code length. Rissanen's idea is to quantize the parameters to obtain a finite code length, taking into account the fact that real numbers cannot be estimated completely [27]. The quantization width is determined by attainable estimation accuracy, which in turn is determined by the data length n. Since the code length diverges as $n \to \infty$, asymptotic approximation comes into play. In this sense, the "minimum description length" actually means the "minimum growth rate" of the description length.

Suppose we hypothetically repeat independent observations, although the actual observation is done only once. The accuracy of estimation increases as the number of hypothetical observations, so we can define the MDL by asymptotic analysis. But increasing the number n of observations effectively reduces the noise level ϵ to $O(1/\sqrt{n})$. It follows that we can define the MDL as the "growth rate" of the description length as $\epsilon \to 0$. The final form is as follows (we omit the details of the code length analysis):

$$\text{G-MDL} = \hat{J} - k\epsilon^2 \log \epsilon^2. \tag{14}$$

We call this criterion the *geometric MDL*[4]. This form can also be obtained from Rissanen's MDL by replacing n by $1/\epsilon^2$ and is different from any MDLs used in statistics and vision applications [8, 11, 21, 22, 31] in that ours does *not* contain the logarithm of the number of the data.

6.2 Degeneracy detection

If degeneracy occurs, the confidence interval (12) expands infinitely wide if no noise exist. In the presence of noise, it has a finite width. We decide that degeneracy has occurred if the confidence interval (12) contains negative values of f. This means that we adopt the following criterion:

$$V[\hat{\phi}] > \frac{\hat{\phi}^2}{9}. \tag{15}$$

The variance $V[\hat{\phi}]$ equals the (1,1) element of the covariance matrix $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ given by eq. (11), so it is equal to $2\hat{\epsilon}^2(\nabla^2\hat{J})^{\dagger}_{11}/N \det(\nabla^2\hat{J})$, where $(\nabla^2\hat{J})^{\dagger}_{11}$ is the

[4] Since the additive terms can be ignored when $\epsilon \ll 1$, changing the unit of length does not affect the relative comparison of models asympotitically.

(1,1)-cofactor of the Hessian $\nabla^2 \hat{J}$ (the determinant of the submatrix obtained by removing the first row and the first column of $\nabla^2 \hat{J}$). Hence, eq. (15) can be rewritten in the form

$$\frac{18\hat{\epsilon}^2}{N}\left(\nabla^2 \hat{J}\right)^{\dagger}_{11} - \hat{\phi}^2 \det\left(\nabla^2 \hat{J}\right) > 0. \tag{16}$$

Since matrix inversion is no longer involved, this expression can always be stably evaluated.

6.3 Models of zooming and motion of the camera

We predict the focal length f and the motion parameters $\{\boldsymbol{t},\ \boldsymbol{R}\}$ in the next frame from the values f_i and $\{\boldsymbol{t}_i,\ \boldsymbol{R}_i\}$ of the current frame and the values f_{i-1} and $\{\boldsymbol{t}_{i-1},\ \boldsymbol{R}_{i-1}\}$ of the preceding frame. Here, we consider the following six models:

Stationary model: We assume that the camera is stationary: $f = f_i$, $\boldsymbol{t} = \boldsymbol{t}_i$, and $\boldsymbol{R} = \boldsymbol{R}_i$. Let $\hat{J}_*$ be the corresponding residual. This model has zero degrees of freedom.

***t*-fixed model:** We assume that the camera only rotates. We let $f = f_i$ and $\boldsymbol{t} = \boldsymbol{t}_i$ and optimally compute the rotation $\boldsymbol{R}$ by Newton iterations starting from $\boldsymbol{R}_i$. Let $\hat{J}_{s'}$ be the corresponding residual. This model has three degrees of freedom.

***t*-predicted model:** Assuming that the zooming does not change, we linearly extrapolate the camera position and let $\boldsymbol{t} = 2\boldsymbol{t}_i - \boldsymbol{t}_{i-1}$. Then, we optimally compute the rotation $\boldsymbol{R}$ by Newton iterations starting from $\boldsymbol{R}_i\boldsymbol{R}_{i-1}^{\top}\boldsymbol{R}_i$. Let $\hat{J}_{p'}$ be the corresponding residual. This model has three degrees of freedom.

***f*-fixed model:** Assuming that the zooming does not change, we optimally compute the motion parameters $\{\boldsymbol{t},\ \boldsymbol{R}\}$ by Newton iterations starting from $\{\boldsymbol{t}_i,\ \boldsymbol{R}_i\}$. Let $\hat{J}_s$ be the corresponding residual. This model has six degrees of freedom. The square noise level ϵ^2 is estimated by

$$\hat{\epsilon}_s^2 = \frac{\hat{J}_s}{2 - 6/N}. \tag{17}$$

***f*-predicted model:** We linearly extrapolate the focal length and let $f = 2f_i - f_{i-1}$. Then, we optimally compute the motion parameters $\{\boldsymbol{t},\ \boldsymbol{R}\}$ by Newton iterations starting from $\{2\boldsymbol{t}_i - \boldsymbol{t}_{i-1},\ \boldsymbol{R}_i\boldsymbol{R}_{i-1}^{\top}\boldsymbol{R}_i\}$. Let $\hat{J}_p$ be the corresponding residual. This model has six degrees of freedom. The square noise level ϵ^2 is estimated by

$$\hat{\epsilon}_p^2 = \frac{\hat{J}_p}{2 - 6/N}. \tag{18}$$

General model: We optimally compute the focal length f and the motion parameters $\{\boldsymbol{t},\ \boldsymbol{R}\}$ by Newton iterations starting from the solution obtained from the f-predicted model. Let $\hat{J}_g$ be the corresponding residual. This model has seven degrees of freedom.

Degeneracy is detected from the f-predicted model. Namely, we estimate the square noise level ϵ^2 by eq. (18) and evaluate the criterion (16). If degeneracy is not detected, we compare the stationary model, the f-fixed model, the f-predicted model, and the general model. Estimating the square noise level ϵ^2 by eq. (10), we evaluate the geometric AICs and the geometric MDLs of these models in the following form:

$$\text{G-AIC}_* = \hat{J}_*, \quad \text{G-AIC}_s = \hat{J}_s + \frac{12}{N}\hat{\epsilon}^2, \quad \text{G-AIC}_p = \hat{J}_p + \frac{12}{N}\hat{\epsilon}^2,$$

$$\text{G-AIC}_g = \hat{J}_g + \frac{14}{N}\hat{\epsilon}^2, \quad \text{G-MDL}_* = \hat{J}_*, \quad \text{G-MDL}_s = \hat{J}_s - \frac{6}{N}\hat{\epsilon}^2 \log \hat{\epsilon}^2,$$

$$\text{G-MDL}_p = \hat{J}_p - \frac{6}{N}\hat{\epsilon}^2 \log \hat{\epsilon}^2, \quad \text{G-MDL}_g = \hat{J}_g - \frac{7}{N}\hat{\epsilon}^2 \log \hat{\epsilon}^2. \tag{19}$$

The model that gives the smallest AIC or the smallest MDL is chosen.

If degeneracy is detected, we compare the stationary model, the $\boldsymbol{t}$-fixed model, the $\boldsymbol{t}$-predicted model, and the f-fixed model. Estimating the square noise level ϵ^2 by eq. (17), we evaluate the geometric AICs and the geometric MDLs of these models in the following form:

$$\text{G-AIC}_* = \hat{J}_*, \quad \text{G-AIC}_{s'} = \hat{J}_{s'} + \frac{6}{N}\hat{\epsilon}_s^2, \quad \text{G-AIC}_{p'} = \hat{J}_{p'} + \frac{6}{N}\hat{\epsilon}_s^2,$$

$$\text{G-AIC}_s = \hat{J}_s + \frac{12}{N}\hat{\epsilon}_s^2, \quad \text{G-MDL}_* = \hat{J}_*, \quad \text{G-MDL}_{s'} = \hat{J}_{s'} - \frac{3}{N}\hat{\epsilon}_s^2 \log \hat{\epsilon}_s^2,$$

$$\text{G-MDL}_{p'} = \hat{J}_{p'} - \frac{3}{N}\hat{\epsilon}_s^2 \log \hat{\epsilon}_s^2, \quad \text{G-MDL}_s = \hat{J}_s - \frac{6}{N}\hat{\epsilon}_s^2 \log \hat{\epsilon}_s^2. \tag{20}$$

The model that gives the smallest AIC or the smallest MDL is chosen.

7 Model Selection Examples

7.1 Numerical simulation

We simulate a camera motion in a plane perpendicular to a 3×3 grid pattern. In the course of its motion, the camera is rotated so that the center of the pattern is always fixed at the center of the image frame. First, the camera moves along a circular trajectory as shown in Fig. 8(a). It perpendicularly faces the pattern at frame 13 and stops at frame 20. The camera stays there for five frames (frames $20 \sim 24$) and then recedes backward for another five frames (frames $25 \sim 30$).

Adding random Gaussian noise of mean 0 and standard deviation 1 (pixel) to each coordinate of the grid points independently at each frame, we compute the focal length and the trajectory of the camera (Figs. 6(b) and 6(c)). Degeneracy is detected at frames 12 and 13. In order to emphasize the fact that the frame-wise estimation fails, we let f be ∞ and the camera position be at the center of the grid pattern in Figs. 6(b) and 6(c) when degeneracy is detected.

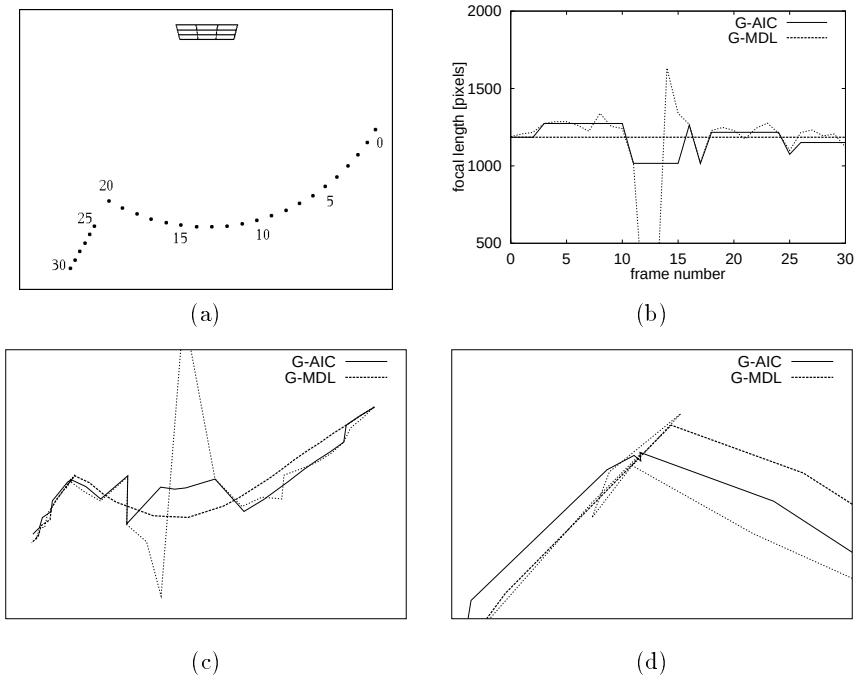


Fig. 6. (a) Simulated camera motion. (b) Estimated focal lengths. (c) Estimated camera trajectory. (d) Magnification of the portion of (c) for frames $20 \sim 24$. In (b)$\sim$(d), the solid lines indicate model selection by the geometric AIC; the thick dashed lines indicate model selection by the geometric MDL; the thin dotted lines indicate frame-wise estimation.

As we can see, both the geometric AIC and the geometric MDL produce a smoother trajectory than frame-wise estimation and that the computed trajectory smoothly passes through the degenerate configuration. Fig. 6(d) is a magnification of the portion for frames $20 \sim 24$ in Fig. 6(c). We can observe that statistical fluctuations exist if the camera position is estimated at each frame independently and that the fluctuations are removed by model selection.

From these results, it is clearly seen that the geometric MDL has a stronger smoothing effect than the geometric AIC. This is because the penalty $-\epsilon^2 \log \epsilon^2$ for each degree of freedom in the geometric MDL is generally larger than the penalty $2\epsilon^2$ in the geometric AIC (see eq. (13) and eq. (14)) so the geometric MDL tends to select a simpler model than the geometric AIC.

7.2 Virtual studio

Fig. 7 shows five sampled frames from a real image sequence obtained in the setting described in Section 5.3. The camera moves from right to left with a fixed focal length. The camera optical axis becomes almost perpendicular to the

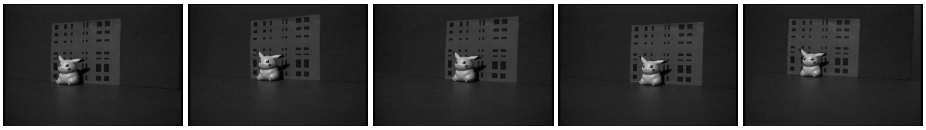
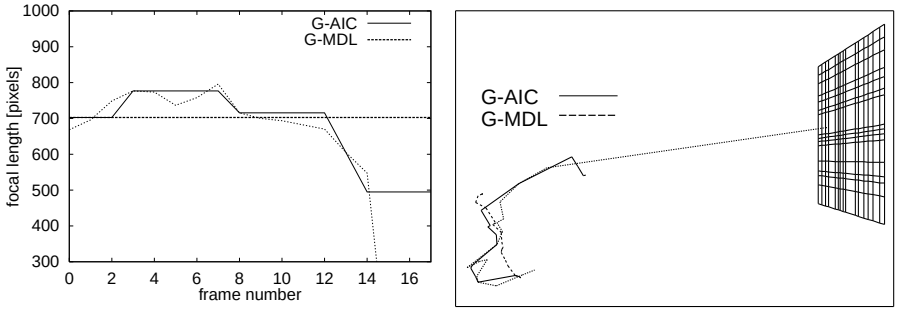


Fig. 7. Sampled frames from a real image sequence.



(a) (b)

Fig. 8. (a) Estimated focal lengths. (b) Estimated camera trajectory. In (a) and (b), the solid lines indicate model selection by the geometric AIC; the thick dashed lines indicate model selection by the geometric MDL; the thin dotted lines indicate frame-wise estimation.

grid pattern in the 15th frame. Degeneracy is detected there and thereafter.

Fig. 8(a) shows the estimated focal lengths; Fig. 8(b) shows the estimated camera trajectory viewed from above. The frame-wise estimation fails when degeneracy occurs. In this case, the estimation by the geometric MDL is more consistent with the actual camera motion than the geometric AIC. But this is because we fixed the zooming and moved the camera smoothly. If we added variations to the zooming and the camera motion, the geometric MDL would still prefer a smooth motion. So, we cannot say which solution should be closer to the true solution; it depends on what kind of solution *we expect* is desirable for the application in question.

8 Concluding Remarks

Motivated by virtual studio applications, we have studied the technique for "simultaneous calibration" for computing the 3-D position and focal length of a continuously moving and continuously zooming camera from an image of a planar pattern placed behind the object. We have described a procedure for computing an optimal solution that provides an evaluation of the reliability of the solution.

Then, we showed that degeneracy of the solution and statistical fluctuations of computation can be avoided by model selection: we predict the 3-D position and focal length of the camera in multiple ways and select the best model using

the geometric AIC and the geometric MDL. Doing numerical and real-image experiments, we have observed that the geometric MDL tends to select a simpler model than the geometric AIC, thereby producing a smoother and more cohesive estimation.

References

1. J. Batista, H. Araújo and A. T. de Almeida, Iterative multistep explicit camera calibration, *IEEE Trans. Robotics Automation*, **15**-5 (1999), 897–917.
2. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Automation Control*, **19**-6 (1974), 716–723.
3. S. Bougnoux, From projective to Euclidean space under any practical situation, a criticism of self calibration, *Proc. 6th Int. Conf. Comput. Vision.*, January 1998, Bombay, India, pp. 790–796.
4. K. L. Boyer, M. J. Mirza and G. Ganguly, The robust sequential estimator: A general approach and its application to surface organization in range data, *IEEE Trans. Patt. Anal. Mach. Intell.*, **16**-10 (1994), 987–1001.
5. A. Criminisi, I. Reid and A. Zisserman, Duality, rigidity and planar parallax, *Proc. 5th Euro. Conf. Comput. Vision.*, June 1998, Freiburg, Germany, pp. 846–861.
6. L. de Agapito, R. I. Hartley and E. Hayman, Linear self-calibration of a rotating and zooming camera, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, June 1999, Fort Collins, CO, U.S.A., pp. 15–21.
7. S. Gibbs, C. Arapis, C. Breiteneder, V. Lalioti, S. Mostafawy and J. Speier, Virtual studios: An overview. *IEEE Multimedia*, **5**-1 (1998), 24–35.
8. H. Gu, Y. Shirai and M. Asada, MDL-based segmentation and motion modeling in a long image sequence of scene with multiple independently moving objects, *IEEE Trans. Patt. Anal. Mach. Intell.*, **18**-1 (1996), 58–64.
9. R. I. Hartley, Self-calibration of stationary cameras, *Int. J. Comput. Vision*, **22**-1 (1997), 5–23.
10. M. Irani, P. Anandan and D. Weinshall, From reference frames to reference planes: Multi-view parallax geometry and applications, *Proc. 5th Euro. Conf. Comput. Vision.*, June 1998, Freiburg, Germany, pp. 829–845.
11. Y. G. Leclerc, Constructing simple stable descriptions for image partitioning, *Int. J. Comput. Vision*, **3**-1 (1989), 73–102.
12. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, The Netherlands, 1996.
13. K. Kanatani, Self-evaluation for active vision by the geometric information criterion, *Proc.7th Int. Conf. Computer Analysis of Images and Patterns (CAIP'97)*, September 1997, Kiel, Germany, pp. 247–254.
14. K. Kanatani, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.
15. K. Kanatani and N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, *Proc. 7th Int. Conf. Comput. Vision*, September, 1999, Kerkya, Greece, pp. 73–78.
16. Y. Kanazawa and K. Kanatani, Infinity and planarity test for stereo vision, *IEICE Trans. Inf. & Syst.* **E80-D**-8 (1997), 774–779.
17. Y. Kanazawa, C. Matsunaga and K. Kanatani, Best marker pattern design for recognition by cross ratio (in Japanese), *IPSJ SIG Notes*, 99-CVIM-115-13, March 1999, pp. 97–104.
18. D. Liebowitz and A. Zisserman, Metric rectification for perspective images of planes, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, June 1998, Santa Barbara, CA, U.S.A., pp. 482–488.

19. C. Matsunaga, K. Niijima and K. Kanatani, Best marker pattern design for recognition by cross ratio: Experimental investigation (in Japanese), *IPSJ SIG Notes*, 99-CVIM-115-14, March 1999, pp. 105–110.
20. C. Matsunaga and K. Kanatani, Stabilizing moving camera calibration from images by the geometric AIC, *Proc. 4th Asian Conf. Comput. Vision*, January 2000, Taipei, Taiwan, pp. 1168–1173.
21. B. A. Maxwell, Segmentation and interpretation of multicolored objects with highlights, *Comput. Vision Image Understand.*, **77**-1 (2000), 1–24.
22. S. J. Maybank and P. F. Sturm, MDL, collineations and the fundamental matrix, *Proc. 10th British Machine Vision Conference*, September 1999, Nottingham, U.K., pp. 53–62.
23. N. Ohta and K. Kanatani, Moving object detection from optical flow without empirical thresholds, *IEICE Trans. Inf. & Syst.*, **E81-D**-2 (1998), 243–245.
24. M. Pollefeys, R. Koch and L. Van Gool, Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, *Int. J. Comput. Vision*, **32**-1 (1999), 7–26.
25. I. Reid and A. Zisserman, Goal-directed video metrology, *Proc. 4th Euro. Conf. Comput. Vision*, April 1996, Cambridge, U.K., Vol. II, pp. 647–658.
26. J. Rissanen, Modeling by shortest data description, *Automatica*, **14** (1978), 465–471.
27. J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.
28. Y. Seo and K. S. Hong, About the self-calibration of a rotating and zooming camera: Theory and practice, *Proc. 7th Int. Conf. Comput. Vision*, September 1999, Kerkyra, Greece, pp. 183–189.
29. P. Sturm and S. Maybank, On plane-based camera calibration: A general algorithm, singularities, applications, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, June 1999, Fort Collins, CO, U.S.A., pp. 432–437.
30. M. Tamir, The Orad virtual set, *Int. Broadcast Eng.*, March (1996), 16–18.
31. P. H. S. Torr, An assessment for information criteria for motion model selection, *Proc. IEEE Conf. Comput. Vision Patt. Recog.*, June 1997, Puerto Ricco, pp. 47–52.
32. P. H. S. Torr, A. W. Fitzgibbon and A. Zisserman, Maintaining multiple motion model hypotheses over many views to recover matching and structure, *Proc. 6th Int. Conf. Comput. Vision*, January 1998, Bombay, India, pp. 485–491.
33. P. H. S. Torr, Geometric motion segmentation and model selection, *Phil. Trans. Roy. Soc.*, A-**356**-1740 (1998), 1321–1340.
34. B. Triggs, Autocalibration from planar scenes, *Proc. 5th Euro. Conf. Comput. Vision*, June 1998, Freiburg, Germany, pp. 89–105.
35. Iman Triono, N. Ohta and K. Kanatani, Automatic recognition of regular figures by geometric AIC, *IEICE Trans. Inf. & Syst.*, **E81-D**-2 (1998), 246–248.
36. R. Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision methodology using off-the-shelf TV cameras and lenses, *IEEE J. Robotics Automation*, **3**-4 (1987), 323–344.
37. Z. Zhang, Flexible camera calibration by viewing a plane from unknown orientations, *Proc. 7th Int. Conf. Comput. Vision*, September, 1999, Kerkya, Greece, pp. 666–673.

This article was processed using the LaTeX macro package with LLNCS style