Stabilizing Moving Camera Calibration from Images by the Geometric AIC

Chikara Matsunaga
Broadcast Division, FOR-A Co. Ltd., 2–3–3 Ohsaku, Sakura, Chiba 285-0802 Japan
Kenichi Kanatani
Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515 Japan

ABSTRACT

We study the scheme for "simultaneous calibration" for computing the focal length and the position of a continuously moving and continuously zooming camera from an image of a planar pattern placed behind an object. We describe a procedure for computing an optimal solution that attains the theoretical accuracy bound in the presence of nosie. This procedure also evaluates the reliability of the solution in the course of its computation. We also show that degeneracy of the solution and statistical fluctuations of computation can be avoided by "model selection". We confirm the effectiveness of our method by simulations and show a real image application.

Keywords: mixed reality, camera calibration, homography, reliability evaluation, model selection, geometric AIC.

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*[8], 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 is *pre-calibration*: we estimate the internal parameters by taking images of objects or patterns of known 3-D shape in a controlled environment in advance. The best known procedure for this is the method of Tsai [12]. Recently, techniques for computing both the camera parameters and the 3-D positions of the camera by observing multiple images of the same scene about which we have no prior knowledge have intensively been studied [9, 13]. Such a technique, known as *self-calibration*, may be very useful in unknown environments such as outdoors. For stable reconstruction, however, it requires a long sequence of images of the same scene taken from different and unconstrained camera positions and feature matching among them. As a result, the amount of computation required is too large for real-time applications. Also, it cannot be applied if the camera motion is constrained or if the scene changes as the camera moves.

In this paper, we focus on *virtual studio* applications [1]: 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 the internal parameters of a moving camera. Since the camera parameters, such as zooming, may change from frame to frame, the calibration must be done for each frame. Hence, we cannot pre-calibrate or self-calibrate the camera. This difficulty can be avoided by placing an easily distinguishable planar pattern with a known geometry in the scene (Fig. 1): we detect the image of the pattern in each frame, compute the camera position and the internal parameters from it, and removing the image of the pattern by image segmentation. We call this *simultaneous calibration*.

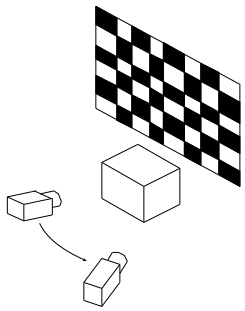


Figure 1: Simultaneous calibration of a moving camera.

This type of calibration has many elements that do not appear in pre-calibration:

1. While manual interventions can be employed in pre-calibration, simultaneous calibration must be automated. In particular, we must automatically find the 3-D positions of the marker points that are visible in each frame.
2. Since the number of marker points that are visible in each frame 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, a degenerate configuration can occur: when the camera optical axis is perpendicular to the pattern, the camera position and the focal length are indeterminate because zooming out and moving the camera forward cause the same visual effect.
4. Since the visible marker points change as the object moves in the scene, the computed camera position may be slightly different from frame to frame due to noise even if the camera is stationary in the scene. This type of statistical fluctuations is conspicuous when the camera motion is small.

To resolve the first issue, we designed a special marker pattern in such a way that it can be detected with theoretically maximum robustness [5, 6].

In order to deal with the second issue, we present a technique for computing a statistically optimal solution and evaluating its reliability at the same time.

As to the third and fourth issues, we show that we can avoid both the degeneracy and the statistical fluctuations by incorporating *model selection*: at each frame, we predict the zooming and the position of the camera from their history and compare the goodness of that prediction, or *model*, with the goodness of the computed solution by means of the *geometric AIC* [3]. We can stabilize the computed solution by adopting the predicted model unless the evidence to the contrary is sufficiently strong. We confirm the effectiveness of our procedure by simulations and real-image experiments.

2. BASIC PRINCIPLE

We fix an XYZ world coordinate system and place a planar pattern in parallel to the XY plane at a known distance d. We take the focal length f as the only unknown camera parameter, assuming that other parameters have already been pre-calibrated so that the imaging geometry can be modeled as a perspective projection.

Suppose we place a hypothetical camera with a known focal length f_0 in such a way that the center of projection is at the origin, the Z-axis coincides with the optical axis, and the world X- and Y-axes are parallel to the image x- and y-axes. Then, we rotate the camera by an unknown rotation matrix $\boldsymbol{R}$, translate it by an unknown vector $\boldsymbol{t}$, and change the focal length into an unknown value f. We call $\{\boldsymbol{t}, \boldsymbol{R}\}$ the *motion parameters*. 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, $\alpha = 1, ..., N$. 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 [3]:

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

Here, $Z[\,\cdot\,]$ denotes normalization to make the third component 1, and $\boldsymbol{H}$ is the matrix in the following form [3]:

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

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 $\cdots$. Eq. (2) defines an image transformation called *homography* [3]. A statistically optimal algorithm for computing it from feature correspondences was presented by Kanatani and Ohta [4], and its C++ code is publicly available via the Web[1] .

Since the unknowns $\{\boldsymbol{t}, \boldsymbol{R}\}$ and f have 7 degrees of freedom, we can determine them in theory if four or more points $\{(x_\alpha, y_\alpha)\}$ are observed.

3. OPTIMAL COMPUTATION

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

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

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

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 low and the third column. If the noise has no particular dependence on positions and orientations, we use the default value $\mathrm{diag}(1,1,0)$. An optimal estimate of $\boldsymbol{H}$ is obtained by minimizing the sum of the squared *Mahalanobis distances*

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

where $(\,\cdot\,)^-$ denotes the (Moore-Penrose) generalized inverse. We view J as a function of non-dimensional variables

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

The first order perturbations of ϕ and $\boldsymbol{\tau}$ are written, respectively, as $\phi \to \phi + \Delta\phi$ and $\boldsymbol{\tau} \to \boldsymbol{\tau} + \Delta\boldsymbol{\tau}$. 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 $\boldsymbol{R}$. 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

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

where $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. 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 $\nabla^2 J$.
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 3.

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

4. RELIABILITY EVALUATION

The squared noise level ϵ^2 can be estimated from the residual $\hat{J}$ in the following form [3]:

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

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 follwoing form [3]:

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

If the true values of $\{\phi,\ \boldsymbol{\tau},\ \boldsymbol{R}\}$ are substituted, this gives the *Cramer-Rao lower bound* on $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ [3].

The (1,1) element $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ gives the variance $V[\hat{\phi}]$ of ϕ. If the error distibution is approximated to be Gaussian, the 99.7% confidence interval of f is given by

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

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 $\bar{\boldsymbol{R}}$ be the true rotation. Let $\Delta\Omega$ and $\boldsymbol{l}$ be the angle and axis of the relative rotation $\hat{\boldsymbol{R}}\bar{\boldsymbol{R}}^\top$. If we define vector

$$\Delta\boldsymbol{\Omega} = \Delta\Omega \boldsymbol{l}, \tag{13}$$

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 $\Delta\boldsymbol{\Omega}$.

5. ANALYTICAL SOLUTION

The initial guess of ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ can be obtained by the following analytical procedure, which produces an exact solution in the absence of noise. First, we compute the homography matrix $\boldsymbol{H}$ by least squares as follows:

$$J_{LS} = \frac{1}{N}\sum_{\alpha=1}^{N} \|\boldsymbol{x}_\alpha \times \boldsymbol{H}\bar{\boldsymbol{x}}_\alpha\|^2 \to \min. \tag{14}$$

Since J_{LS} is a quadratic form in $\boldsymbol{H}$, the solution is immediately obtained up to scale by solving the eigenvalue problem. It follows that there exits a constant k such that

$$\boldsymbol{H} = k\boldsymbol{A}^{-1}\boldsymbol{R}^\top(\boldsymbol{I} - \boldsymbol{\tau}\boldsymbol{k}^\top), \tag{15}$$

where we define $\boldsymbol{A} = \mathrm{diag}(1, 1, \phi)$. Multiplying each side by $\boldsymbol{A}/k$ from left and then multiplying each side by its transpose, we obtain

$$\frac{1}{k^2}\boldsymbol{H}^\top\boldsymbol{A}^2\boldsymbol{H} = (\boldsymbol{I} - \boldsymbol{k}\boldsymbol{\tau}^\top - \boldsymbol{\tau}\boldsymbol{k}^\top + \|\boldsymbol{\tau}\|^2\boldsymbol{k}\boldsymbol{k}^\top). \tag{16}$$

Letting

$$\xi = \frac{1}{k^2}, \qquad \eta = \frac{\phi^2 - 1}{k^2}, \tag{17}$$

and equating (1,1), (2,2), and (1,2) elements on both sides of eq. (16), we obtain the following linear equation in ξ and η:

$$\begin{pmatrix} \|\boldsymbol{H}\boldsymbol{i}\|^2 & H_{31}^2 \\ \|\boldsymbol{H}\boldsymbol{j}\|^2 & H_{32}^2 \\ (\boldsymbol{H}\boldsymbol{i}, \boldsymbol{H}\boldsymbol{j}) & H_{31}H_{32} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}. \tag{18}$$

Solving this by least squares, we obtain the non-dimensional focal length ϕ in the following form:

$$\phi = \sqrt{1 + \frac{\eta}{\xi}}. \tag{19}$$

Equating (1,3) and (2,3) elements on both sides of eq. (16), we obtain τ_1 and τ_2 in the form

$$\begin{aligned} \tau_1 &= -\xi(\boldsymbol{H}\boldsymbol{i}, \boldsymbol{H}\boldsymbol{k}) - \eta H_{31}H_{33}, \\ \tau_2 &= -\xi(\boldsymbol{H}\boldsymbol{j}, \boldsymbol{H}\boldsymbol{k}) - \eta H_{32}H_{33}. \end{aligned} \tag{20}$$

Substituting these into the equation obtained by equating the (3,3) elements of eq. (16) on both sides, we obtain the following quadratic equation in τ_3:

$$\tau_3^2 - 2\tau_3 + (1 + \tau_1^2 + \tau_2^2 - \xi\|\boldsymbol{H}\boldsymbol{k}\|^2 - \eta H_{33}^2) = 0. \tag{21}$$

Since the camera is in front of the planar pattern, we choose from among the two solutions the one that satisfies $\tau_3 \leq 1$. The rotation $\boldsymbol{R}$ can be obtained from eq. (15). However, the resulting solution need not be an exact orthogonal matrix if there is noise in the data. So, we apply the following least-squares minimization:

$$\left\|\boldsymbol{R} - \frac{1}{k}(\boldsymbol{I} - \boldsymbol{k}\boldsymbol{\tau}^\top)^{-1}\boldsymbol{H}^\top\boldsymbol{A}\right\|^2 \to \min. \tag{22}$$

The solution is obtained analytically as follows. First, compute the following singular value decomposition:

$$\left(\boldsymbol{H}^\top + \frac{\boldsymbol{k}(\boldsymbol{H}\boldsymbol{\tau})^\top}{1 - (\boldsymbol{k}, \boldsymbol{\tau})}\right)\boldsymbol{A} = \boldsymbol{V}\,\mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)\boldsymbol{U}^\top. \tag{23}$$

Here, $\boldsymbol{V}$ and $\boldsymbol{U}$ are orthogonal matrix, and $\sigma_1 \geq \sigma_2 \geq \sigma_3$ (≥ 0) are the singular values. Then, the rotation matrix $\boldsymbol{R}$ that minimizes eq. (23) is given as follows [2]:

$$\boldsymbol{R} = \boldsymbol{V}\,\mathrm{diag}(1, 1, \det \boldsymbol{V}\boldsymbol{U}^\top)\boldsymbol{U}^\top. \tag{24}$$

6. SIMULATION 1

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, each time using different noise. Fig. 3(a) is the histogram of the computed focal length $\hat{f}$, and Fig. 3(b) is the corresponding histogram of the analytical solution. The vertical lines in the figures indicate the standard deviation (12) computed from the theoretical lower bound.

Fig. 4(a) is a 3-D plot of the distribution of the error vector $\Delta\boldsymbol{t} = \hat{\boldsymbol{t}} - \bar{\boldsymbol{t}}$ of translation, and Fig. 4(b) is the corresponding result for the analytical solution. The ellipses in the figures indicate the theoretical lower bound on the standard deviation in each orientation given by

$$(\Delta\boldsymbol{t}, V[\hat{\boldsymbol{\tau}}]^{-1}\Delta\boldsymbol{t}) = d^2, \tag{25}$$

where the covariance matrix $V[\hat{\boldsymbol{\tau}}]$ is obtained from eq. (11).

Fig. 5(a) is a 3-D plot of the error vector $\Delta\boldsymbol{\Omega}$ of rotation defined by eq. (13), and Fig. 5(b) is the corresponding result for the analytical solution. The ellipses in the figures indicate the theoretical lower bound on the standard deviation in each orientation given by

$$(\Delta\boldsymbol{\Omega}, V[\hat{\boldsymbol{R}}]^{-1}\Delta\boldsymbol{\Omega}) = 1. \tag{26}$$

Table 1 lists the standard deviations of the optimal and analytical solutions for 1,000 trials and their corresponding theoretical lower bounds. From these

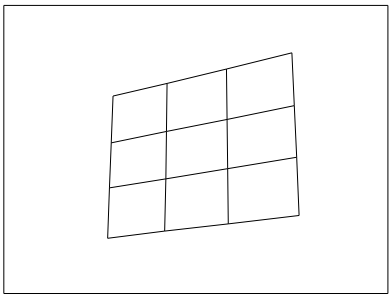
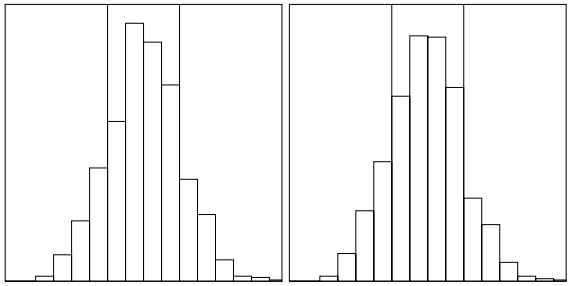
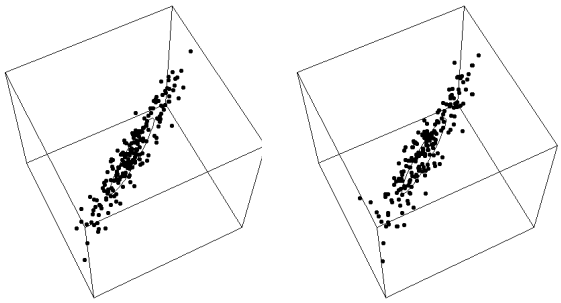


Figure 2: A simulated image of a grid pattern.



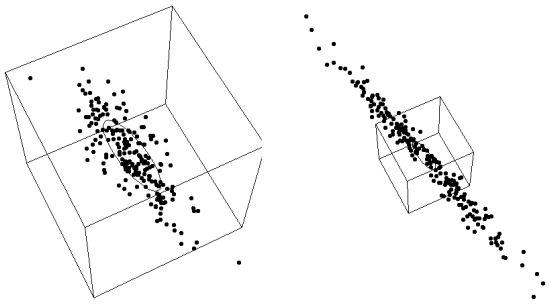
(a) (b)

Figure 3: Histogram of the computed focal length. (a) Optimal solution. (b) Analytical solution.



(a) (b)

Figure 4: Error distribution of the computed translation. (a) Optimal solution. (b) Analytical solution.



(a) (b)

Figure 5: Error distribution of the computed rotation. (a) Optimal solution. (b) Analytical solution.

Table 1: Standard deviations of the optimal and analytical solutions and their corresponding theoretical lower bounds.

	optimal	analytical	bound
focal length (pixles)	33.4	39.3	34.0
translation (cm)	32.9	38.5	32.6
rotation (deg)	0.413	1.292	0.414

results, we can see that our optimization scheme improves the accuracy as compared with the initial analytical solution. We can also see that the accuracy almost attains the theoretical lower bound. This means that the reliability of the solution can be evaluated by computing the lower bounds.

7. REAL IMAGE APPLICATION

Fig. 6 show consecutive real images of a tennis court taken by a continuously moving camera with varying zooming. Since the size of the tennis court is stipulated by an international rule, we can compute the focal length and the camera position by using that knowledge. Fig. 7(a) shows the estimated focal lengths. The vertical line segments indicate the confidence intervals given by eq. (12). Fig. 7(b) shows the computed heights of the camera. The confidence intervals are similarly indicated.

Fig. 8 shows the camera trajectory computed from them and viewed from above. The ellipses there indicate three times the standard deviation of the estimation in each orientation (actually they are elliposids viewed from above). We have superimposed the top view of the tennis court by mapping the five images via the estimated homography.

8. DEGENERACY CRITERION

If the camera optical axis is perpendicular to the planar pattern, the Hessian $\nabla^2 J$ is a singular matrix, so the inverse matrix in eq. (8) cannot be computed. This does not occur in practice due to image noise, but the computed inverse becomes very large, and the resulting solution is numerically very unstable. We decide that degeneracy has ocurred and the solution is meaningless if the confidence interval (12) contains negative values of f. This means that we adopt the following criterion:

$$V[\hat{\phi}] > \frac{\hat{\phi}^2}{9}. \tag{27}$$

The variance $V[\hat{\phi}]$ is given by the (1,1) element of the covariance matrix $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ given by eq. (11), so it equals $2\hat{\epsilon}^2(\nabla^2 \hat{J})^{\dagger}_{11}/N \det(\nabla^2 \hat{J})$, where $(\nabla^2 \hat{J})^{\dagger}_{11}$ is the (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. (27) can be rewritten in the following form:

$$\frac{18\hat{\epsilon}^2}{N}\left(\nabla^2 \hat{J}\right)^{\dagger}_{11} - \hat{\phi}^2 \det\left(\nabla^2 \hat{J}\right) > 0. \tag{28}$$

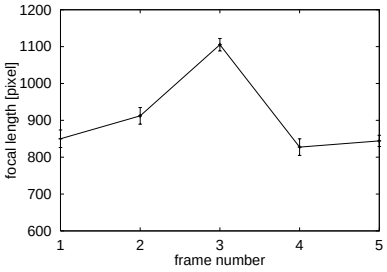
Since matrix inversion is no longer involved, this expression can be stably computed.

9. TRAJECTORY STABILIZATION

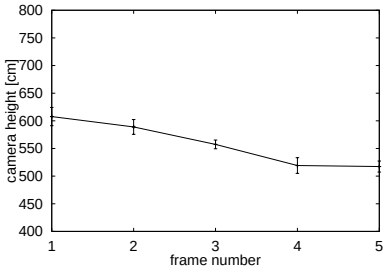
In order to stabilize the computaiton of the camera trajectory, we predict the focal length f and the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ in the next frame from the values f_i and $\{\boldsymbol{t}_i, \boldsymbol{R}_i\}$ of the current frame and the values f_{i-1} and $\{\boldsymbol{t}_{i-1}, \boldsymbol{R}_{i-1}\}$ 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.

Figure 6: Consecutive images of a tennis court.



(a)



(b)

Figure 7: (a) Estimated focal lengths and their confidence intervals. (b) Estimated camera heights and their confidence intervals.

***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

$$\hat{\epsilon}_s^2 = \frac{\hat{J}_p}{2 - 6/N}. \tag{29}$$

***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

$$\hat{\epsilon}_s^2 = \frac{\hat{J}_s}{2 - 6/N}. \tag{30}$$

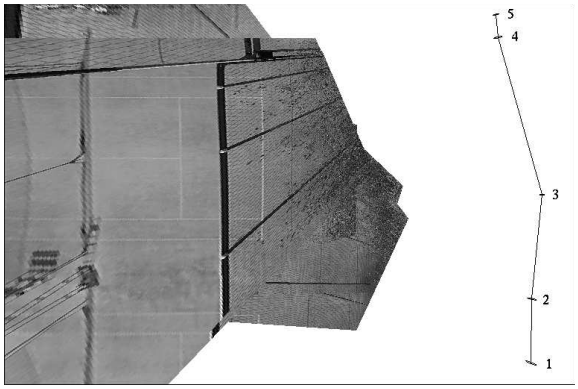


Figure 8: Estimated camera trajectory and confidence regions viewed from above.

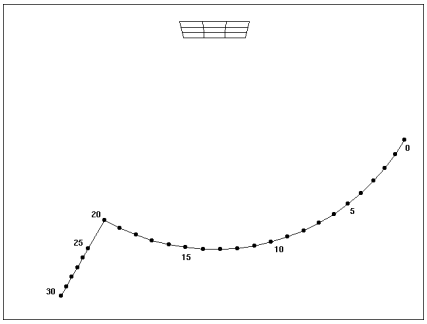


Figure 9: Sumulated camera motion.

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. (30) and evaluate the criterion (28). 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 of these models in the form

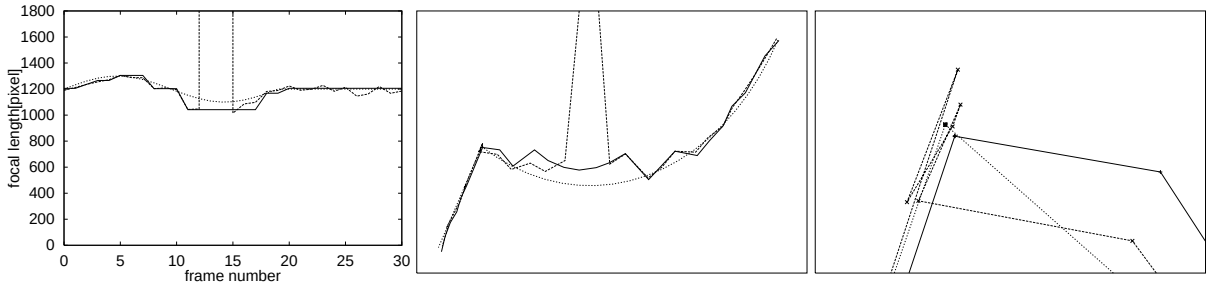
$$AIC_* = \hat{J}_*, \qquad AIC_s = \hat{J}_s + \frac{12}{N}\hat{\epsilon}^2,$$

$$AIC_p = \hat{J}_p + \frac{12}{N}\hat{\epsilon}^2, \qquad AIC_g = \hat{J}_g + \frac{14}{N}\hat{\epsilon}^2, \tag{31}$$

and choose the model that gives the smallest AIC.

If degeneracy is detected, we compare the stationary model, the ***t*-fixed model**, the ***t*-predicted model**, and the f-fixed model. Estimating the square noise level ϵ^2 by eq. (29), we evaluate the geometric AIC of these models in the form

$$AIC_* = \hat{J}_*, \qquad AIC_{s'} = \hat{J}_{s'} + \frac{6}{N}\hat{\epsilon}_s^2,$$



(a) (b) (c)

Figure 10: (a) Estimated focal lengths. The dashed lines indicate the values estimated independently at each frame. The dotted lines indicate the true values. (b) Estimated camera positions. The dashed lines indicate the camera trajectory computed independently at each frame. The dotted curve is the true trajectory. (c) Magnification of the portion for frames $20 \sim 24$.

$$AIC_{p'} = \hat{J}_{p'} + \frac{6}{N}\hat{\epsilon}_s^2, \qquad AIC_s = \hat{J}_s + \frac{12}{N}\hat{\epsilon}_s^2, \quad (32)$$

and choose the model that gives the smallest AIC.

By model selection, not only can we remove statistical fluctuations that occur when the camera motion is small but also obtain a stable trajectory when the camera undergoes a degenerate configuration. Also, we need not compute the initial guess for the Newton iterations frame by frame.

10. SIMULATION 2

We simulate a camera motion in a plane perpendicular to a 3×3 grid pattern. The camera is rotated so that the center of the pattern is fixed at the center of the image frame. The camera first moves along a circular trajectory as shown in Fig. 9. It comes in front of the pattern at frame 13. The camera stops at frame 20 and stays in the same position for five frames (frames $20 \sim 24$). Then, it recedes backward for another five frames (frames $25 \sim 30$).

Adding random Gaussian noise of mean 0 and standard deviation 0.5 (pixels) to each of the coordinate of the grid points independently at each frame, we compute the focal length f and the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ (Figs. 10(a) and 10(b)). Degeneracy is detected at frames 13 and 14. In order to emphasize the fact that the independent estimation fails, we let f be ∞ and the camera position be at the center of the grid pattern when degenerarcy is detected. We can see that the trajectory computed by model selection smoothly passes through the degenerate configuration.

Fig. 10(c) is a magnification of the portion for frames $20 \sim 24$ in Fig. 10(b). We can observe that statisfical fluctuations exist even when the camera is stationary if the camera position is computed independently at each frame. We can also observe that the statistical fluctuations are removed by model selection.

11. CONCLUDING REMARKS

We have studied the scheme for "simultaneous calibration" for computing the focal length and the position 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 attains the theoretical accuracy bound in the presence of nosie. This procedure also evaluates the reliability of the solution in the course of its computation. Finally, we showed that degeneracy of the solution and statistical fluctuations of computation can be avoided by predicting the focal length and the position of the camera from their history and comparing the predicted model with the actually computed model by the geometric AIC. We have confirmed the effectiveness of our method by simulations and shown a real image application. Our method is effective in analyzing 3-D motion from video images of live sports [7, 10, 11] and stably fusing CG with real human images in "virtual studio" applications [1].

Acknowledgments. This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research C(2) (No. 11680377).

References

[1] S. Gibbs, C. Arapis, C. Breiteneder, V. Lalioti, S. Mostafawy and J. Speier, Virtual studios: An overview. *IEEE Multimedia*, **5**-1 (1998), 24–35.

[2] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, 1993.

[3] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier Science, Amsterdam, 1996.

[4] K. Kanatani and N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, *Proc. 7th Int. Conf. Computer Vision*, September, 1999, Kerkya, Greece, pp. 73–78.

[5] 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.

[6] 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.

[7] C. Matsunaga and K. Kanatani, Calibration of a moveing camera using a planar pattern, *5th Symposium on Sensing via Image Information (SII'99)*, June 1999, Yokohama, Japan, G-1, pp. 255–260.

[8] P. Milgram and F. Kishino, A Taxonomy of Mixed Reality Visual Display, *IEICE Trans. Inf. & Sys.* **E77-D**-12 (1994).

[9] M. Polleyfes, R. Koch and L. Van Gool, Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, *Proc. Int. Conf. Comput. Vision*, January 1998, Bombay, India, pp. 90–95.

[10] 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.

[11] Y. Seo and K. S. Hong, Auto-calibration of a rotating and zooming camera, *Proc. IAPR Workshop on Machine Vision Applications*, November 1998, Makuhari, Chiba, Japan, pp. 274–277.

[12] 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.

[13] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*, Kluwer Academic, Dordrecht, The Netherlands, 1996.