Electronics and Communications in Japan, Part 3, Vol. 84, No. 7, 2001
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J83-A, No. 6, June 2000, pp. 694–701

Calibration of a Moving Camera Using a Planar Pattern: Optimal Computation, Reliability Evaluation, and Stabilization by the Geometric AIC

Chikara Matsunaga[1] and Kenichi Kanatani[2]

[1]Broadcast Division, FOR-A Co. Ltd., Chiba, 285-0802 Japan

[2]Department of Computer Science, Gunma University, Gunma, 376-8515 Japan

SUMMARY

We propose a "simultaneous calibration" scheme for computing the parameters of a continuously moving and continuously zooming camera by placing a planar pattern behind the object. We describe a procedure for computing an analytical solution which is exact in the absence of noise, estimating an optimal solution that attains the theoretical accuracy bound in the presence of noise, and evaluating the reliability of the computed solution. Finally, we show that degenerate configurations in which the solution is indeterminate and statistical fluctuations of computation can be avoided by "model selection." This is done by comparing the model predicted from the history of the camera motion with the actually computed model by the "geometric AIC." We confirm the effectiveness of our method by simulations and show real image examples. © 2001 Scripta Technica, Electron Comm Jpn Pt 3, 84(7): 12–21, 2001

Key words: Mixed reality; camera calibration; homography; reliability evaluation; model selection; geometric AIC.

1. Introduction

Visually presenting 3D shapes of real objects is one of the main goals of many Internet applications such as network cataloging and virtual museums [14], and generating virtual images by embedding graphics objects in real scenes or real objects in graphics scenes, known as *mixed reality* [18], is one of the central themes of image and media applications today [2, 10, 11, 19]. In order to reconstruct the 3D shapes of real objects or scenes for such applications, we need to know the 3D position of the camera that we use and its internal parameters, that is, we need to *calibrate* the camera.

The standard method is *precalibration*: we estimate the internal parameters in advance. Recently, techniques for computing both the camera parameters and the 3D positions of the camera from multiple images of the scene about which we have no prior knowledge have been studied intensively [12, 15, 21]. Such a technique, known as *self-calibration*, may be very useful in unknown environments such as outdoors, but it requires complicated procedures. In indoor environments where human control is possible, doing 3D reconstruction after calibrating the camera is much easier and more accurate.

In many applications, however, the camera position and the zooming change from frame to frame. Thus, the calibration must be done for each frame. This can be done by placing an easily distinguishable planar pattern with a known geometry in the scene (Fig. 1). We call this *simultaneous calibration*. Although the basic principle has been well known [3, 20], simultaneous calibration has many elements that do not appear in precalibration:

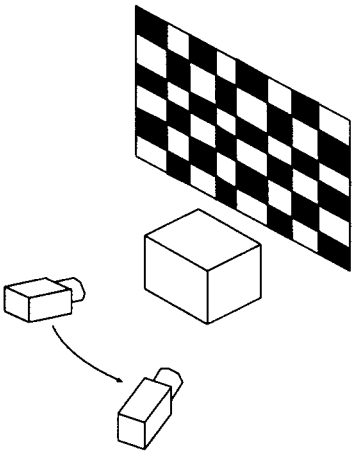


Fig. 1. Calibration of a moving camera.

- While manual interventions can be employed in precalibration, simultaneous calibration must be completely automated. For this purpose, we have designed a special marker pattern that makes automatic recognition easy [6, 8].
- In precalibration, the marker pattern can be placed in such a way that the computation becomes easy, whereas in simultaneous calibration the position of the pattern relative to the camera is not known a priori. Thus, the calibration accuracy is different from frame to frame.
- A sufficient number of marker points can be given in precalibration, while in simultaneous calibration only a small number of marker points may be visible depending on the camera position.
- When the camera optical axis is perpendicular to the pattern, zooming out cannot be distinguished from moving the camera forward as long as we use a planar pattern.
- Visible marker points change as the object moves in the scene, so the computed camera position may be slightly different from frame to frame even if the camera is stationary in the scene. This type of statistical fluctuation is conspicuous when the camera motion is small.

In this paper, we describe a procedure for maximizing the accuracy and evaluating the reliability of the computed solution by applying the statistical optimization theory [4]. Then, we show that the degeneracy and statistical fluctuations can be avoided by *model selection*. This is done by comparing the model predicted from the history of the camera motion with the actually computed model by the *geometric AIC* [5]. We confirm the effectiveness of our method by simulations and show real image examples.

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 assume that all camera parameters other than the focal length have already been precalibrated so that the imaging geometry can be modeled as a perspective projection.

We place a hypothetical camera of 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 *f*. We call $\{\boldsymbol{t}, \boldsymbol{R}\}$ the *motion parameters*.

Suppose a point on the planar pattern with coordinates (X, Y, d) is observed at (x, y) in the image. Define the 3D vectors

$$\boldsymbol{X} = \begin{pmatrix} X/d \\ Y/d \\ 1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x/f_0 \\ y/f_0 \\ 1 \end{pmatrix} \tag{1}$$

We have the following relationship in the absence of noise [4]:

$$\boldsymbol{x} = Z[\boldsymbol{H}\boldsymbol{X}] \tag{2}$$

The operation $Z[\cdot]$ denotes normalization to make the third component 1. The matrix $\boldsymbol{H}$ has the following form [4]:

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

The superscript $\top$ denotes transpose, and we put $\boldsymbol{k} = (0, 0, 1)^\top$; diag(···) denotes the diagonal matrix with diagonal elements ···. Equation (2) defines an image transformation called *homography* [4]; the matrix $\boldsymbol{H}$ is also called the "homography" when no confusion can occur. A program for optimally computing the homography and evaluating its reliability is publicly available [7, 17].

We set the distance *d* of the pattern from the hypothetical camera and its focal length f_0 to standard values. The coordinates (X, Y) of marker points are known. The image coordinates (x, y) are observed data. Hence, the unknowns are the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ and the focal length *f* after the camera motion. Since the rotation matrix $\boldsymbol{R}$ has only three degrees of freedom, the total number of unknowns is seven. Hence, they can be determined if four or more marker points are observed.

3. Optimal Computation

Suppose we observe N marker points. Let $\boldsymbol{X}_\alpha$ and $\boldsymbol{x}_\alpha$ be the values of $\boldsymbol{X}$ and $\boldsymbol{x}$ of the α-th point, respectively. Let $V[\boldsymbol{x}_\alpha]$ be the covariance matrix of $\boldsymbol{x}_\alpha$. Since its absolute magnitude is difficult to predict, we assume that it is known up to scale and write

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

We call the unknown constant ϵ the *noise level* and the known matrix $V_0[\boldsymbol{x}_\alpha]$, which describes the relative characteristics of the noise, 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 positions or orientations, we use the default value diag(1, 1, 0).

An optimal estimate of $\boldsymbol{H}$ is obtained by minimizing the sum of the squared *Mahalanobis distances*

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

Throughout this paper, we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$, and $(\cdot)^-$ means the (Moore–Penrose) generalized inverse.

Let us introduce the following nondimensional variables:

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

If no noise exists, an exact solution that minimizes Eq. (5) can be obtained analytically (Appendix 1). In the presence of noise, we apply Newton iterations starting from the analytical solution. We write the gradient ∇J and the Hessian $\nabla^2 J$ of the function J with respect to the unknowns ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ as follows:

$$\nabla J = \begin{pmatrix} \partial J/\partial\phi \\ \nabla_\tau J \\ \nabla_{\mathbf{R}} J \end{pmatrix} \tag{7}$$

$$\nabla^2 J = \begin{pmatrix} \partial^2 J/\partial^2\phi & (\nabla_\tau \partial J/\partial\phi)^\top & (\nabla_{\mathbf{R}} \partial J/\partial\phi)^\top \\ \nabla_\tau \partial J/\partial\phi & \nabla^2_{\tau\tau} J & (\nabla^2_{\mathbf{R}\tau} J)^\top \\ \nabla_{\mathbf{R}} \partial J/\partial\phi & \nabla^2_{\mathbf{R}\tau} J & \nabla^2_{\mathbf{R}\mathbf{R}} J \end{pmatrix} \tag{8}$$

The procedure for the Newton iterations is as follows:

(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 the update increments $\Delta\phi$, $\Delta\boldsymbol{\tau}$, and $\Delta\boldsymbol{\Omega}$ by

$$\begin{pmatrix} \Delta\phi \\ \Delta\boldsymbol{\tau} \\ \Delta\boldsymbol{\Omega} \end{pmatrix} = -\left(\nabla^2 J\right)^{-1} \nabla J \tag{9}$$

(4) Update ϕ, $\boldsymbol{\tau}$, and $\boldsymbol{R}$ as follows and go back to Step 2 until $|\Delta\phi| < \varepsilon_\phi$, $\|\Delta\boldsymbol{\tau}\| < \varepsilon_\tau$, and $\|\Delta\boldsymbol{\Omega}\| < \varepsilon_{\mathbf{R}}$:

$$\phi \leftarrow \phi + \Delta\phi, \quad \boldsymbol{\tau} \leftarrow \boldsymbol{\tau} + \Delta\boldsymbol{\tau}, \quad \boldsymbol{R} \leftarrow \mathcal{R}(\Delta\boldsymbol{\Omega})\boldsymbol{R} \tag{10}$$

Here, $\mathcal{R}(\Delta\boldsymbol{\Omega})$ denotes the rotation matrix of angle $\|\Delta\boldsymbol{\Omega}\|$ around $\Delta\boldsymbol{\Omega}$, and ε_ϕ, ε_τ, and $\varepsilon_{\mathbf{R}}$ are thresholds for convergence. See Ref. 9 for the actual expressions for the elements of the gradient ∇J and the Hessian $\nabla^2 J$.

4. Reliability Evaluation

The squared noise level ε^2 can be estimated from the residual $\hat{J}$ as follows [4] (Appendix 2):

$$\hat{\varepsilon}^2 = \frac{\hat{J}}{2 - 7/N} \tag{11}$$

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 [4] (if the true values of $\{\phi, \boldsymbol{\tau}, \boldsymbol{R}\}$ are substituted, this gives a theoretical lower bound on the covariance matrix $V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}]$ [4]):

$$V[\hat{\phi}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{R}}] = \frac{2\hat{\varepsilon}^2}{N}\left(\nabla^2 \hat{J}\right)^{-1} \tag{12}$$

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

$$\hat{\phi} - 3\sqrt{V[\hat{\phi}]} < \frac{f}{f_0} < \hat{\phi} + 3\sqrt{V[\hat{\phi}]} \tag{13}$$

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 the (normalized) translation $\boldsymbol{\tau}$. Let $\overline{\boldsymbol{R}}$ be the true rotation. Let $\Delta\boldsymbol{\Omega}$ and $\boldsymbol{l}$ be, respectively, the angle and axis of the relative rotation $\hat{\boldsymbol{R}}\overline{\boldsymbol{R}}^\top$. We define the error vector of the rotation by

$$\Delta\boldsymbol{\Omega} = \Delta\Omega\boldsymbol{l} \tag{14}$$

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. Simulation 1

Figure 2 shows a simulated image of a grid pattern viewed from an angle. We added Gaussian random noise of

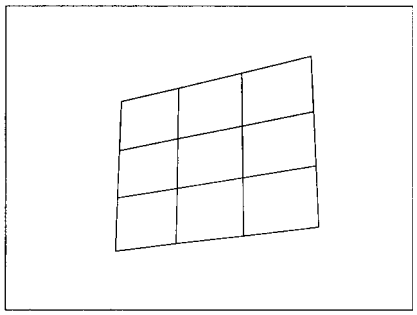
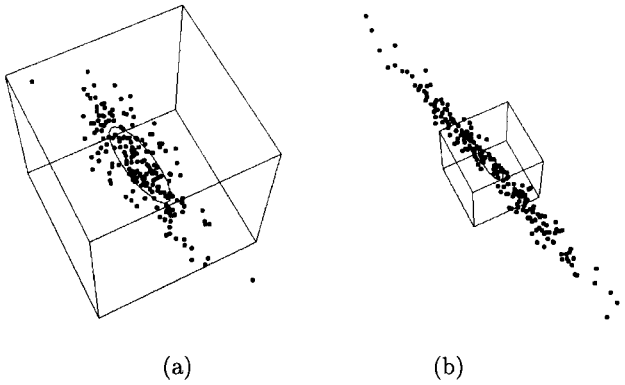
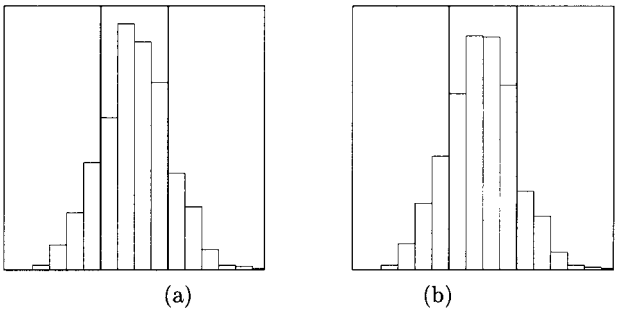


Fig. 2. A simulated image of a grid pattern.



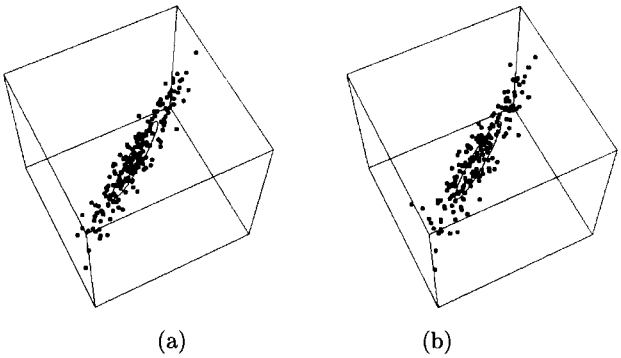
(a) (b)

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



(a) (b)

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



(a) (b)

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

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 1000 times.

Figure 3(a) is the histogram of the computed focal length $\hat{f}$; Fig. 3(b) is the corresponding histogram of the analytical solution. The vertical lines in the figures indicate the theoretical lower bound on the standard deviation around the true value $\bar{f}$.

Figure 4(a) is the 3D plot of the error vectors of translation; 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.

Figure 5(a) is the 3D plot for the error vector $\Delta\boldsymbol{\Omega}$ of rotation defined by Eq. (14); 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.

Table 1 lists standard deviations of the optimal and analytical solutions for 1000 trials and their corresponding theoretical lower bounds.

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

	Optimal	Analytical	Bound
Focal length (pixels)	33.4	39.3	34.0
Translation (cm)	32.9	38.5	32.6
Rotation (deg)	0.413	1.292	0.414

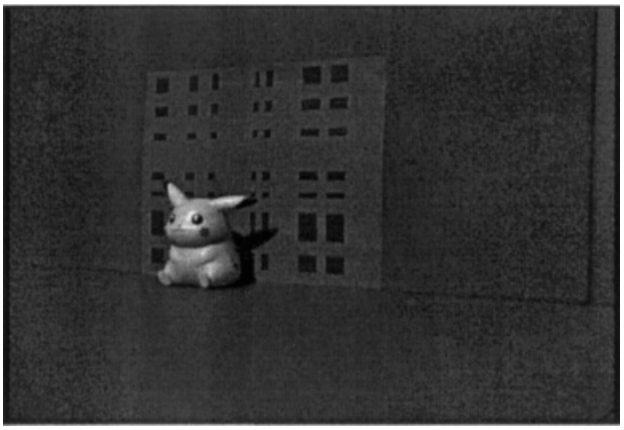


Fig. 6. An object and a grid pattern behind it.

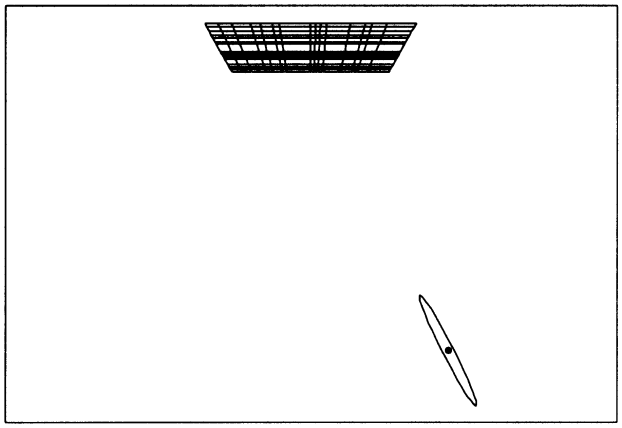


Fig. 8. Estimated camera position and its reliability.

From these 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.

6. Real Image Experiment 1

Figure 6 is a real image of an object, behind which is placed a grid pattern perpendicularly on the floor. The camera optical axis is almost parallel to the floor. This pattern is so designed that the image grid can be matched to the original grid in an optimal way by measuring the cross ratio of adjacent vertices [6, 8].

The pattern is colored light and dark blue, so the object image can be segmented by using a chromakey technique. We calibrated the camera, using the grid pattern image. See Ref. 8 for the image processing details. The focal length is estimated to be 576.1 pixels. The standard deviations of the focal length, the translation, and the rotation are evaluated to be 38.32 pixels, 5.73 cm, and 0.812 deg, respectively.

Figure 7 is a composition of the toy image and a graphics scene generated by VRML [1]. Figure 8 is the top view of the estimated camera position and its uncertainty ellipsoid (three times the standard deviation in each orientation).

7. Real Image Experiment 2

Figure 9 is a real image of a tennis court. Since the size of the court is stipulated by an international rule, we

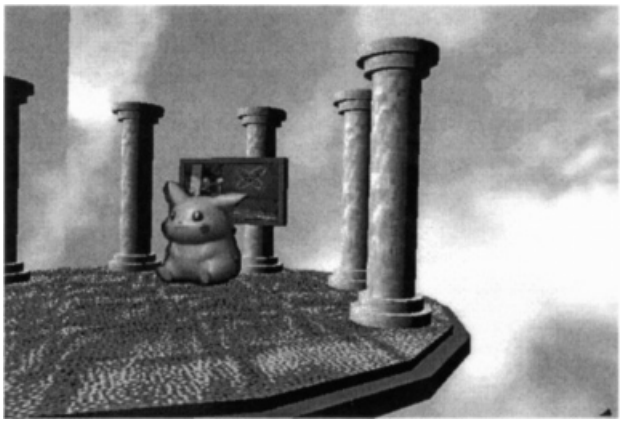


Fig. 7. Composition with a virtual scene.

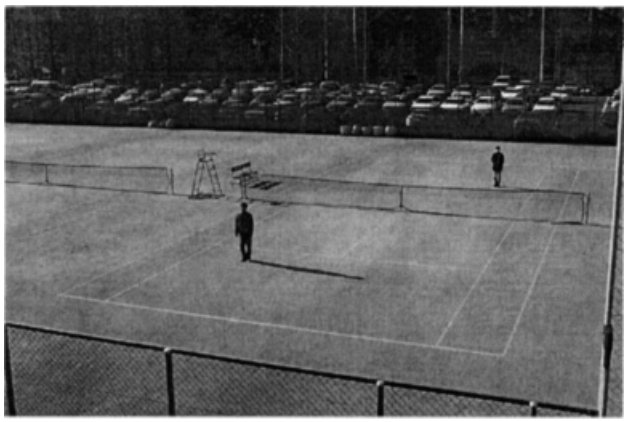


Fig. 9. A real image of a tennis court.

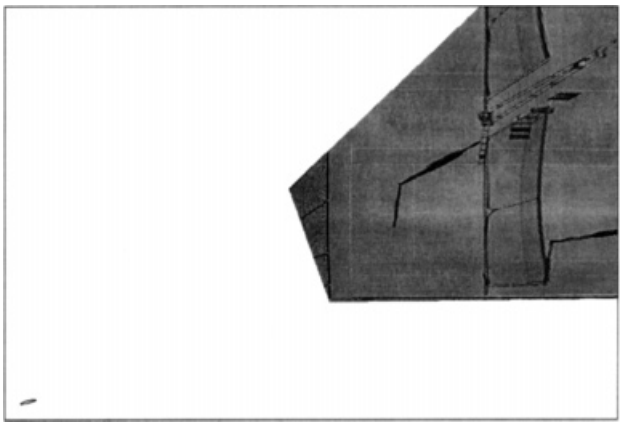


Fig. 10. The computed camera position viewed from above.

can compute the 3D camera position and the focal length by using this knowledge. Figure 10 shows the top view of the tennis court generated from Fig. 9 by applying the computed homography. The estimated camera position and the uncertainty ellipsoid are shown in the same way as in Fig. 8.

The focal length is estimated to be 954.8 pixels. The camera is estimated to be 627.1 cm above the ground. The standard deviations of the focal length, the translation, and the rotation are evaluated to be 6.99 pixels, 16.14 cm, and 0.151 deg, respectively.

The images of the poles and the persons in Fig. 10 can be regarded as their "shadows" on the ground cast by hypothetical light emitted from the camera, so we can compute their heights. The right pole is estimated to be 113.4 cm high. The person near the camera is estimated to be 170.5 cm tall. Using this technique, we can compute the 3D positions and shapes of the scenes and persons from one image in sports broadcasting [13, 16].

8. Degeneracy Criterion

As pointed out in Section 1, the Hessian $\nabla^2 J$ in Eq. (8) becomes singular when the camera optical axis is perpendicular to the planar pattern, so the inverse in Eq. (9) cannot be computed. It can be computed in the presence of noise, but the resulting value is unreliable. We decide that the solution is meaningless due to degeneracy if the confidence interval (13) contains negative values. This criterion is written as $V[\hat{\phi}] > \hat{\phi}^2/9$. Since the variance $V[\hat{\phi}]$ is given by the (1,1) element of Eq. (12), it equals $2\hat{\varepsilon}^2(\nabla^2\hat{J})^{\dagger}_{11}/N \det(\nabla^2\hat{J})$, 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, our criterion is

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

This expression can be computed stably because no matrix inversion is involved, but the solution to be substituted cannot be computed stably. Therefore, we substitute an approximation to it instead of the exactly optimal solution (the details are shown shortly).

9. Trajectory Stabilization

As pointed out in Section 1, statistical fluctuations cannot be avoided if image processing is done independently at each frame even when the camera motion is continuous. The fluctuations are particularly conspicuous when the camera motion is small. In order to prevent them, we predict the camera motion and do model selection by the geometric AIC [5] (Appendix 2). Let f_i and $\{\boldsymbol{t}_i, \boldsymbol{R}_i\}$, be the focal length and the motion parameters of the current frame, respectively, and f_{i-1} and $\{\boldsymbol{t}_{i-1}, \boldsymbol{R}_{i-1}\}$ those of the preceding frame. Here, we consider the following six models:

- Stationary model (zero degrees of freedom): We fix $f = f_i$, $\boldsymbol{t} = \boldsymbol{t}_i$, and $\boldsymbol{R} = \boldsymbol{R}_i$. Let $\hat{J}_*$ be the corresponding residual.
- $\boldsymbol{t}$-fixed model (three degrees of freedom): We fix $f = f_i$ and $\boldsymbol{t} = \boldsymbol{t}_i$ and optimize $\boldsymbol{R}$ by iterations from the initial value $\boldsymbol{R}_i$. Let $\hat{J}_{s'}$ be the corresponding residual.
- $\boldsymbol{t}$-predicted model (three degrees of freedom): We fix $f = f_i$, predict $\boldsymbol{t} = 2\boldsymbol{t}_i - \boldsymbol{t}_{i-1}$, and optimize $\boldsymbol{R}$ by iterations from the initial value $\boldsymbol{R}_i\boldsymbol{R}_{i-1}^{\top}\boldsymbol{R}_i$. Let $\hat{J}_{p'}$ be the corresponding residual.
- f-fixed model (six degrees of freedom): We fix $f = f_i$ and optimize $\{\boldsymbol{t}, \boldsymbol{R}\}$ by iterations from the initial values $\{\boldsymbol{t}_i, \boldsymbol{R}_i\}$. Let $\hat{J}_s$ be the corresponding residual. We estimate the square noise level ε^2 by

$$\hat{\varepsilon}_s^2 = \frac{\hat{J}_s}{2 - 6/N} \tag{16}$$

- f-predicted model (six degrees of freedom): We predict $f = 2f_i - f_{i-1}$ and optimize $\{\boldsymbol{t}, \boldsymbol{R}\}$ by iterations from the initial values $\{2\boldsymbol{t}_i - \boldsymbol{t}_{i-1}, \boldsymbol{R}_i\boldsymbol{R}_{i-1}^{\top}\boldsymbol{R}_i\}$. Let $\hat{J}_p$ be the corresponding residual. We estimate the square noise level ε^2 by

$$\hat{\varepsilon}_p^2 = \frac{\hat{J}_p}{2 - 6/N} \tag{17}$$

- General model (seven degrees of freedom): We optimize f and $\{\boldsymbol{t}, \boldsymbol{R}\}$ by iterations starting from the prediction of the f-predicted model. Let $\hat{J}_g$ be the corresponding residual.

Since the solution of the general model cannot be determined stably when degeneracy occurs, the occurrence of degeneracy is judged using the f-predicted model. Namely, we estimate the square noise level ε^2 by Eq. (17) and evaluate the criterion (15), approximating the solution of the general model by the solution of the f-predicted model.

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. (11), we evaluate the geometric AICs of these models in the following form (Appendix 2):

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

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. (16), we evaluate the geometric AIC of these models in the following form (Appendix 2):

$$AIC_* = \hat{J}_*, \quad AIC_{s'} = \hat{J}_{s'} + \frac{6}{N}\hat{\varepsilon}_s^2,$$
$$AIC_{p'} = \hat{J}_{p'} + \frac{6}{N}\hat{\epsilon}_s^2, \quad AIC_s = \hat{J}_s + \frac{12}{N}\hat{\varepsilon}_s^2 \qquad (19)$$

In either case, we choose the model that gives the smallest geometric AIC. By model selection, not only can we remove statistical fluctuations that occur when the camera motion is small, we can also obtain a stable trajectory when the camera passes through a degenerate configuration. Also, we need not compute the initial guess analytically frame by frame.

10. Simulation 2

We simulate a camera motion in a plane perpendicular to a 3×3 grid pattern as shown in Fig. 11; the camera rotates so that it always gazes at the center of the grid. The camera faces the pattern perpendicularly at frame 13 and stops at frame 20. The camera stays there over frames 20 to 24 and then recedes backward over frames 25 to 30. Adding random Gaussian noise of mean 0 and standard deviation 0.5 (pixel) to each coordinate of the grid points independently, we compute the focal length and the camera trajectory.

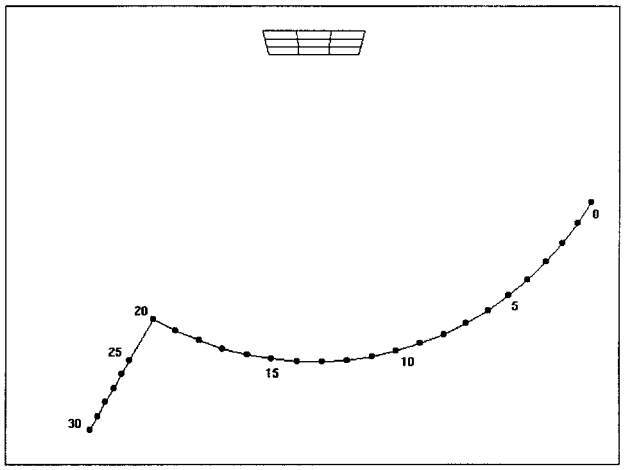


Fig. 11. Simulated camera motion.

Figures 12 and 13 show the computed focal length and the camera trajectory, respectively. Degeneracy is detected at frames 13 and 14. In order to emphasize the fact that the independent estimation fails, we let f be ∞ in Fig. 12 and the camera position be at the center of the grid in Fig. 13 when degeneracy is detected.

The models selected by the geometric AIC are, in the order of the camera motion, 3, 4, 5, 4, 3, 5, 3, 3, 5, 3, 3, 5, 3, 3^*, 3^*, 4, 3, 4, 5, 3, 5, 0, 0, 0, 0, 3, 3, 3, 3, 3, 3, where 0, 1, 2, 3, 4, and 5 stand for the stationary model, the $\boldsymbol{t}$-fixed model, the $\boldsymbol{t}$-predicted model, the f-fixed model, the f-pre-

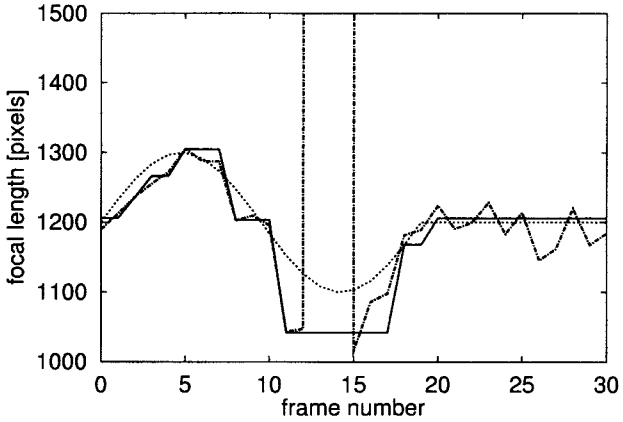


Fig. 12. Estimated focal length. Model selection (solid lines), independent estimation (chained lines), and the true values (dotted lines).

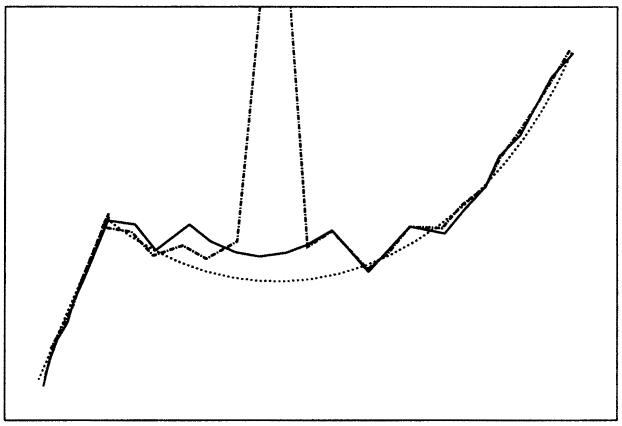


Fig. 13. Estimated trajectory. Model selection (solid lines), independent estimation (chained lines), and the true trajectory (dotted lines).

dicted model, and the general model, respectively (the asterisks indicate degeneracy).

Figure 14 is a magnification of the trajectory over frames 20 to 24. Although the camera is stationary, we can observe irregular fluctuations when the computation is done independently at each frame. Our model selection chooses the stationary model there, thereby removing the fluctuations.

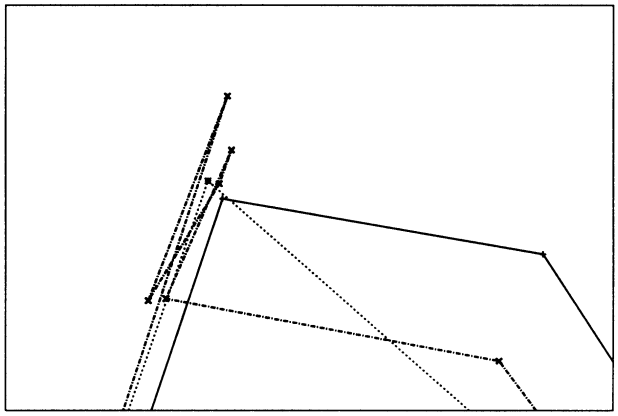


Fig. 14. Magnification of Fig. 13.

11. Concluding Remarks

With a view to mixed reality applications, we have proposed a scheme for "simultaneous calibration" for computing the camera parameters of a continuously moving and continuously zooming camera by placing a planar pattern behind the object.

We have described a procedure for computing an analytical solution which is exact in the absence of noise, estimating an optimal solution that attains the theoretical accuracy bound in the presence of noise, and evaluating the reliability of the computed solution. Finally, we have shown that degenerate configurations in which the solution is indeterminate and statistical fluctuations of computation can be avoided by model selection. This is done by comparing the model predicted from the history of the camera motion with the actually computed model by the geometric AIC. We have confirmed the effectiveness of our method by simulations and shown real image examples.

In this paper, the camera motion models were introduced according to our empirical intuition. The question as to how many models and what models we should introduce needs to be studied further. Also, much remains to be studied about the criterion for degeneracy detection and the use of information criteria other than the geometric AIC (e.g., MDL-like ones).

REFERENCES

1. Carey R, Bell G. The annotated Vrml 2.0 reference manual. Addison Wesley Longman; 1997.
2. Gibbs S, Arapis C, Breiteneder C, Lalioti V, Mostafawy S, Speier J. Virtual studios: An overview. IEEE Multimedia 1998;5:24–35.
3. Kanatani K. Geometric computation for machine vision. Oxford University Press; 1993.
4. Kanatani K. Statistical optimization for geometric computation. Elsevier Science; 1996.
5. Kanatani K. Geometric model selection by an information criterion. Trans Inf Process Soc Japan 1996;37:1073–1080. (in Japanese)
6. Kanazawa Y, Matsunaga C, Kanatani K. Best marker pattern design for recognition by cross ratio. IPSJ SIG Notes, 99-CVIM-115-13, p 97–104, 1999. (in Japanese)
7. Kanazawa Y, Ohta N, Kanatani K. Image mosaicing by optimally computed homography. IPSJ SIG Notes, 99-CVIM-116-2, p 9–16, 1999. (in Japanese)
8. Matsunaga C, Niijima K, Kanatani K. Best marker pattern design for recognition by cross ratio: Experimental investigation. IPSJ SIG Notes, 99-CVIM-115-14, p 105–110, 1999. (in Japanese)

9. Matsunaga C, Kanatani K. Calibration of a moving camera using a planar surface. IPSJ SIG Notes, 99-CVIM-116-1, p 1–8, 1999. (in Japanese)
10. Miyata K, Kurokawa M, Hirota G. Fusion of image and graphics. J IPSJ 1997;38:189–194. (in Japanese)
11. Nakajima M. Advanced digital image processing: Basic theories and applications (II): Introduction to 2-dimensional digital image processing. J IEICE 1998;18:68–75. (in Japanese)
12. Pollefeys M, Koch R, Van Gool L. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. Int J Comput Vision 1999;32:7–26.
13. Reid I, Zisserman A. Goal-directed video metrology. Proc 4th Eur Conf Comput Vision, 1996, Cambridge, U.K., Vol. 2, p 647–658.
14. Sakamura K. Digital museum—Construction of a new type of museum that fully utilizes computers. J IPSJ 1998;39:385–392. (in Japanese)
15. Sato J. Computer vision—Geometry of vision. Corona Publishing; 1999. (in Japanese)
16. Seo Y, Hong KS. Auto-calibration of a rotating and zooming camera. Proc IAPR Workshop Machine Vision Appl, 1998, Makuhari, p 274–277.
17. Shimizu Y, Ohta N, Kanatani K. Optimal homography computation program with a reliability measure. IPSJ SIG Notes, 99-CVIM-111-5, p 33–40, 1999. (in Japanese)
18. Tamura H. Mixed reality systems: A perspective view from image sensing. Proc 5th Symp Sensing via Image Information, 1999, Yokohama, p 169–178. (in Japanese)
19. Tamir M. The Orad virtual set. Int Broadcast Eng, March, 1996, p 16–18.
20. Tsai RY. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. J Robotics Automation 1987;3:323–344.
21. Xu G, Tsuji S. 3D vision. Kyoritsu Shuppan; 1998. (in Japanese)

APPENDIX

1. Analytical Solution

Given a homography $\boldsymbol{H} = (H_{ij})$ up to scale, it can be analytically decomposed into the focal length f, the translation $\boldsymbol{t}$, and the rotation $\boldsymbol{R}$ that satisfy Eq. (3) as follows (see Ref. 9 for the derivation). First, we solve the following linear equation in ξ and η, say, by least squares:

$$\begin{pmatrix} \|\boldsymbol{Hi}\|^2 & H_{31}^2 \\ \|\boldsymbol{Hj}\|^2 & H_{32}^2 \\ (\boldsymbol{Hi}, \boldsymbol{Hj}) & H_{31}H_{32} \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \tag{A.1}$$

Here, we let $\boldsymbol{i} = (1, 0, 0)^\top$ and $\boldsymbol{j} = (0, 1, 0)^\top$. From the first of Eqs. (6), the normalized focal length ϕ is given by

$$\phi = \sqrt{1 + \frac{\eta}{\xi}} \tag{A.2}$$

The normalized translations τ_1 and τ_2 defined by the second of Eqs. (6) are given by

$$\begin{aligned} \tau_1 &= -\xi(\boldsymbol{Hi}, \boldsymbol{Hk}) - \eta H_{31}H_{33} \\ \tau_2 &= -\xi(\boldsymbol{Hj}, \boldsymbol{Hk}) - \eta H_{32}H_{33} \end{aligned} \tag{A.3}$$

The third element τ_3 is given as the solution of the quadratic equation

$$\tau_3^2 - 2\tau_3 + (1 + \tau_1^2 + \tau_2^2 - \xi\|\boldsymbol{Hk}\|^2 - \eta H_{33}^2) = 0 \tag{A.4}$$

where from among the two solutions the one that satisfies $\tau_3 \leq 1$ is chosen. For computing the rotation $\boldsymbol{R}$, we first compute the following singular value decomposition:

$$\left(\boldsymbol{H}^\top + \frac{\boldsymbol{k}(\boldsymbol{H\tau})^\top}{1 - (\boldsymbol{k}, \boldsymbol{\tau})}\right) \boldsymbol{A} = \boldsymbol{V\Lambda U}^\top \tag{A.5}$$

Here, $\boldsymbol{A}$ = diag(1, 1, ϕ), and $\boldsymbol{V}$ and $\boldsymbol{U}$ are orthogonal matrices, while $\boldsymbol{\Lambda}$ is the diagonal matrix having the singular values as its diagonal element in the order of magnitude. The rotation matrix $\boldsymbol{R}$ is given by

$$\boldsymbol{R} = \boldsymbol{V}\mathrm{diag}(1, 1, \det(\boldsymbol{VU}^\top))\boldsymbol{U}^\top \tag{A.6}$$

The solution is exact if there is no noise in the data. In the presence of noise, it is an approximation in the least-squares sense; $\boldsymbol{R}$ is always an exact rotation matrix.

2. Geometric Fitting

Geometric fitting is the problem of estimating the parameter $\boldsymbol{u}$ given by a *constraint*

$$F^{(k)}(\boldsymbol{a}, \boldsymbol{u}) = 0, \quad k = 1, ..., L \tag{A.7}$$

and multiple realizations $\{\boldsymbol{a}_\alpha\}$ of the variable $\boldsymbol{a}$ [4]. We assume that the noise is Gaussian and the normalized covariance matrix $V_0[\boldsymbol{a}_\alpha]$ (a singular matrix in general) is known. The *maximum likelihood estimation* is to minimize the sum of the square Mahalanobis distances

$$J = \sum_{\alpha=1}^{N} (\boldsymbol{a}_\alpha - \bar{\boldsymbol{a}}_\alpha, V_0[\boldsymbol{a}_\alpha]^{-1}(\boldsymbol{a}_\alpha - \bar{\boldsymbol{a}}_\alpha)) \tag{A.8}$$

with respect to the unknowns $\boldsymbol{u}$ and $\{\bar{\boldsymbol{a}}_\alpha\}$ subject to the constraint $F^{(k)}(\bar{\boldsymbol{a}}_\alpha, \boldsymbol{u}) = 0$. Let $\hat{J}$ be the residual (the mini-

mum of J), and ε the noise level. It can be shown that $\hat{J}/\varepsilon^2$ is subject to a χ^2 distribution with $rN - n'$ degrees of freedom in the first order [4], where r is the rank (the number of independent equations) of the constraint (A.7); n' is the degree of freedom of the unknown $\boldsymbol{u}$. An unbiased estimator of the square noise level ε^2 is obtained in the following form:

$$\hat{\varepsilon}^2 = \frac{\hat{J}}{rN - n'} \tag{A.9}$$

If the constraint (A.7) defines a d-dimensional model (manifold) in the data space (the space of the variable $\boldsymbol{a}$), the geometric AIC is defined as follows [4, 5]:

$$AIC = \hat{J} + 2(Nd + n')\varepsilon^2 \tag{A.10}$$

Using this, we can detect degeneracy of the model by comparing the geometric AICs of the general model and its degeneracy, where the noise level is estimated from the general model by Eq. (A.9).

The constraint (2) has rank $r = 2$; the dimension of the model is $d = 0$. The degree of freedom of the unknowns $\{\boldsymbol{t}, \boldsymbol{R}\}$ and f is $n' = 7$ if the camera motion and zooming are unconstrained. If they are constrained, we obtain the expressions given in Section 9. The function J given in Eq. (5) involves division by N for computational convenience, so $\hat{J}$ in Eqs. (A.9) and (A.10) is replaced by $N\hat{J}$. Also, we defined Eq. (A.10) divided by N to be the geometric AIC in Section 9.

AUTHORS (from left to right)

Chikara Matsunaga received his B.E. and M.E. degrees in electronic engineering from Muroran Institute of Technology in 1988 and 1990. Currently, he is at the Broadcast Division of FOR-A Co. Ltd. He is developing digital video equipment for broadcasting. His research interests include image processing, computer vision, and computer graphics.

Kenichi Kanatani received his Ph.D. degree in applied mathematics from the University of Tokyo in 1979. He is currently professor of computer science at Gunma University. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford University Press, 1993), and *Statistical Optimization for Geometric Computation* (Elsevier, 1996).