Systems and Computers in Japan, Vol. 33, No. 9, 2002
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J84-D-II, No. 8, August 2001, pp. 1655–1662

3D Reconstruction from Uncalibrated-Camera Optical Flow and Its Reliability Evaluation

Kenichi Kanatani,[1] Naoya Ohta,[2] and Yoshiyuki Shimizu[3]

[1]Department of Information Technology, Okayama University, Okayama, 700-8530 Japan

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

[3]Audio-Visual Systems Group, Sharp, Ltd., Tochigi, 329-2193 Japan

SUMMARY

We present a scheme for reconstructing a 3D structure from optical flow observed by a camera with an unknown focal length in a statistically optimal way as well as evaluating the reliability of the computed shape. First, the flow fundamental matrices are optimally computed from the observed flow. They are then decomposed into the focal length, its rate of change, and the motion parameters. Next, the flow is optimally corrected so that it satisfies the epipolar equation exactly. Finally, the 3D positions are computed, and their covariance matrices are evaluated. By simulations and real-image experiments, we test the performance of our system and observe how the normalization (gauge) for removing indeterminacy affects the description of uncertainty. © 2002 Wiley Periodicals, Inc. Syst Comp Jpn, 33(9): 1–10, 2002; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.10133

Key words: optical flow; 3D reconstruction; self-calibration; reliability evaluation; gauge transformation.

Contract grant sponsors: Supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C(2) (No. 13680432), the Support Center for Advanced Telecommunications Technology Research, and the Kayamori Foundation of Informational Science Advancement.

1. Introduction

Traditionally, there exist two approaches for extracting 3D information from motion images: one is based on the "fundamental matrix" [14, 17] computed from point correspondences over two images; the other is to regard the displacements of points in the image as an instantaneous velocity field ("optical flow"), on which the 3D analysis is based. In this paper, we experimentally evaluate the performance of the latter.

Optical flow is usually computed from the gray levels of the images using the so-called *gradient constraint* [6, 12]. This computation is relatively easy, and the flow is obtained at all pixels. However, the accuracy is not very high, because derivatives are approximated by differences and arbitrary constraints for smoothness and regularization are imposed. The resulting flow may be useful for some applications, such as image segmentation, that do not require high accuracy. For accurate 3D reconstruction, however, we need to limit the number of feature points and compute their displacements with high accuracy. Here, we assume that individual feature points are accurately tracked using template matching or spatiotemporal filtering. Of course, our method can be applied to optical flow computed using the gradient constraint if high accuracy is not required.

© 2002 Wiley Periodicals, Inc.

Considerable research has been done in the past for 3D reconstruction from optical flow [6, 11, 13]. In most cases, however, the camera is assumed to be calibrated. Recently, significant progress has been made in the study of self-calibration of uncalibrated cameras using image sequences [14, 17], and Brooks and colleagues [2] and Viéville and Faugeras [16] presented a method for 3D reconstruction from optical flow observed by an uncalibrated camera.

In this paper, we follow Brooks's method [2] with a special emphasis on the following facts that have not received much attention in the past:

- While Brooks and colleagues [2] simply showed the possibility of 3D reconstruction, we give a statistically optimal reconstruction method that attains a theoretical accuracy by introducing a mathematical model of noise.
- Not only do we reconstruct the 3D shape but also evaluate the reliability of the solution to see how accurate the reconstructed shape is.
- We study how the normalization (gauge) for removing indeterminacy affects the description of uncertainty.

Optimal 3D reconstruction from optical flow observed by a calibrated camera has already been studied [13], and extension to the uncalibrated case is straightforward. However, the increase in the number of unknowns considerably complicates the reliability evaluation. In this paper, we introduce realistic first approximations to the reliability evaluation while optimizing the reconstruction computation exactly. Our system consists of pairs of computation and reliability evaluation at each stage as follows:

1. Detect optical flow at multiple (at least 8) pixels in the image.

2. Optimally compute the flow fundamental matrices from the detected flow, and evaluate their reliability.

3. Decompose the computed flow fundamental matrices into the focal length, its rate of change, and the camera motion parameters.

4. Optimally correct the optical flow so that it satisfies the epipolar equation exactly, and evaluate the reliability of the corrected flows.

5. Reconstruct the 3D positions from the corrected flow, and evaluate their reliability.

6. Evaluate the error in the flow fundamental matrices, and compute the covariance matrices of the reconstructed points.

The component techniques used in this process have been known in various forms [5–7, 15]. We integrate them in the most effective way for our purpose and confirm the effectiveness of our system by experiments using synthetic and real images.

2. Epipolar Equation and Flow Fundamental Matrices

Suppose the camera is translating with translation velocity $\boldsymbol{v}$ and rotating with rotation velocity $\boldsymbol{\omega}$ (around axis $\boldsymbol{\omega}$ with angular velocity $\|\boldsymbol{\omega}\|$) relative to the scene. We call $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ the *motion parameters*. Define an arbitrary coordinate system in the image, and let $(\dot{u}, \dot{v})$ be the displacement velocity (optical flow) at (u, v). We express them in vectors

$$\boldsymbol{x} = \begin{pmatrix} u/f_0 \\ v/f_0 \\ 1 \end{pmatrix}, \quad \dot{\boldsymbol{x}} = \begin{pmatrix} \dot{u}/f_0 \\ \dot{v}/f_0 \\ 0 \end{pmatrix} \tag{1}$$

where f_0 is an appropriate approximation to the camera focal length. If image noise does not exist, the following *epipolar equation* holds [2, 7, 16]:

$$(\boldsymbol{x}, \boldsymbol{W}\dot{\boldsymbol{x}}) + (\boldsymbol{x}, \boldsymbol{C}\boldsymbol{x}) = 0 \tag{2}$$

Throughout this paper, $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. In the above equation, $\boldsymbol{W}$ and $\boldsymbol{C}$ are antisymmetric and symmetric matrices, respectively. Since they correspond to the *fundamental matrix* [14, 17] that determines the epipolar equation for two images, we call $\{\boldsymbol{W}, \boldsymbol{C}\}$ the *flow fundamental matrices* [7]. Define the following vector from the elements of the antisymmetric matrix $\boldsymbol{W} = (W_{ij})$:

$$\boldsymbol{w} = \begin{pmatrix} W_{32} \\ W_{13} \\ W_{21} \end{pmatrix} \tag{3}$$

It can be shown [7] that matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ satisfy the *decomposability constraint*

$$(\boldsymbol{w}, \boldsymbol{C}\boldsymbol{w}) = 0 \tag{4}$$

which corresponds to the constraint that the fundamental matrix computed from two images have determinant zero [14, 17]. If the camera is calibrated, a stronger decomposability constraint is imposed [6, 13].

3. Computing Flow Fundamental Matrices

Brooks and colleagues [2] and Viéville and Faugeras [16] computed the flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ by a simple least-squares method, but the solution is known to have a large statistical bias [6]. In order to remove the bias, the authors presented an algorithm for optimally computing the flow fundamental matrices by introducing a statistical model of image noise and optimally fitting the epipolar equation (2) to the flow. This algorithm computes the

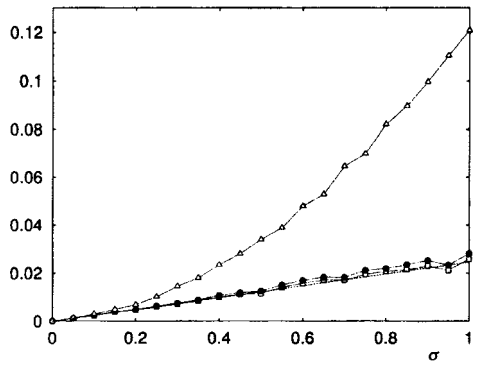


Fig. 1. Root-mean-square error of flow fundamental matrix computation. Δ: least squares; ●: renormalization; □: renormalization and optimal correction. The dotted lines indicate the theoretical lower bound.

matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ by a technique called *renormalization* [6] and optimally corrects the solution so that the decomposability constraint (4) is exactly satisfied. A C++ program package of the algorithm is publicly available.*

Figure 1 shows the root-mean-square error (corresponding to the standard deviation) of the flow fundamental matrices computed from the synthetic example given in Ref. 15. The abscissa is for the standard deviation of the random noise added to the data points. We did 100 trials using different noise for each standard deviation. The symbol Δ denotes solutions computed by the simple least-squares method of Refs. 2 and 16; ● denotes renormalization solutions; □ denotes renormalization solutions with optimal correction. The dotted line indicates the theoretical lower bound, which can be estimated in the course of computing the fundamental matrices [15]. We can see that our solution falls in the vicinity of the lower bound.

This bound allows us to evaluate the reliability of quantities derived from the flow fundamental matrices. Figure 2(a) shows the optical flow computed from a real image sequence by the method described in Ref. 12. The covariance matrix of the detected flow can be evaluated from the gray levels of the images in the course of detecting the optical flow [12]. Figure 2(b) shows the reliability of the epipole (the point in the direction of the camera motion) estimated from the optimally computed flow fundamental matrices; the ellipse shows the standard deviation from the estimated epipole in each orientation.

Our method computes not only optimal estimate $\{\hat{\boldsymbol{C}}, \hat{\boldsymbol{W}}\}$ of the flow fundamental matrices but also their primary deviation pairs $\{\boldsymbol{C}^{(+)}, \boldsymbol{W}^{(+)}\}$ and $\{\boldsymbol{C}^{(-)}, \boldsymbol{W}^{(-)}\}$. They are the values in the parameter space that are separated from

*http://www.ail.cs.gunma-u.ac.jp/Labo/e-programs.html

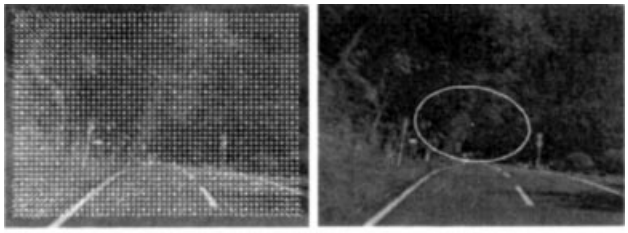


Fig. 2. (a) Optical flow computed from road scene images. (b) The estimated epipole and its uncertainty evaluation.

$\{\hat{\boldsymbol{C}}, \hat{\boldsymbol{W}}\}$ by the standard deviation in the two directions along which errors are most likely to occur; they correspond to the theoretical accuracy bound. If $\boldsymbol{C}^{(+)}$ and $\boldsymbol{C}^{(-)}$ agree up to, say, three significant digits, the solution $\hat{\boldsymbol{C}}$ is guaranteed to have accuracy up to approximately three significant digits [6, 15]. This information can be used to evaluate the reliability of the reconstructed 3D shape (see Section 9).

4. Optimal Correction of Optical Flow

Let $V_0[\boldsymbol{x}]$ and $V_0[\dot{\boldsymbol{x}}]$ be the *normalized covariance matrices* of the feature position $\boldsymbol{x}$ and the flow $\dot{\boldsymbol{x}}$ there; they are determined up to a scale factor, describing the qualitative characteristics of their accuracy. They are computed from the gray levels of the images if the feature points are detected by an image processing operation [10].

If $(\dot{u}, \dot{v})$ and (u, v) in Eqs. (1) are interpreted to be, respectively, the displacement vector connecting the corresponding points over the two images and their midpoint, and if the noise is assumed independent and isotropic (i.e., the standard deviation is independent of orientations) at each point, the covariance matrices of $\dot{\boldsymbol{x}}$ and $\boldsymbol{x}$ are written, up to a scale factor, in the form

$$V_0[\dot{\boldsymbol{x}}] = 2\mathrm{diag}(1,1,0), \quad V_0[\boldsymbol{x}] = \frac{1}{2}\mathrm{diag}(1,1,0) \tag{5}$$

where diag(. . .) denotes a diagonal matrix with diagonal elements . . . . If no particular information is available about the accuracy of the feature points, we use these as the default values of the covariance matrices.

Even though the flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ are optimally computed, $\dot{\boldsymbol{x}}$ and $\boldsymbol{x}$ do not necessarily satisfy the epipolar equation (2) exactly. So, we correct $\dot{\boldsymbol{x}}$ and $\boldsymbol{x}$ so that Eq. (2) is exactly satisfied. This is done as follows [6]:

$$\hat{\dot{\boldsymbol{x}}} = \dot{\boldsymbol{x}} + \frac{E(\dot{\boldsymbol{x}}, \boldsymbol{x})}{V(\dot{\boldsymbol{x}}, \boldsymbol{x})} V_0[\dot{\boldsymbol{x}}]\boldsymbol{W}\boldsymbol{x}$$

$$\hat{\boldsymbol{x}} = \boldsymbol{x} - \frac{E(\dot{\boldsymbol{x}}, \boldsymbol{x})}{V(\dot{\boldsymbol{x}}, \boldsymbol{x})} V_0[\boldsymbol{x}](\boldsymbol{W}\dot{\boldsymbol{x}} + 2\boldsymbol{C}\boldsymbol{x}) \tag{6}$$

$$E(\dot{\boldsymbol{x}}, \boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{W}\dot{\boldsymbol{x}}) + (\boldsymbol{x}, \boldsymbol{C}\boldsymbol{x})$$

$$V(\dot{\boldsymbol{x}}, \boldsymbol{x}) = (\boldsymbol{W}\boldsymbol{x}, V_0[\dot{\boldsymbol{x}}]\boldsymbol{W}\boldsymbol{x}) + (\boldsymbol{W}\dot{\boldsymbol{x}} + 2\boldsymbol{C}\boldsymbol{x}, V_0[\boldsymbol{x}](\boldsymbol{W}\dot{\boldsymbol{x}} + 2\boldsymbol{C}\boldsymbol{x})) \tag{7}$$

Equations (6) are iterated in the form $\boldsymbol{x} \leftarrow \hat{\boldsymbol{x}}$ and $\boldsymbol{x} \leftarrow \hat{\boldsymbol{x}}$ until $E(\hat{\boldsymbol{x}}, \hat{\boldsymbol{x}}) = 0$ is sufficiently satisfied. Since the convergence is quadratic as Newton iterations, one iteration is almost sufficient in practice.

Because the corrected values $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}$ should satisfy the epipolar equation, the degrees of freedom of their (normalized) covariance matrices are constrained to lower ranks. Thus, we replace them by the following (normalized) *a posterior covariance matrices* [6]:

$$V_0[\hat{\boldsymbol{x}}] = V_0[\dot{\boldsymbol{x}}] - \frac{(V_0[\dot{\boldsymbol{x}}]\boldsymbol{W}\boldsymbol{x})(V_0[\dot{\boldsymbol{x}}]\boldsymbol{W}\boldsymbol{x})^\top}{V(\dot{\boldsymbol{x}}, \boldsymbol{x})}$$

$$V_0[\hat{\boldsymbol{x}}] = V_0[\boldsymbol{x}] - \frac{V_0[\boldsymbol{x}](\boldsymbol{W}\dot{\boldsymbol{x}}+2\boldsymbol{C}\boldsymbol{x})(V_0[\boldsymbol{x}](\boldsymbol{W}\dot{\boldsymbol{x}}+2\boldsymbol{C}\boldsymbol{x}))^\top}{V(\dot{\boldsymbol{x}}, \boldsymbol{x})} \tag{8}$$

Although the noise in the position $\boldsymbol{x}$ and the noise in the flow $\dot{\boldsymbol{x}}$ are independent by assumption, the corrected values $\hat{\boldsymbol{x}}$ are $\hat{\hat{\boldsymbol{x}}}$ are no longer independent. They have the following (normalized) *correlation matrix* [6]:

$$V_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}] = \frac{V_0[\boldsymbol{x}](\boldsymbol{W}\dot{\boldsymbol{x}}+2\boldsymbol{C}\boldsymbol{x})(V_0[\dot{\boldsymbol{x}}]\boldsymbol{W}\boldsymbol{x})^\top}{V(\dot{\boldsymbol{x}}, \boldsymbol{x})} \tag{9}$$

The amount of the optimal correction described above is usually subpixel, but it significantly affects the accuracy of the reconstructed 3D points that are far from the camera. Also, we can evaluate the reliability of the reconstructed shape from the amount of this correction (see Section 8).

5. Decomposition of Flow Fundamental Matrices

The flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ are defined up to a scale factor and constrained by the decomposability constraint (4). In addition, $\boldsymbol{C}$ is symmetric and $\boldsymbol{W}$ is antisymmetric. Thus, they have seven degrees of freedom. The absolute scale of the translation velocity $\boldsymbol{v}$ is indeterminate due to the scale indeterminacy of $\{\boldsymbol{C}, \boldsymbol{W}\}$. Hence, the motion parameters $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ have five degrees of freedom. It follows that at most two camera parameters can be determined if the camera motion is unconstrained.

A realistic choice for the two parameters is the focal length f and its rate of change $\dot{f}$, because other intrinsic parameters can be calibrated in advance. With today's high-quality cameras, we will not have serious problems if we assume the standard values, that is, assume that the principal point (the point where the optical axis passes through) is at the center of the frame, the aspect ratio (the ratio of vertical to horizontal sizes of the pixel array) is 1, and the skew angle (the angle made between the rows and columns of the photo cell array) is 90°. However, the focal length (zooming) frequently changes in the course of shooting.

Assuming the standard values for all the parameters except the focal length f and its rate of change $\dot{f}$, Brooks and colleagues [2] presented a complicated algebraic procedure for computing f and $\dot{f}$. However they can be obtained by a simple computation in terms of complex numbers based on the group representation theory [3, 4] as follows [7]:

1. Compute the following quantities from the matrix $\boldsymbol{C} = (C_{ij})$:

$$A = C_{11} + C_{22}, \qquad \tilde{B} = (C_{11} - C_{22}) + 2iC_{12}$$
$$\tilde{C} = 2(C_{13} + iC_{23}), \qquad D = C_{33} \tag{10}$$

2. Compute the vector $\boldsymbol{w} = (w_i)$ defined in Eq. (3) from the matrix $\boldsymbol{W}$.

3. Compute the complex numbers $\tilde{w}$ and $\tilde{\omega}$ as follows:

$$\tilde{w} = w_1 + iw_2, \qquad \tilde{\omega}' = \frac{\tilde{B}}{\tilde{w}} \tag{11}$$

4. Compute ω_1' and ω_2' as follows:

$$\omega_1' = \Re[\tilde{\omega}'], \qquad \omega_2' = \Im[\tilde{\omega}'] \tag{12}$$

5. Compute ω_3 and f' as follows:

$$\omega_3 = -\frac{A + (\tilde{w}, \tilde{\omega}')}{2w_3}, \qquad f' = \sqrt{-\frac{D}{(\tilde{w}, \tilde{\omega}')}} \tag{13}$$

6. Compute the complex number $\tilde{\phi}$ as follows:

$$\tilde{\phi} = \frac{\tilde{C} - f'^2 w_3 \tilde{\omega}'}{\tilde{w}} \tag{14}$$

7. Compute ω_3 and $\dot{f}'$ as follows:

$$\omega_3 = \Re[\tilde{\phi}], \qquad \dot{f}' = -f'\Im[\tilde{\phi}] \tag{15}$$

8. Compute ω_1 and ω_2 as follows:

$$\omega_1 = f'\omega_1', \qquad \omega_2 = f'\omega_2' \tag{16}$$

9. Compute the focal length f and its rate of change $\dot{f}$ as follows:

$$f = f' f_0, \qquad \dot{f} = \dot{f}' f_0 \tag{17}$$

10. Compute the translation velocity $\boldsymbol{v}$ as follows:

$$\boldsymbol{v} = N\left[\begin{pmatrix} w_1 \\ w_2 \\ (f/f_0)w_3 \end{pmatrix}\right] \tag{18}$$

In the above equations, i denotes the imaginary unit, and the quantities with tildes are complex numbers. The real and imaginary parts of a complex number are denoted by $\Re[\cdot]$ and $\Im[\cdot]$, respectively, and the "inner product" of complex numbers $z = x + iy$ and $z' = x' + iy'$ is defined to be $(z, z') = xx' + yy'$. The operation $N[\cdot]$ designates normalization to a unit vector ($N[\boldsymbol{a}] = \boldsymbol{a}/\|\boldsymbol{a}\|$). Note that ω_3 is obtained in two ways by the first of Eqs. (13) and the first of Eqs. (15). This is a consequence of the decomposability constraint (4): the two values do not agree if Eq. (4) is not satisfied. This fact can be used for error checking. The solution is uniquely determined if the camera motion is such that the optical axis moves into a skewed position [7].

6. Changing the Focal Length

Now that the focal length f and its rate of change $\dot{f}$ are determined, we modify $\hat{\hat{\boldsymbol{x}}}$ and $\hat{\boldsymbol{x}}$ in the form

$$\hat{\hat{\boldsymbol{x}}} \leftarrow \frac{f_0}{f}\left(\hat{\hat{\boldsymbol{x}}} - \frac{\dot{f}}{f}\boldsymbol{P}\hat{\boldsymbol{x}}\right), \quad \hat{\boldsymbol{x}} \leftarrow \mathrm{diag}\left(\frac{f_0}{f}, \frac{f_0}{f}, 1\right)\hat{\boldsymbol{x}} \tag{19}$$

where $\boldsymbol{P} = \mathrm{diag}(1, 1, 0)$. The second equation is to replace the approximate focal length f_0 by the true value f. The first equation is to subtract the apparent flow due to zoom change from the observed flow. The modified values $\hat{\boldsymbol{x}}$ and $\hat{\hat{\boldsymbol{x}}}$ can be respectively interpreted to be the orientation of the line of sight starting from the center of the lens and its rate of change. Their (normalized a posteriori) covariance matrices $V_0[\hat{\boldsymbol{x}}]$ and $V_0[\hat{\hat{\boldsymbol{x}}}]$ and (normalized) correlation matrix $V_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}]$ are now replaced in the form

$$\begin{aligned} V_0[\hat{\hat{\boldsymbol{x}}}] &\leftarrow \frac{f_0^2}{f^2}\left(V_0[\hat{\hat{\boldsymbol{x}}}] - \frac{2\dot{f}}{f}S[V_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}]] + \frac{\dot{f}^2}{f^2}V_0[\hat{\boldsymbol{x}}]\right) \\ V_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}] &\leftarrow \frac{f_0^2}{f^2}\left(V_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}] - \frac{\dot{f}}{f}V_0[\hat{\boldsymbol{x}}]\right) \\ V_0[\hat{\boldsymbol{x}}] &\leftarrow \frac{f_0^2}{f^2}V_0[\hat{\boldsymbol{x}}] \end{aligned} \tag{20}$$

where $S[\cdot]$ designates the symmetrization operation ($S[\boldsymbol{A}] = (\boldsymbol{A} + \boldsymbol{A}^\top)/2$).

7. Depth Computation

The position $\boldsymbol{r}$ in the scene of the feature point corresponding to the vector $\boldsymbol{x}$ is expressed in the form

$$\boldsymbol{r} = Z\boldsymbol{x} \tag{21}$$

where Z is the depth from the camera coordinate origin (the center of the lens) measured along the optical axis. It is computed as follows [6, 13]:

$$Z = -\frac{(\boldsymbol{v}, \boldsymbol{S}_{\boldsymbol{X}}\boldsymbol{v})}{(\boldsymbol{v}, \boldsymbol{S}_{\boldsymbol{X}}(\dot{\boldsymbol{x}} + \boldsymbol{\omega} \times \boldsymbol{x}))} \tag{22}$$

Here, we define $\boldsymbol{Q}_{\boldsymbol{X}}$ and $\boldsymbol{S}_{\boldsymbol{X}}$ as follows ($\boldsymbol{I}$ denotes the unit matrix, and we let $\boldsymbol{k} = (0, 0, 1)^\top$):

$$\boldsymbol{Q}_{\boldsymbol{X}} = \boldsymbol{I} - \boldsymbol{x}\boldsymbol{k}^\top, \qquad \boldsymbol{S}_{\boldsymbol{X}} = \boldsymbol{Q}_{\boldsymbol{X}}^\top \boldsymbol{Q}_{\boldsymbol{X}} \tag{23}$$

At this stage, we need to choose the signs of the depth. Since the flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ are determined only up to a scale factor, their signs are indeterminate. This indeterminacy stems from the fact that the perspective projection has the same mathematical form if the scene is "behind" the camera. Thus, we compute the depths $\hat{Z}_\alpha$ of the points $\hat{\boldsymbol{x}}_\alpha$, $\alpha = 1, \ldots, N$, and change the signs of $\hat{Z}_\alpha$ and $\boldsymbol{v}$ if

$$\sum_{\alpha=1}^{N} \mathrm{sgn}[\hat{Z}_\alpha] < 0 \tag{24}$$

where $\mathrm{sgn}[\cdot]$ designates the sign function taking 1, 0, or –1 for $x > 0$, $x = 0$, or $x < 0$, respectively. This operation is necessary because we may not select the correct solution if we simply compute $\Sigma_{\alpha=1}^{N} \hat{Z}_\alpha$; a very large positive depth may turn out to be close to $-\infty$ due to noise [6].

8. Reliability Evaluation of 3D Reconstruction 1

Given the estimated depth $\hat{Z}$, the corresponding 3D position $\hat{\boldsymbol{r}}$ is determined by Eq. (21). Its normalized covariance matrix has the following form [6, 13]:

$$V_0[\hat{\boldsymbol{r}}] = \hat{Z}^2 V_0[\hat{\boldsymbol{x}}] + 2\hat{Z}S[V_0[\hat{Z}, \hat{\boldsymbol{x}}]\hat{\boldsymbol{x}}^\top] + V_0[\hat{Z}]\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^\top \tag{25}$$

The (normalized a posteriori) covariance matrix $V_0[\hat{\boldsymbol{x}}]$ in the right-hand side expression is given by the first of Eqs. (20). The normalized variance $V_0[\hat{Z}]$ and the normalized correlation vector $V_0[\hat{Z}, \hat{\boldsymbol{x}}]$ are obtained from Eq. (22) as follows [6, 13]:

$$\begin{aligned} V_0[\hat{Z}] = \frac{\hat{Z}^4}{(\boldsymbol{v}, \boldsymbol{S}_{\hat{\boldsymbol{X}}}\boldsymbol{v})}\mathrm{tr}\Big(&V_0[\hat{\hat{\boldsymbol{x}}}] + 2S[\boldsymbol{Q}_{\hat{\boldsymbol{X}}}(\boldsymbol{\omega} \times V_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}])] \\ &- aV_0[\hat{\boldsymbol{x}}, \hat{\hat{\boldsymbol{x}}}] + \boldsymbol{Q}_{\hat{\boldsymbol{X}}}(\boldsymbol{\omega} \times V_0[\hat{\boldsymbol{x}}] \times \boldsymbol{\omega})\boldsymbol{Q}_{\hat{\boldsymbol{X}}}^\top \\ &- 2aS[\boldsymbol{Q}_{\hat{\boldsymbol{X}}}(\boldsymbol{\omega} \times V_0[\hat{\boldsymbol{x}}])] + a^2 V_0[\hat{\boldsymbol{x}}]\Big) \end{aligned} \tag{26}$$

$$V_0[\hat{Z}, \hat{\boldsymbol{x}}] = \frac{\hat{Z}^2}{(\boldsymbol{v}, \boldsymbol{S}_{\hat{\boldsymbol{x}}}\boldsymbol{v})}\Big(V_0[\hat{\boldsymbol{x}}, \hat{\dot{\boldsymbol{x}}}]\boldsymbol{Q}_{\hat{\boldsymbol{x}}}\boldsymbol{v} - V_0[\hat{\boldsymbol{x}}](\boldsymbol{\omega} \times \boldsymbol{S}_{\hat{\boldsymbol{x}}}\boldsymbol{v} + a\boldsymbol{Q}_{\hat{\boldsymbol{x}}}\boldsymbol{v})\Big) \quad (27)$$

Here, tr designates the trace. We also define

$$a = \frac{(\boldsymbol{k}, \boldsymbol{v})}{\hat{Z}} + |\boldsymbol{k}, \boldsymbol{\omega}, \hat{\boldsymbol{x}}| \quad (28)$$

where $|\boldsymbol{k}, \boldsymbol{\omega}, \hat{\boldsymbol{x}}|$ denotes the scalar triple product of $\boldsymbol{k}$, $\boldsymbol{\omega}$, and $\hat{\boldsymbol{x}}$.

9. Reliability Evaluation of 3D Reconstruction 2

The reliability analysis so far has been done on the assumption that the computed flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ are correct: we optimally corrected the flow by Eqs. (6) using $\{\boldsymbol{C}, \boldsymbol{W}\}$, applied the transformation of Eqs. (19) using the values f and $\dot{f}$ resulting from $\{\boldsymbol{v}, \boldsymbol{\omega}\}$, and computed the depths from the motion parameters $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ obtained from $\{\boldsymbol{C}, \boldsymbol{W}\}$. The normalized covariance matrix $V_0[\hat{\boldsymbol{r}}]$ of Eq. (25) describes how the reconstructed position $\hat{\boldsymbol{r}}$ is affected by the errors in $\hat{\boldsymbol{x}}$ and $\hat{\dot{\boldsymbol{x}}}$.

By "normalized" we mean the use of a scale such that the noise level ε is 1; the absolute value of the covariance matrix $V[\hat{\boldsymbol{r}}]$ is given by $\varepsilon^2 V_0[\hat{\boldsymbol{r}}]$. The absolute magnitude of ε is automatically estimated in the course of computing $\{\boldsymbol{C}, \boldsymbol{W}\}$ by the renormalization method [15].

However, the flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ also have errors because they are computed from noisy data. In fact, we can evaluate their accuracy in terms of their covariance tensor [15], from which we could, in principle, analyze the errors in f, $\dot{f}$, and $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ as well as their correlations. However, this would entail a very complicated procedure. In order to avoid this, we adopt the following compromise.

As stated in Section 3, the renormalization computation of the flow fundamental matrices also produces their primary deviation pairs $\boldsymbol{C}^{(\pm)}$ and $\boldsymbol{W}^{(\pm)}$. From them, we compute the corresponding $f^{(\pm)}$, $\dot{f}^{(\pm)}$, $\boldsymbol{v}^{(\pm)}$, and $\boldsymbol{\omega}^{(\pm)}$. Let the resulting 3D positions be $\boldsymbol{r}^{(\pm)}$. We approximate the effect of errors in the flow fundamental matrices by the line segment connecting the two points $\boldsymbol{r}^{(+)}$ and $\boldsymbol{r}^{(-)}$. Since the reconstructed position $\hat{\boldsymbol{r}}$ is the midpoint of it to a first approximation, we can regard $(\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})(\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})^\top$ as its covariance matrix (we only need to compute $\boldsymbol{r}^{(+)}$).

Equation (25) describes how the reconstructed position $\hat{\boldsymbol{r}}$ is affected by the errors in the corresponding point $\boldsymbol{x}$ and the flow $\dot{\boldsymbol{x}}$ there. This relationship is direct. In contrast, the flow fundamental matrices $\{\boldsymbol{C}, \boldsymbol{W}\}$ are optimized over all the points and all the flow vectors. Since this relationship is indirect, the correlation between $\{\boldsymbol{C}, \boldsymbol{W}\}$ and individual points is expected to be small. Thus, we evaluate the covariance matrix of the reconstructed position, to a first approximation, as the sum of the covariance matrices obtained separately, taking into account one effect only for each:

$$V[\hat{\boldsymbol{r}}] = \hat{\epsilon}^2 V_0[\hat{\boldsymbol{r}}] + (\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})(\boldsymbol{r}^{(+)} - \hat{\boldsymbol{r}})^\top \quad (29)$$

The omitted terms are of the order of their product or higher. If the error distribution is approximated by a Gaussian distribution, the deviations from the reconstructed position $\hat{\boldsymbol{r}}$ by less than the standard deviation in each orientation are inside the following ellipsoid, which we call the *standard region* [6]:

$$(\boldsymbol{r} - \hat{\boldsymbol{r}}, V[\hat{\boldsymbol{r}}]^{-1}(\boldsymbol{r} - \hat{\boldsymbol{r}})) = 1 \quad (30)$$

It has been confirmed by simulations of many geometric fitting problems that this type of evaluation can well approximate the underlying uncertainty [6].

10. Simulation Experiments

Figure 3 shows simulated 512×512-pixel images of a 3D grid environment. We added random Gaussian noise of mean 0 and standard deviation 0.5 (pixel) to each of the x and y coordinates of the grid points and reconstructed the 3D shape by regarding the displacements of points as the flow vectors at their midpoints and using the default noise model of Eq. (5).

Figure 4(a) shows the reconstructed shape (solid lines) superimposed on the true shape (dotted lines) rescaled to $\|\boldsymbol{t}\| = 1$ viewed from an angle. Figure 4(b) shows three times the standard regions defined by Eq. (30) centered on the reconstructed vertices. They look like thin needles, meaning that the uncertainty is large in the depth orientation. We also observe that the points farther from the camera have larger uncertainty. Comparing this with Fig.

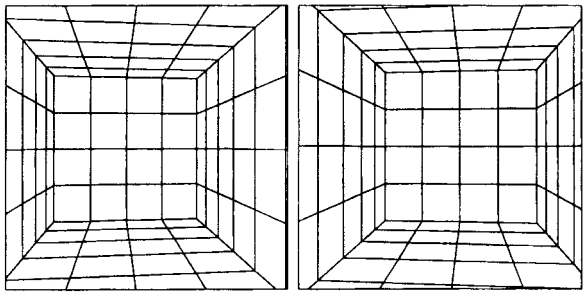
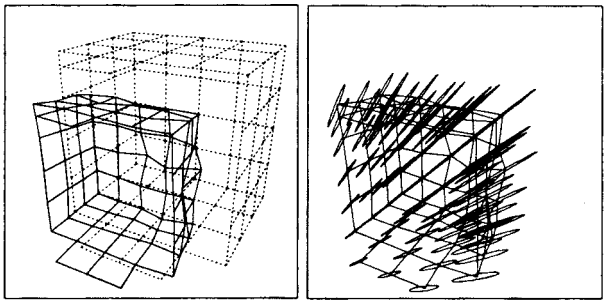


Fig. 3. Simulated images of a 3D scene.



(a) (b)

Fig. 4. (a) Reconstructed shape (solid lines) and the true shape (broken lines). (b) The standard regions of the grid points.

4(a), we see that these ellipsoids approximately describe the magnitude of the deviations from their true positions.

However, they are an underestimation if viewed as representing the covariance matrix of a probability distribution. We repeated the experiment many times using different noise, and we always observed underestimation to this degree. This is probably because we identified finite displacements with an instantaneous velocity field, omitted higher-order terms, and approximated the variance of the flow fundamental matrices by their primary deviation pairs. Rigorous error analysis is generally a very difficult problem, but this type of gross evaluation can be of sufficient help for seeing the general tendency of the errors and comparing the accuracy in relative terms.

11. Real-Image Experiments

Figure 5 shows a pair of real images (512 × 768 pixels) of an indoor scene. We manually selected feature points as marked in the images and reconstructed the 3D shape by regarding the displacements of points as the flow vectors at their midpoints and using the default noise model of Eq. (5) (we assumed the standard values for the camera parameters other than the focal length). Figure 6 shows a

Fig. 5. Real images of an indoor scene.

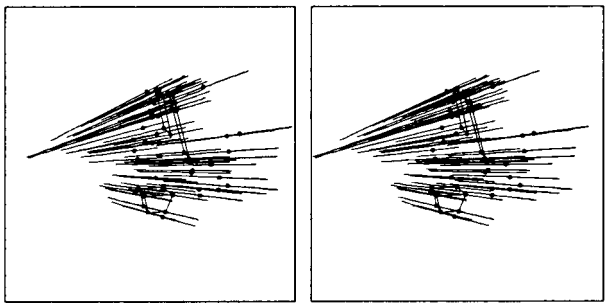
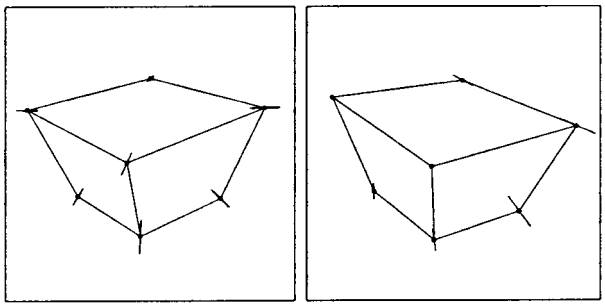


Fig. 6. Reconstructed points and their standard regions (stereogram).

side view of the reconstructed points as a stereogram. On each point is centered the standard region defined by Eq. (30). Wireframes consisting of some of the reconstructed points are shown for visual aid.

Seeing this, we have an impression that the reconstructed shape has large uncertainty in the depth orientation. However, this is due to the uncertainty of the camera translation velocity; the shape itself is not so very uncertain. To see this, we picked out the polyhedral object in the scene and displayed its shape after placing the centroid of the object at the coordinate origin and scaling the root-mean-square distance of the vertices from the origin to unity [Fig. 7(a)]. We see that the standard regions are very small compared with Fig. 6. Alternatively, we can place one of the vertices at the coordinate origin, another at (1, 0, 0), and yet another on the *XY* plane [Fig. 7(b)]. The points placed at the origin and (1, 0, 0) have no uncertainty by definition, so their standard regions are nil.

Thus, even if the reconstructed shape itself is the same, the reliability evaluation changes depending on the normalization (or *gauge*). According to the *gauge theory*



(a) (b)

Fig. 7. Standard regions of the vertices.
(a) Normalization of the centroid and the size.
(b) Normalization of the three vertices.

Table 1. Reliability of the ratio of edge lengths and the angle

	computed value	measured value	theoretical standard dev.
ratio	1.02	1.00	0.08
angle (deg)	95.1	90.0	17.0

[9], which systematically describes this effect, the description of uncertainty of an indeterminate system cannot be given absolute meanings because it depends on the normalization involved. Absolute meanings can be given only to the uncertainty of *gauge invariants* that are not affected by the change of normalization (or *gauge transformation*). Typical gauge invariants include ratios of lengths and angles between lines.

We actually measured the size of the polyhedral object with a rule and the angles at the vertices with a protractor. Table 1 lists the ratio of the length of one of the edges near the viewer to that of the other and the angle they make; we compare the values computed from the reconstructed shape, the measurement values, and the theoretical standard deviations predicted by Eq. (29). The description of uncertainty is meaningful only for this type of gauge invariants in evaluating the reliability of 3D reconstruction.

The theoretical standard deviations given in Table 1 look like somewhat overestimating the deviations from the measurement values. As in Fig. 4, this is probably because we identified the finite displacement with an instantaneous motion, omitted higher-order terms, and approximated the variance of the flow fundamental matrices by their primary deviation pairs. There is also a possibility that the physical positions supposedly representing the reconstructed points may not exactly coincide with those in the scene. Again, rigorous analysis is very difficult, and this type of rough evaluation can be of sufficient help for grasping the general tendency.

Figure 8 shows real images (512×768 pixels) of a car. We manually selected feature points as marked in the images and reconstructed the 3D shape, using the default noise model of Eq. (5) (we assumed the standard values for the camera parameters other than the focal length). Figure 9 shows new views generated by creating a wireframe model from the reconstructed points and mapping the texture to it. We see that the 3D shape is not very accurate in the part far from the viewer compared with the front part, which is fairly accurate.

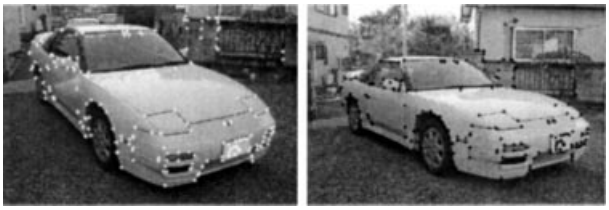


Fig. 8. Real images of a car.

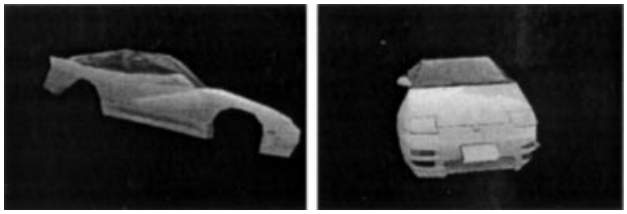


Fig. 9. 3D reconstruction of the car.

12. Concluding Remarks

We have presented a scheme for reconstructing the 3D shape in a statistically optimal way from optical flow observed by a camera with an unknown focal length as well as evaluating the reliability of the reconstructed shape to a first approximation. We have tested the performance of our system by doing experiments using synthetic and real images. We have also studied how the normalization (or gauge) for removing indeterminacy affects the description of uncertainty.

Doing comparative experiments using the same image data, we have found that the accuracy of our system is somewhat inferior to the method based on the fundamental matrix [8]. Baumela and colleagues [1] also obtained a similar conclusion. It follows that the method based on the fundamental matrix should be used if the 3D shape reconstruction is the sole purpose. Optical flow, on the other hand, has many applications including image segmentation and moving object detection, and our system may be useful in acquiring auxiliary 3D information in such applications.

Acknowledgments. The authors thank Mike Brooks, Wojciech Chojnacki, and Anton van den Hengel of the University of Adelaide, Australia, for collaboration in this research. They also thank Luis Baumela of Madrid Technical University, Spain, Lourdes de Agapito of the University of Oxford, U.K., and Toshiharu Mukai of Riken Institute, Japan, for helpful discussions. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C(2) (No. 13680432), the Support Center for Advanced Telecommunications Technology Research, and the Kayamori Foundation of Informational Science Advancement.

REFERENCES

1. Baumela L, de Agapito L, Bustos P, Reid I. Motion estimation using the differential epipolar equation. Proc 15th Int Conf Patt Recog, 2000, Barcelona, Spain, Vol. 3, p 848–851.
2. Brooks MJ, Chojnacki W, Baumera L. Determining the egomotion of an uncalibrated camera from instantaneous optical flow. J Opt Soc Am 1997;A-14:2670–2677.
3. Kanatani K. Coordinate rotation invariance of image characteristics for 3D shape and motion recovery. Trans IEICE 1987;J70-D:937–945. (in Japanese)
4. Kanatani K. Group-theoretical methods in image understanding. Springer; 1990.
5. Kanatani K. Geometric computation for machine vision. Oxford University Press; 1993.
6. Kanatani K. Statistical optimization for geometric computation: Theory and practice. Elsevier Science; 1996.
7. Kanatani K, Brooks MJ. Decomposition of the flow fundamental matrices: Group-theoretical method for self-calibration. IPSJ SIG Notes, 2000-CVIM-120-8, p 57–64, 2000. (in Japanese)
8. Kanatani K, Mishima H. 3-D reconstruction from two uncalibrated views and its reliability evaluation. Trans IPSJ: CVIM 2001;42:1–8. (in Japanese)
9. Kanatani K, Morris DD. Gauges and gauge transformations for uncertainty description of geometric structure with indeterminacy. IEEE Trans Inf Theory 2001;47:2017–2928.
10. Kanazawa Y, Kanatani K. Do we really have to consider covariance matrices for image features? IPSJ SIG Notes, 2001-CVIM-126-1, p 1–8, 2001. (in Japanese)
11. Mukai T, Ohnishi N. Recovery of three-dimensional motion parameters and structure from an optical flow image by linear calculation. Trans SICE 1998;34:438–444. (in Japanese)
12. Ohta N. Image movement detection with reliability indices. Trans IEICE 1991;E74:3379–3388.
13. Ohta N, Kanatani K. Optimal structure-from-motion algorithm for optical flow. Trans IEICE 1995;E78-D:1559–1566.
14. Sato J. Computer vision: Geometry of visual perception. Corona Publishing; 1999. (in Japanese)
15. Shimizu Y, Kanatani K. Optimal computation of the optical flow fundamental matrix and its reliability evaluation. IPSJ SIG Notes, 99-CVIM-118-11, p 75–82, 1999. (in Japanese)
16. Viéville T, Faugeras OD. The first order expansion of motion equations in the uncalibrated case. Comput Vision Image Understand 1996;64:128–146.
17. Xu G, Tsuji S. 3D vision. Kyoritsu Shuppan; 1998. (in Japanese)

AUTHORS

Kenichi Kanatani received his M.Sc. and Ph.D. degrees in applied mathematics from the University of Tokyo in 1974 and 1979. After serving as a professor of computer science at Gunma University, he is currently a professor of information technology at Okayama University. He was a visiting researcher at the University of Maryland, the University of Copenhagen, the University of Oxford, and INRIA. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford, 1993), and *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier, 1996). His research interests include mathematical analysis of image processing and computer vision systems. He is an IEEE Fellow.

AUTHORS (continued) (from left to right)

Naoya Ohta received his M.Sc. degree in information science from Tokyo Institute of Technology in 1985 and his Ph.D. degree in applied mathematics from the University of Tokyo in 1998. He engaged in research and development of image processing systems at the Pattern Recognition Research Laboratories of NEC, Japan. He was a visiting researcher at the Media Laboratory and the University of Adelaide. He is currently an associate professor of computer science at Gunma University. His research interests include image processing and computer vision technologies.

Yoshiyuki Shimizu received his M.Sc. degree in computer science from Gunma University in 2000. He is currently engaged in research and development of liquid-crystal mobile systems in the Audio-Visual Systems Group. Sharp, Ltd.