１９９９年情報論的学習理論ワークショップ
1999 Workshop on Information-Based Induction Sciences (IBIS'99)
Izu, Japan, August 26-27, 1999.

Stabilizing Moving Camera Calibration from Images by the Geometric AIC

Chikara Matsunaga * Kenichi Kanatani †

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. This is one of the most fundamental techniques for mixed reality applications. We first point out that this scheme entails degeneracy of the solution and statistical fluctuations of computation. We then show that these can be avoided by predicting the focal length and the position of the camera from their values in preceding frames and comparing the predicted model with the actually computed model by the geometric AIC. We confirm the effectiveness of our method by simulations and show a real image application. Our method is effective in analyzing 3-D motion from video images of live sports and stably fusing CG with real human images in "virtual studio" applications.

Key words: 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, and 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 today. 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 the first step in all vision and media applications.

The standard method for it is *pre-calibration*: we estimate the camera 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 [15]. If the camera parameters are known, we can compute the 3-D camera position and the 3-D shape of the object up to scale by a technique called *structure from motion* [2, 16]. If we know the 3-D shapes of the objects or the 3-D structure of the scene that we are observing, we can compute the 3-D camera position from a single image. This problem has been studied by many researchers as *self-localization* for mobile robot applications [1, 6, 13, 14].

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 [10, 16]. Such a technique, known as *self-calibration*, may be very useful in unknown environments such as outdoors, but for stable reconstruction it requires a long sequence of images of the same scene taken from different and unconstrained camera positions and feature matching among them. It cannot be applied if the camera motion is constrained or if the scene changes as the camera moves, and the amount of computation required is too large for real-time applications.

In this paper, we focus on *virtual studio* applications: 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 position 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*.

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 completely 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

*Broadcast Division, FOR-A Co. Ltd., 2–3–3 Ohsaku, Sakura, Chiba 285-0802 Japan, Tel: (043)498-1239, Fax: (043)498-2223, E-mail:matsunaga@for-a.co.jp

† Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515 Japan, Tel: (0277)30-1842, Fax: (0277)30-1801, E-mail: kanatani@cs.gunma-u.ac.jp

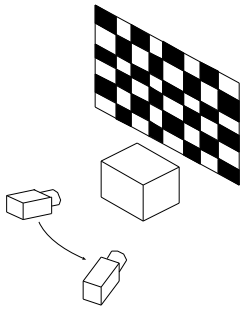


Figure 1: Simultaneous calibration of a moving camera.

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 [7, 8]. In order to deal with the second issue, we presented in [9] a technique for computing a statistically optimal solution and evaluating its reliability at the same time by introducing a mathematical model of image noise and applying the optimization theory of Kanatani [3]. The main purpose of this paper is to overcome the third and the fourth issues.

The degenerate configuration can be avoid by placing the pattern in such a way that it is unlikely to occur: if we know that the camera moves with its optical axis kept approximately horizontal, the pattern is slanted by some angle from its vertical position; if we take images of objects on a horizontal floor or on the ground with a hand-held camera, we can place the pattern horizontally. However, the possibility that the camera directly faces the planer pattern is never zero.

In this paper, we will 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, 4]. 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, respectively. 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]:

$$\boldsymbol{H} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & f_0/f \end{pmatrix} \boldsymbol{R}^\top \left(\boldsymbol{I} - \frac{\boldsymbol{t}\boldsymbol{k}^\top}{d} \right). \tag{3}$$

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. Eq. (2) defines an image transformation called *homography* [3]. A statistically optimal algorithm for computing it from feature correspondences by a technique called *renormalization* was presented by Kanatani [5], and its C++ code is publicly available via the Web[1].

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

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

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

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

the third column. If the noise has no particular dependence on positions and orientations, we use the default value diag$(1, 1, 0)$ (the diagonal matrix with diagonal elements 1, 1, and 0 in that order). If the noise is Gaussian, 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 $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, and $(\,\cdot\,)^-$ 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 the 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 following form:

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

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

$$\begin{pmatrix} \Delta\phi \\ \Delta\boldsymbol{\tau} \\ \Delta\boldsymbol{\Omega} \end{pmatrix} = -\left(\nabla^2 J\right)^{-1} \nabla J. \tag{8}$$

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.

The initial guess of ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ for the above iterations can be obtained by first computing the homography $\boldsymbol{H}$, say, by the renormalization procedure of Kanatani [5] and then decomposing it into ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ in the form of eq. (3), say, by the analytical procedure that we presented in [9]. However, this procedure is necessary only for the initial frame: for the subsequent frames, we can use as the initial guess 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 $\hat{J}$ 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

$$f_0\left(\hat{\phi} - 3\sqrt{V[\hat{\phi}]}\right) < f < f_0\left(\hat{\phi} + 3\sqrt{V[\hat{\phi}]}\right). \tag{12}$$

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

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

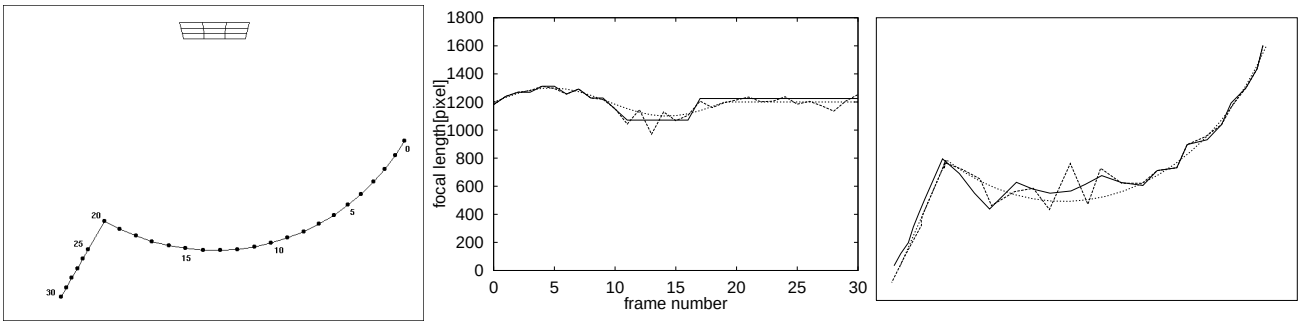
the submatrix of $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ defined by its fifth to seventh rows and columns gives the covariance matrix $V[\hat{\boldsymbol{R}}]$ of $\Delta\boldsymbol{\Omega}$.

5. Degeneracy Criterion

If the camera optical axis is perpendicular to the planar pattern, the Hessian $\nabla^2 J$ is a singular matrix. Hence, 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. So, 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:

$$V[\hat{\phi}] > \frac{\hat{\phi}^2}{9}. \tag{14}$$

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



(a) (b) (c)

Figure 2: (a) Sumulated camera motion. (b) Estimated focal lengths for $\beta = 80^\circ$. The dashed lines indicate the values estimated independently at each frame. The dotted lines indicate the true values. (c) Estimated camera positions for $\beta = 80^\circ$. The dashed lines indicate the camera trajectory computed independently at each frame. The dotted curve is the true trajectory.

(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. (14) can be rewritten in the following form:

$$\frac{18\hat{\epsilon}^2}{N}\left(\nabla^2 \hat{J}\right)^{\dagger}_{11} - \hat{\phi}^2 \det\left(\nabla^2 \hat{J}\right) > 0. \qquad (15)$$

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

6. 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.

***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

$$\hat{\epsilon}_s^2 = \frac{\hat{J}_p}{2 - 6/N}. \qquad (16)$$

***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

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

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.

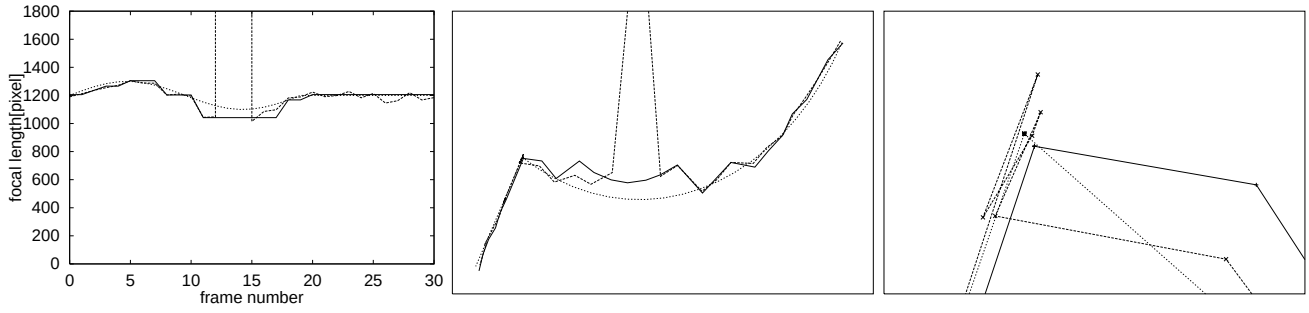
Degeneracy is detected from the f-predicted model. Namely, we estimate the square noise level ϵ^2 by eq. (17) and evaluate the criterion (15). If degeneracy is not detected, we compare the stationary model, the f-fixed model, the f-predicted model, and the general model. Estimating the square noise level ϵ^2 by eq. (10), we evaluate the geometric AICs of these models in the form

$$AIC_* = \hat{J}_*, \qquad AIC_s = \hat{J}_s + \frac{12}{N}\hat{\epsilon}^2,$$
$$AIC_p = \hat{J}_p + \frac{12}{N}\hat{\epsilon}^2, \qquad AIC_g = \hat{J}_g + \frac{14}{N}\hat{\epsilon}^2, \qquad (18)$$

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. (16), 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 3: (a) Estimated focal lengths for $\beta = 90^\circ$. The dashed lines indicate the values estimated independently at each frame. The dotted lines indicate the true values. (b) Estimated camera positions for $\beta = 90^\circ$. 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 $20 \sim 24$ ($\beta = 90^\circ$).

$$AIC_{p'} = \hat{J}_{p'} + \frac{6}{N}\hat{\epsilon}_s^2, \qquad AIC_s = \hat{J}_s + \frac{12}{N}\hat{\epsilon}_s^2, \tag{19}$$

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.

7. Simulations

We simulate a camera motion in a plane that makes angle β with 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. 2(a). 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 $25 \sim 30$). The focal length (in pixels) changes as follows (k indicates the frame number):

$$f = \begin{cases} 1200 + 100\sin 2\pi k/19 & k = 0, 1, ..., 19 \\ 1200 & k = 20, ..., 30 \end{cases}. \tag{20}$$

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 $\{\boldsymbol{t}, \boldsymbol{R}\}$. Fig. 2(b) shows the computed focal lengths in solid lines for $\beta = 80^\circ$. The dotted lines indicate the true values, and the dashed lines indicate the values estimated independently at each frame. Fig. 2(c) shows the corresponding camera positions viewed from above. From Figs. 2(b) and 2(c), we can see that statistical fluctuations are considerably supressed by model selection as compared with independent estimations.

Figs. 3(a) and 3(b) show the corresponding result for $\beta = 90^\circ$. 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. 3(c) is a magnification of the portion for frames $20 \sim 24$ in Fig. 3(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 completely removed by model selection.

8. Real Image Application

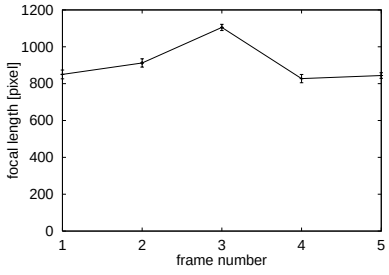
The first five images in Fig. 4 are consecutive real images of a tennis court taken by a continuously moving camera with varying zooming. The last one shows the camera trajectory computed from them and viewed from above. The ellipses there indicate three times the standard deviation of the estimation 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.

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. 5(a) shows the estimated focal lengths. The vertical line segments indicate the confidence intervals given by eq. (12). Fig. 5(b) shows the computed heights of the camera. The confidence intervals are similarly indicated.

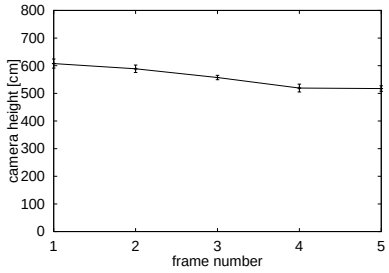
9. Concluding Remarks

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. This is one of the most fundamental techniques for mixed reality applications. We first pointed out that this scheme entails degeneracy of the solution and statistical fluctuations of computation. We then showed that these can be avoided by predicting the focal length and the position of the

Figure 4: Consecutive images of a tennis court and estimated camera trajectory and confidence regions viewed from above.



(a)



(b)

Figure 5: (a) Estimated focal lengths and their confidence intervals. (b) Estimated camera heights and their confidence intervals.

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 [9, 11, 12] and stably fusing CG with real human images in "virtual studio" applications.

References

[1] M. Betke and L. Gurvits, Mobile robot localization using landmarks, *IEEE Trans. Robotics Automation*, **13**-2 (1997), 251–263.

[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, Geometric information criterion for model selection, *Int. J. Comput. Vision*, **26**-3 (1998), 171–189.

[5] K. Kanatani, Optimal homography computation with a reliability measure, *Proc. IAPR Workshop on Machine Vision Applications*, November 1998, Makuhari, Japan, pp. 426–429.

[6] K. Kanatani and N. Ohta, Optimal robot self-localization and accuracy bounds, *IEICE Trans. Inf. & Syst.*, **E82-D**-2 (1999), 447–452.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] K. T. Sutherland and W. B. Thompson, Localizing in unconstrained environment: Dealing with the errors, *IEEE Trans. Robotics Automation*, **10**-6 (1994), 740–754.

[14] R. Talluri and J. K. Aggarwal, Mobile robot self-location using model-image feature correspondence, *IEEE Trans. Robotics Automation*, **12**-1 (1996), 63–77.

[15] 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.

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