Memoirs of the Faculty of Engineering, Okayama University, Vol.36, No.1, pp.91-106, December 2001

Comparing Optimal 3-D Reconstruction for Finite Motion and Optical Flow

Kenichi KANATANI*
Department of Information Technology
Okayama University
Okayama 700-8530 Japan

Naoya OHTA
Department of Computer Science
Gunma University
Kiryu, Gunma 376-8515 Japan

(Received November 9, 2001)

We present a theoretically optimal linear algorithm for 3-D reconstruction from point correspondences over two views. We also present a similarly constructed optimal linear algorithm for 3-D reconstruction from optical flow. We then compare the performance of the two algorithms by simulation and real-image experiments using the same data. This is the first impartial comparison ever done in the sense that the two algorithms are both optimal, extracting the information contained in the data to a maximum possible degree. We observe that the finite motion solution is always superior to the optical flow solution and conclude that the finite motion algorithm should be used for 3-D reconstruction.

1. INTRODUCTION

Computing the 3-D structure and motion from an image sequence is one of the most important of computer vision tasks and also one of the research areas that are making the rapidest progress today. Already, a vast literature has appeared on this subject (see, e.g., [8, 13] for the latest developments). There have been two approaches to this problem since 1970s when 3-D reconstruction from images began to attract interest: that based on point correspondences over different views, which we call the *finite motion approach*; that based on instantaneous image motion, which we call the *optical flow approach*. Early literature includes [21, 31, 32] for the former and [4, 9, 22] for the latter. This division has lasted to date because of different mathematical and technical disciplines involved.

Mathematically, the finite motion approach is based on vector calculus of triangulation (or *epipolar geometry*); the optical flow approach is based on differential calculus of the gray levels (or the *gradient constraint*). Technically, the former uses template matching for feature correspondence detection; the latter uses filter operations with smoothness constraints for optical flow detection. The latter has been closely associated with human perception psychology (see [9, 22]).

This division, however, cannot not be crucial if we note that all the computation is based on images. Although optical flow is mathematically modeled as a velocity field, all that can actually be computed is the displacements over consecutive frames. It follows that once optical flow (= a set of point correspondences) is obtained, one could run a finite motion algorithm for reconstructing the 3-D structure and motion. The finite motion approach can apply to any motion, large or small, while the optical flow approach can apply only to a small motion. Then, what is the benefit of the optical flow approach?

This question has not been fully answered yet, partly because researchers in the past favored one approach over the other, demonstrating the merits of their own approaches, and partly because the techniques used were often tuned to the image data used, making it difficult to compare the two approaches for the same data. But more

*E-mail kanatani@suri.it.okayama-u.ac.jp

fundamental is the difficulty of finding a common ground for fair comparison; either of the approaches works very well if properly implemented, the performance depending largely on implementation.

The aim of this paper is to make a fair comparison of the two approaches from a *theoretical standpoint*. Introducing a statistical model of uncertainty for feature locations, we present an optimal algorithm for both finite motion and optical flow; by "optimal", we mean that the accuracy already reaches a theoretical bound and hence *no other algorithms could possibly outperform it*. The optimization techniques we use are already known but have been published only in fragments [15, 16, 17], so we give a self-consistent description of our algorithm for finite motion. For optical flow, the major part is described in our previous paper [19]. Here, we focus on only those parts that are not given there.

We then compare the performance of the two algorithms by simulation and real-image experiments using the same data. This is the first impartial comparison ever done in the sense that the two algorithms are both optimal, extracting the information contained in the data to a maximum possible degree. Although finite motion analysis can be extended to multiple images [8, 13], we concentrate on two-view analysis for the sake of comparison with optical flow, which is defined over two images.

In Sec. 2, we introduce the fundamental matrices, which are the basis of 3-D reconstruction from images for both finite motion and optical flow. In Sec. 3, we define a statistical model of feature uncertainty and give a theoretical bound on the estimation accuracy. In Sec. 4, we present an optimal algorithm for computing the fundamental matrix for finite motion. We also show simulation results that confirm that the solution indeed falls in the vicinity of the bound. In Sec. 5, we describe the 3-D reconstruction procedure for both finite motion and optical flow. In Sec. 6, we compare the results delivered by the finite motion and optical flow algorithms. Sec. 7 gives our conclusions and discussions on the raison d'être of the optical flow approach.

2. FUNDAMENTAL MATRICES

2.1 Definition

Let $\{(x_\alpha, y_\alpha)\}$ and $\{(x'_\alpha, y'_\alpha)\}$, $\alpha = 1, ..., N$, be image coordinates of two sets of N points on two different images; the image coordinate system is defined arbitrarily for each camera. Define vectors

$$\boldsymbol{x}_\alpha = \begin{pmatrix} x_\alpha/f_0 \\ y_\alpha/f_0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{x}'_\alpha = \begin{pmatrix} x'_\alpha/f_0 \\ y'_\alpha/f_0 \\ 1 \end{pmatrix}, \tag{1}$$

where f_0 is an appropriate scale factor[1] chosen so that x_α/f_0, y_α/f_0, x'_α/f_0, and y'_α/f_0 have an order 1. For brevity, we call the point having coordinates (x_α, y_α) simply "point $\boldsymbol{x}_\alpha$".

If $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$ are images of the same point in the scene, they must satisfy the constraint

$$(\boldsymbol{x}_\alpha, \boldsymbol{F}\boldsymbol{x}'_\alpha) = 0, \tag{2}$$

known as the *epipolar equation* [8, 13]. Here, $\boldsymbol{F}$ is a matrix of determinant 0, called the *fundamental matrix* [8, 13]. Throughout this paper, we denote by $(\boldsymbol{a}, \boldsymbol{b})$ the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Since the absolute scale of $\boldsymbol{F}$ is unconstrained, we normalize it to $\|\boldsymbol{F}\| = 1$, where the norm of a matrix $\boldsymbol{A} = (A_{ij})$ is defined by $\|\boldsymbol{A}\| = \sqrt{\sum_{i,j=1}^{3} A_{ij}^2}$.

If the motion is small, we can write

$$\boldsymbol{x}'_\alpha = \boldsymbol{x}_\alpha + \dot{\boldsymbol{x}}_\alpha \Delta t, \tag{3}$$

to a first approximation, where Δt is the interframe time lapse[2]. The vectors $\{\dot{\boldsymbol{x}}_\alpha\}$ describing the velocities on the image plane are called the *optical flow*. Substituting eq. (3) into eq. (2) and taking a first approximation, we obtain the following *flow epipolar equation* [3, 19, 33]:

$$(\boldsymbol{x}_\alpha, \boldsymbol{W}\dot{\boldsymbol{x}}_\alpha) + (\boldsymbol{x}_\alpha, \boldsymbol{C}\boldsymbol{x}_\alpha) = 0. \tag{4}$$

[1] For example, we can take it to be the size of the image frame.

[2] In practice, the interframe laps is taken to be unit time for convenience.

Here, $\boldsymbol{W} = (W_{ij})$ is an antisymmetric matrix, and $\boldsymbol{C} = (C_{ij})$ is a symmetric matrix. If we define

$$\boldsymbol{w} = \begin{pmatrix} W_{32} \\ W_{13} \\ W_{21} \end{pmatrix}, \tag{5}$$

the following *decomposability condition* holds [3, 19, 33]:

$$(\boldsymbol{w}, \boldsymbol{C}\boldsymbol{w}) = 0. \tag{6}$$

This constraint results from the rank constraint $\det \boldsymbol{F} = 0$ for finite motion. The matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ are called the *flow fundamental matrices* [19].

2.2 Computation

Computation of the fundamental matrix $\boldsymbol{F}$ has been studied by many researchers [1, 11, 12, 23, 24, 28, 29, 34, 35]. Common approaches are the *bundle-adjustment* and the *linear algorithm*.

The bundle-adjustment is known to be optimal under Gaussian noise, satisfying the *Cramer-Rao lower bound* (*CRLB*). However, a parameter space in very high dimensions need to be searched, and a good initial guess is required so that the search is not trapped into local minima [30].

The linear algorithm is based on the observation that the epipolar constraint (2) is linear in $\boldsymbol{F}$. This allows us to compute the solution by simply solving an eigenvalue problem, provided the rank constraint $\det \boldsymbol{F} = 0$ is ignored [11, 21, 31]. However, this efficiency sacrifices the accuracy, causing a large statistical bias [16].

In this paper, we present a modification to the linear algorithm such that the resulting performance is comparable to the bundle-adjustment: we first apply a technique called *renormalization* [16], which iteratively removes the statistical bias inherent to the linear algorithm; we then impose the rank constraint in a statistically optimal manner. We also give a (non-CRLB type) theoretical bound on the accuracy of $\boldsymbol{F}$ by generalizing the uncertainty analysis of Csurka et al. [7]. We demonstrate that the resulting solution indeed falls in the vicinity of the accuracy bound.

The same strategy has already been applied to the computation of the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$, and an optimal algorithm that delivers results in the vicinity of the theoretical accuracy bound has been obtained (see [19]).

3. UNCERTAINTY MODEL AND ACCURACY BOUND

3.1 Statistical Model of Uncertainty

We view $\{\boldsymbol{x}_\alpha\}$ and $\{\boldsymbol{x}'_\alpha\}$ as perturbed from their true locations $\{\bar{\boldsymbol{x}}_\alpha\}$ and $\{\bar{\boldsymbol{x}}'_\alpha\}$ that satisfy the epipolar constraint (2) exactly. We write

$$\boldsymbol{x}_\alpha = \bar{\boldsymbol{x}}_\alpha + \Delta\boldsymbol{x}_\alpha, \qquad \boldsymbol{x}'_\alpha = \bar{\boldsymbol{x}}'_\alpha + \Delta\boldsymbol{x}'_\alpha, \tag{7}$$

and regard $\Delta\boldsymbol{x}_\alpha$ and $\Delta\boldsymbol{x}'_\alpha$ as independent Gaussian random variables of mean $\boldsymbol{0}$ but not necessarily isotropic or homogeneous. We call $V[\boldsymbol{x}_\alpha] = E[\Delta\boldsymbol{x}_\alpha \Delta\boldsymbol{x}_\alpha^\top]$ and $V[\boldsymbol{x}'_\alpha] = E[\Delta\boldsymbol{x}'_\alpha \Delta\boldsymbol{x}'^\top_\alpha]$ the covariance matrices of $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$, respectively, where $E[\,\cdot\,]$ denotes expectation. In practice, we need not know the absolute covariance values; it suffices to know them *up to scale*. So, we write

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha], \qquad V[\boldsymbol{x}'_\alpha] = \epsilon^2 V_0[\boldsymbol{x}'_\alpha], \tag{8}$$

and assume that $V_0[\boldsymbol{x}_\alpha]$ and $V_0[\boldsymbol{x}'_\alpha]$, which we call the *normalized covariance matrices*, are known but the constant ϵ, which we call the *noise level*, is unknown. The normalized covariance matrices can be estimated from the Hessian of the residual surface of template matching [20].

Since the third components of $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$ are identically 1, the matrices $V_0[\boldsymbol{x}_\alpha]$ and $V_0[\boldsymbol{x}'_\alpha]$ are singular with third columns and third rows filled with zeros. If the noise has the same isotropic distribution everywhere, we have

$$V_0[\boldsymbol{x}_\alpha] = V_0[\boldsymbol{x}'_\alpha] = \mathrm{diag}(1,1,0), \tag{9}$$

where $\mathrm{diag}(\cdots)$ denotes the diagonal matrix with diagonal elements $\cdots$. We use eq. (9) as the default value when no information is available about the noise behavior.

If a corresponding pair of $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$ are identified with "optical flow" $\dot{\boldsymbol{x}}_\alpha = \boldsymbol{x}'_\alpha - \boldsymbol{x}_\alpha$ at the "midpoint" $(\boldsymbol{x}_\alpha + \boldsymbol{x}'_\alpha)/2$, the flow $\dot{\boldsymbol{x}}_\alpha$ and the location $\boldsymbol{x}_\alpha$ are uncorrelated and have their normalized covariance matrices $V_0[\boldsymbol{x}_\alpha] + V_0[\boldsymbol{x}'_\alpha]$ and $(V_0[\boldsymbol{x}_\alpha] + V_0[\boldsymbol{x}'_\alpha])/4$, respectively. It follows that the corresponding default noise model for the optical flow approach is given as follows [19]:

$$V_0[\dot{\boldsymbol{x}}_\alpha] = 2\mathrm{diag}(1,1,0), \qquad V_0[\boldsymbol{x}_\alpha] = \frac{1}{2}\mathrm{diag}(1,1,0). \tag{10}$$

3.2 Theoretical Accuracy Bound

Let $\hat{\boldsymbol{F}}$ be an estimate of the fundamental matrix, and $\bar{\boldsymbol{F}}$ its true value. The uncertainty of the estimate $\hat{\boldsymbol{F}}$ is measured by its *covariance tensor*

$$\mathcal{V}[\hat{\boldsymbol{F}}] = E[\mathcal{P}\big((\hat{\boldsymbol{F}} - \bar{\boldsymbol{F}}) \otimes (\hat{\boldsymbol{F}} - \bar{\boldsymbol{F}})\big)\mathcal{P}^\top], \tag{11}$$

where the operator $\otimes$ denotes tensor product: for matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$, the $(ijkl)$ element of their tensor product is $A_{ij}B_{kl}$. For tensors $\mathcal{P} = (P_{ijkl})$ and $\mathcal{T} = (T_{ijkl})$, the product $\mathcal{P}\mathcal{T}\mathcal{P}^\top$ is a tensor whose $(ijkl)$ element is $\sum_{m,n,p,q=1}^3 P_{ijmn}P_{klpq}T_{mnpq}$. The $(ijkl)$ element of the tensor $\mathcal{P} = (P_{ijkl})$ in eq. (11) is given by

$$P_{ijkl} = \delta_{ik}\delta_{jl} - \bar{F}_{ij}\bar{F}_{kl}, \tag{12}$$

where δ_{ij} is the Kronecker delta, taking 1 for $i = j$ and 0 otherwise.

Invoking the general theory of statistical optimization [16], we can derive a (non-CRLB type) lower bound on the covariance tensor $\mathcal{V}[\hat{\boldsymbol{F}}]$: If we define the *moment tensor* $\bar{\mathcal{M}} = (\bar{M}_{ijkl})$ by

$$\bar{\mathcal{M}} = \frac{1}{N}\sum_{\alpha=1}^N \bar{W}_\alpha \bar{\boldsymbol{x}}_\alpha \otimes \bar{\boldsymbol{x}}'_\alpha \otimes \bar{\boldsymbol{x}}_\alpha \otimes \bar{\boldsymbol{x}}'_\alpha, \tag{13}$$

$$\bar{W}_\alpha = \frac{1}{(\bar{\boldsymbol{x}}'_\alpha, \bar{\boldsymbol{F}}^\top V_0[\boldsymbol{x}_\alpha]\bar{\boldsymbol{F}}\bar{\boldsymbol{x}}'_\alpha) + (\bar{\boldsymbol{x}}_\alpha, \bar{\boldsymbol{F}}V_0[\boldsymbol{x}'_\alpha]\bar{\boldsymbol{F}}^\top\bar{\boldsymbol{x}}_\alpha)}, \tag{14}$$

the accuracy bound is given in the form

$$\mathcal{V}[\hat{\boldsymbol{F}}] \succ \frac{\epsilon^2}{N}\big(\mathcal{P}^{\mathcal{S}}\bar{\mathcal{M}}\mathcal{P}^{\mathcal{S}\top}\big)_7^-, \tag{15}$$

where $\mathcal{T} \succ \mathcal{S}$ for tensors $\mathcal{T}$ and $\mathcal{S}$ means that $\mathcal{T} - \mathcal{S}$ is a positive semi-definite tensor, and the operation $(\cdot)_r^-$ denotes the (Moore-Penrose) generalized inverse of rank r (discussed later). The $(ijkl)$ element of the tensor $\mathcal{P}^{\mathcal{S}} = (P^{\mathcal{S}}_{ijkl})$ in eq. (15) is given by

$$P^{\mathcal{S}}_{ijkl} = \delta_{ik}\delta_{jl} - \frac{\bar{F}^\dagger_{ji}\bar{F}^\dagger_{lk}}{\|\bar{\boldsymbol{F}}^\dagger\|^2}, \tag{16}$$

where $\bar{\boldsymbol{F}}^\dagger$ is the cofactor matrix of $\bar{\boldsymbol{F}}$.

For a tensor $\mathcal{T} = (T_{ijkl})$, a matrix $\boldsymbol{A} = (A_{ij})$, and a scalar λ, we say that $\boldsymbol{A}$ is an *eigenmatrix* of $\mathcal{T}$ with *eigenvalue* λ if $\mathcal{T}\boldsymbol{A} = \lambda\boldsymbol{A}$, where the product $\mathcal{T}\boldsymbol{A}$ is a matrix whose (ij) element is $\sum_{k,l=1}^3 T_{ijkl}A_{kl}$. The eigenmatrices and eigenvalues of a tensor can be computed by identifying a tensor and a matrix with a 9×9 matrix and a 9-dimensional vector [16].

A tensor $\mathcal{T} = (T_{ijkl})$ is said to be *symmetric* if $T_{ijkl} = T_{klij}$. A symmetric $3 \times 3 \times 3 \times 3$ tensor has nine real eigenvalues $\{\lambda_i\}$. The corresponding eigenmatrices $\{\boldsymbol{U}_i\}$ can be chosen to be an orthogonal system of matrices of unit norm, where the inner product of matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$ is defined by $(\boldsymbol{A}; \boldsymbol{B}) = \sum_{i,j=1}^{3} A_{ij}B_{ij}$. A symmetric tensor is *positive semi-definite* if its eigenvalues are all nonnegative.

Let $\lambda_1 \geq \cdots \geq \lambda_9$ (≥ 0) be the eigenvalues of a positive semi-definite symmetric tensor $\mathcal{T}$, and let $\{\boldsymbol{U}_1, ..., \boldsymbol{U}_9\}$ be the corresponding orthonormal set of eigenmatrices of unit norm. If $\lambda_r > 0$, the (*Moore-Penrose*) *generalized inverse* of $\mathcal{T}$ of rank r is computed as follows:

$$\mathcal{T}_r^- = \sum_{i=1}^{r} \frac{\boldsymbol{U}_i \otimes \boldsymbol{U}_i}{\lambda_i}. \tag{17}$$

The *root-mean-square error* of an estimate $\hat{\boldsymbol{F}}$ is defined by

$$\mathrm{rms}[\hat{\boldsymbol{F}}] = \sqrt{E[\|\mathcal{P}(\hat{\boldsymbol{F}} - \bar{\boldsymbol{F}})\|^2]}. \tag{18}$$

Since $\hat{\boldsymbol{F}}$ and $\bar{\boldsymbol{F}}$ are both normalized to unit norm, we have $0 \leq \mathrm{rms}[\hat{\boldsymbol{F}}] \leq 1$. From eq. (15), we have

$$\mathrm{rms}[\hat{\boldsymbol{F}}] \geq \frac{\epsilon}{\sqrt{N}} \sqrt{\mathrm{tr}\left(\mathcal{P}^S \bar{\mathcal{M}} \mathcal{P}^{S\mathsf{T}}\right)_7^-}, \tag{19}$$

where the *trace* $\mathrm{tr}\mathcal{T}$ of a tensor $\mathcal{T} = (T_{ijkl})$ is defined by

$$\mathrm{tr}\mathcal{T} = \sum_{k,l=1}^{3} T_{klkl}. \tag{20}$$

A similar accuracy bound is obtained for the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$, too (see [19]).

4. OPTIMAL ALGORITHM FOR THE FUNDAMENTAL MATRIX

The algorithm we are presenting for computing the fundamental matrix $\boldsymbol{F}$ has the same mathematical structure as that for computing the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ given in [19]. Namely, we optimally compute $\boldsymbol{F}$ by a technique called *renormalization* [16] without considering the rank constraint $\det \boldsymbol{F} = 0$ and then impose the rank constraint $\det \boldsymbol{F} = 0$ in a statistically optimal manner. We will show that accuracy is not lost by this type of two-stage cascading.

4.1. Renormalization

The renormalization algorithm proceeds as follows:

1. Let $c = 0$, $W_\alpha = 1$, $\alpha = 1, ..., N$, and $J = \infty$.

   # The symbol ∞ means a very large number, e.g., 10^{10}.

2. Compute the tensors $\mathcal{M} = (M_{ijkl})$ and $\mathcal{N} = (N_{ijkl})$ as follows:

$$M_{ijkl} = \frac{1}{N} \sum_{\alpha=1}^{N} W_\alpha x_{\alpha(i)} x'_{\alpha(j)} x_{\alpha(k)} x'_{\alpha(l)}, \tag{21}$$

$$N_{ijkl} = \frac{1}{N} \sum_{\alpha=1}^{N} W_\alpha (V_0[\boldsymbol{x}_\alpha]_{ik} x'_{\alpha(j)} x'_{\alpha(l)} + V_0[\boldsymbol{x}'_\alpha]_{jl} x_{\alpha(i)} x_{\alpha(k)}). \tag{22}$$

   # $x_{\alpha(i)}$ and $x'_{\alpha(i)}$ are the ith components of $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$, respectively, and $V_0[\boldsymbol{x}_\alpha]_{ij}$ and $V_0[\boldsymbol{x}'_\alpha]_{ij}$ are the (ij) elements of $V_0[\boldsymbol{x}_\alpha]_{ij}$ and $V_0[\boldsymbol{x}'_\alpha]_{ij}$, respectively.

3. Compute the nine eigenvalues $\lambda_1 \geq \cdots \geq \lambda_9$ of the tensor $\mathcal{M}$ and the corresponding orthonormal set $\{\boldsymbol{F}_1, ..., \boldsymbol{F}_9\}$ of eigenmatrices of unit norm.

4. Do the following computation:

- Update c as follows:

$$c \leftarrow c + \frac{\lambda_9}{(\boldsymbol{F}_9; \mathcal{N}\boldsymbol{F}_9)}. \tag{23}$$

- Compute W_α, $\alpha = 1, ..., N$, as follows:

$$W_\alpha = \frac{1}{(\boldsymbol{x}'_\alpha, \boldsymbol{F}_9^\top V_0[\boldsymbol{x}_\alpha]\boldsymbol{F}_9\boldsymbol{x}'_\alpha) + (\boldsymbol{x}_\alpha, \boldsymbol{F}_9 V_0[\boldsymbol{x}'_\alpha]\boldsymbol{F}_9^\top \boldsymbol{x}_\alpha)}. \tag{24}$$

- Compute the tensors $\mathcal{M}$ and $\mathcal{N}$ by eqs. (21) and (22).
- Let

$$J' \leftarrow J, \qquad J \leftarrow (\boldsymbol{F}_9, \mathcal{M}\boldsymbol{F}_9). \tag{25}$$

- If $J' < J$, let $J \leftarrow J'$.

  # The variable J stores the minimum residual.

- Else, compute the nine eigenvalues $\lambda_1 \geq \cdots \geq \lambda_9$ of the tensor

$$\hat{\mathcal{M}} = \mathcal{M} - c\mathcal{N}, \tag{26}$$

and the corresponding orthonormal set $\{\boldsymbol{F}_1, ..., \boldsymbol{F}_9\}$ of eigenmatrices of unit norm.

5. Repeat Step 4 until $J' \leq J$ or $|\lambda_9| \approx 0$.

   # This guarantees the iterations to converge.

6. Let $\boldsymbol{F}$ take the value $\boldsymbol{F}_9$, this being our estimate of the fundamental matrix.

The above procedure is based on the observation that, in the absence of noise, $\bar{\boldsymbol{F}}$ is the eigenmatrix of the moment tensor $\mathcal{M}$ defined in eq. (21) with eigenvalue 0. It can be shown that, in the presence of noise, $\mathcal{M}$ is statistically biased from its true value to a first approximation by a constant times the tensor $\mathcal{N}$ defined in eq. (22). By eq. (26), we iteratively remove the bias in $\mathcal{M}$ in such a way that the smallest eigenvalue of $\mathcal{M}$ converges to zero. The nature of this type of algorithm is rigorously analyzed by Chojnacki et al. [5, 6].

For the corresponding renormalization procedure for the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$, see [19].

4.2 Optimal Correction

We next apply a correction to $\boldsymbol{F}$, shifting it iteratively to the nearest value that satisfies the rank constraint $\det \boldsymbol{F} = 0$. Our procedure requires as inputs the eigenvalues and eigenmatrices emerging from the previous scheme. The steps of the method are as follows:

- Compute the normalized *covariance tensor* of $\boldsymbol{F}$ as follows:

$$\mathcal{V}_0[\boldsymbol{F}] = \frac{1}{N}\sum_{i=1}^{8} \frac{\boldsymbol{F}_i \otimes \boldsymbol{F}_i}{\lambda_i}. \tag{27}$$

- Repeat the following computation until $\det \boldsymbol{F} \approx 0$.

  1. Update $\boldsymbol{F}$ as follows:

$$\boldsymbol{F} \leftarrow N[\boldsymbol{F} - \frac{(\det \boldsymbol{F})\mathcal{V}_0[\boldsymbol{F}]\boldsymbol{F}^{\dagger\top}}{(\boldsymbol{F}^{\dagger\top}; \mathcal{V}_0[\boldsymbol{F}]\boldsymbol{F}^{\dagger\top})}]. \tag{28}$$

  # The operator $N[\,\cdot\,]$ denotes normalization of the norm to 1: $N[\boldsymbol{F}] = \boldsymbol{F}/\|\boldsymbol{F}\|$.

2. Compute the projection tensor $\mathcal{P} = (P_{ijkl})$ as follows:

$$P_{ijkl} = \delta_{ik}\delta_{jl} - F_{ij}F_{kl}. \tag{29}$$

3. Update the normalized covariance tensor $\mathcal{V}_0[\boldsymbol{F}]$ as follows:

$$\mathcal{V}_0[\boldsymbol{F}]_{ijkl} \leftarrow \sum_{m,n,p,q=1}^{3} P_{ijmn}P_{klpq}\mathcal{V}_0[\boldsymbol{F}]_{mnpq}. \tag{30}$$

This operation projects the error distribution onto the space orthogonal to $\boldsymbol{F}$.

It can be shown [16] that the estimate $\boldsymbol{F}$ resulting from the renormalization procedure has the normalized covariance tensor $\mathcal{V}_0[\hat{\boldsymbol{F}}]$ given in eq. (27). The above procedure enforces the rank constraint $\det \boldsymbol{F} = 0$ on $\boldsymbol{F}$ by iteratively updating $\boldsymbol{F}$ along the shortest path in the sense of the *Mahalanobis distance* associated with the normalized covariance tensor $\mathcal{V}_0[\hat{\boldsymbol{F}}]$. It can be proved that renormalization coupled with this type of correction produces a solution that attains the theoretical accuracy bound in the first order [16].

For the corresponding optimal correction procedure for the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$, see [19].

4.3. Program Package

The algorithm described above is implemented in C++ and placed in our Web page[3]. It outputs a solution $\hat{\boldsymbol{F}}$ along with its *standard deviation pair* $\{\boldsymbol{F}^{(+)}, \boldsymbol{F}^{(-)}\}$. These are the values in the parameter space that are separated from $\hat{\boldsymbol{F}}$ by the standard deviation in the two directions along which errors implied by eq. (15) are the most likely to occur.

We evaluate the right-hand side of eq. (15) by substituting the data and the estimate $\hat{\boldsymbol{F}}$ for their true values. The square noise level ϵ^2 in the expression can be estimated by

$$\hat{\epsilon}^2 = \frac{J}{1 - 8/N}, \tag{31}$$

using the value J returned by the renormalization procedure of Sec. 4.1. This type of estimation is known to give a good approximation to the true value [16].

Let $\lambda_{\max}$ be the maximum eigenvalue of the thus evaluated tensor on the right-hand side of eq. (15), and let $\boldsymbol{U}_{\max}$ be the corresponding eigenmatrix of unit norm. The standard deviation pair is defined by

$$\boldsymbol{F}^{(+)} = N[\hat{\boldsymbol{F}} + \sqrt{\lambda_{\max}}\boldsymbol{U}_{\max}], \qquad \boldsymbol{F}^{(-)} = N[\hat{\boldsymbol{F}} - \sqrt{\lambda_{\max}}\boldsymbol{U}_{\max}]. \tag{32}$$

If $\boldsymbol{F}^{(+)}$ and $\boldsymbol{F}^{(-)}$ coincide up to, say, three significant digits, the solution $\hat{\boldsymbol{F}}$ is likely to have accuracy up to approximately three significant digits.

The fundamental matrix cannot be defined uniquely if the feature points are in a degenerate configuration. This occurs, for example, when the camera translation is zero or all the feature points are on a special quadric called a *critical surface*, a typical instance of which is a planar surface [15]. If $\lambda_{\max}$ in eqs. (32) is predicted to be approximately 1 in the course of computation, our program judges that degeneracy has occurred and stops the computation after issuing a warning message.

The corresponding program package for computing the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ is also avaliable from the same site.

4.4 Simulated Experiments

Fig. 1 shows simulated 512 × 512-pixel images of a 3-D grid environment. They are supposedly captured by a moving camera with different focal lengths. Some of the epipolars (the images of the lines of sight starting from

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

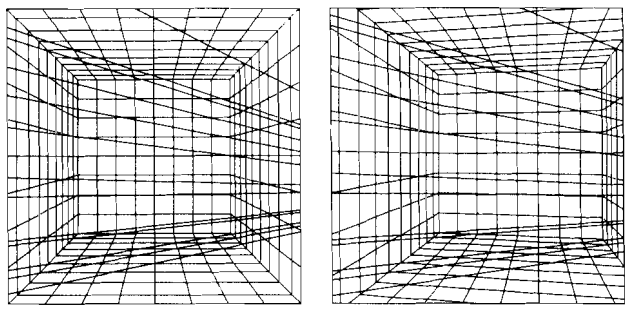
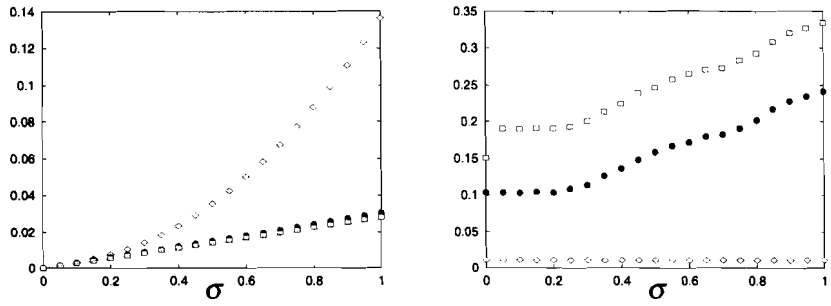


Figure 1: Simulated images of a 3-D scene and epipolars.



(a) (b)

Figure 2: Accuracy and efficiency of computation: our solutions (□), solutions without the optimal correction (●), and least-squares solutions (◇). (a) Root-means-squares error. The dotted line indicates the theoretical lower bound. (b) Average computation time (in seconds).

the projection center of the other camera) are superimposed. Random Gaussian noise of mean 0 and standard deviation σ (pixels) was added to the x and y coordinates of each grid point independently, and the fundamental matrix $\boldsymbol{F}$ was computed by using the default noise model of eq. (9). The renormalization converged after three or four iterations.

Fig. 2(a) shows a plot of the root-mean-squares error $\sqrt{\sum_{a=1}^{100} \|\mathcal{P}(\hat{\boldsymbol{F}}^a - \bar{\boldsymbol{F}})\|^2/100}$ over 100 trials for each σ using different noise each time, where $\hat{\boldsymbol{F}}^a$ is the ath estimate, $\bar{\boldsymbol{F}}$ is the true value, and $\mathcal{P}$ is the projection tensor defined in eq. (12). The symbol □ denotes solutions obtained via the method presented in this paper, and the dotted line indicates the theoretical lower bound derived from eq. (15). The symbol ● denotes renormalization solutions without applying the optimal correction of Section 4.2. The symbol ◇ denotes solutions computed by the widely used linear algorithm, often referred to as the *least-squares method* or the *algebraic distance minimization* [11, 12], which directly minimizes the sum of the squares of the epipolar constraint (2) in the form

$$\frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha, \boldsymbol{F}\boldsymbol{x}'_\alpha)^2 \to \min. \tag{33}$$

As we can see from Fig. 2(a), the errors in our estimates practically fall on the theoretical lower bound, which is known to be attained by the bundle adjustment. This confirms that our linear algorithm indeed achieves the accuracy of the bundle adjustment. Fig. 2(b) shows the average computation time on a Sun Ultra-30 workstation (SunOS 5.6). Naturally, the method takes more time than the naive least-squares method, but the theoretical accuracy bound is attained only at this higher computational cost.

Similar experiments for comfirming the optimality of the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ are given in [19].

5. 3-D RECONSTRUCTION FROM TWO VIEWS

We now describe the 3-D reconstruction procedure for both finite motion and optical flow.

5.1 Finite Motion Approach

After the fundamental matrix $\boldsymbol{F}$ is computed by the procedure described in the preceding section, the image locations $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$ of each feature point are corrected so as to satisfy the epipolar equation (2) exactly in a statistically optimal way [16]:

$$\hat{\boldsymbol{x}}_\alpha = \boldsymbol{x}_\alpha - \frac{E(\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha)}{V(\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha)} V_0[\boldsymbol{x}_\alpha]\boldsymbol{F}\boldsymbol{x}'_\alpha, \qquad \hat{\boldsymbol{x}}'_\alpha = \boldsymbol{x}'_\alpha - \frac{E(\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha)}{V(\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha)} V_0[\boldsymbol{x}'_\alpha]\boldsymbol{F}^\top\boldsymbol{x}_\alpha. \tag{34}$$

Here, we have defined

$$E(\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha) = (\boldsymbol{x}_\alpha, \boldsymbol{F}\boldsymbol{x}'_\alpha), \qquad V(\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha) = (\boldsymbol{x}'_\alpha, \boldsymbol{F}^\top V_0[\boldsymbol{x}_\alpha]\boldsymbol{F}\boldsymbol{x}'_\alpha) + (\boldsymbol{x}_\alpha, \boldsymbol{F}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{F}^\top\boldsymbol{x}_\alpha). \tag{35}$$

Letting $\boldsymbol{x}_\alpha \leftarrow \hat{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}'_\alpha \leftarrow \hat{\boldsymbol{x}}'_\alpha$, we repeat this procedure until the epopolar equation $E(\hat{\boldsymbol{x}}_\alpha, \hat{\boldsymbol{x}}'_\alpha) = 0$ is sufficiently satisfied. The convergence has quadratic speed, so one iteration is almost sufficient. This procedure is equivalent to the *triangulation* of Hartley and Sturm [11], which requires solving a sixth degree polynomial, but the above form is far more efficient, as pointed out by Torr and Zisseramann [28].

In reconstructing the 3-D structure from point correspondences over two images taken by two uncalibrated cameras, all information is encoded in the fundamental matrix $\boldsymbol{F}$ [8, 13]. Since $\boldsymbol{F}$ is defined up to scale and constrained to be $\det \boldsymbol{F} = 0$, it has *seven* degrees of freedom. The relative motion of the two cameras is specified by a translation vector $\boldsymbol{t}$ and a rotation matrix $\boldsymbol{R}$, but the absolute scale of the translation is indeterminate and a 3-D rotation has three degrees of freedom. So, the motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ have *five* degrees of freedom. It follows that only *two* camera parameters can be recovered.

A practical choice for them is the *focal lengths* f and f' of the two cameras, since other parameters can be pre-calibrated and fixed while zooming usually changes freely as the camera moves[4]. Hartley [10] presented an analytic procedure for computing the focal lengths f and f' from the fundamental matrix $\boldsymbol{F}$. The solution is obtained by applying the singular value decomposition (SVD) and solving linear equations in four unknowns. Pan et al. [26, 27] reduced this problem to solving cubic equations. Newsam et al. [25] refined these algorithms into a combination of SVD and linear equations in three unknowns. Kanatani and Matsunaga [17] reduced the problem to solving a quadratic equation in one variable. Bougnoux [2] presented an explicit formula for f in $\boldsymbol{F}$. The degeneracy condition for the solution to be indeterminate has also been analyzed [17, 25].

Among many mathematically equivalent alternatives, the most convenient may be the following modification of the Bougnoux formula [2] given by Kanatani and Matsunaga [17]:

$$f = f_0 \Bigg/ \sqrt{1 + \frac{\|\boldsymbol{F}\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k})\|\boldsymbol{e}' \times \boldsymbol{k}\|^2/(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})}{\|\boldsymbol{e}' \times \boldsymbol{k}\|^2\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2}},$$

$$f' = f_0 \Bigg/ \sqrt{1 + \frac{\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k})\|\boldsymbol{e} \times \boldsymbol{k}\|^2/(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})}{\|\boldsymbol{e} \times \boldsymbol{k}\|^2\|\boldsymbol{F}\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2}}. \tag{36}$$

Here, we put $\boldsymbol{k} = (0, 0, 1)^\top$. The symbols $\boldsymbol{e}$ and $\boldsymbol{e}'$ denote the unit eigenvectors of $\boldsymbol{F}^\top$ and $\boldsymbol{F}$, respectively, for eigenvalue 0; they represent the *epipoles*—the image of the projection center of the second camera in the first image and the image of that of the first camera in the second image, respectively [8, 13].

After the focal lengths f and f' have been obtained, we recompute $\{\hat{\boldsymbol{x}}_\alpha\}$ and $\{\hat{\boldsymbol{x}}'_\alpha\}$ by replacing f_0 in the vector representation of eqs. (1) by the corresponding values f and f' we have just computed. This can be done as follows:

$$\hat{\boldsymbol{x}}_\alpha \leftarrow \mathrm{diag}\Big(\frac{f_0}{f}, \frac{f_0}{f}, 1\Big)\hat{\boldsymbol{x}}_\alpha, \qquad \hat{\boldsymbol{x}}'_\alpha \leftarrow \mathrm{diag}\Big(\frac{f_0}{f'}, \frac{f_0}{f'}, 1\Big)\hat{\boldsymbol{x}}'_\alpha. \tag{37}$$

The motion parameters $\{\boldsymbol{t}, \boldsymbol{R}\}$ are analytically computed as follows (the indeterminate scale of the translation $\boldsymbol{t}$ is normalized to unit length: $\|\boldsymbol{t}\| = 1$) [15]:

[4]Strictly speaking, the *principal point* (the intersection of the optical axis with the image plane) may slightly move as zooming changes, but regarding it as a fixed point is known to be a good approximation.

1. Compute the following *essential matrix*:

$$\boldsymbol{E} = \mathrm{diag}(1, 1, \frac{f_0}{f})\boldsymbol{F}\mathrm{diag}(1, 1, \frac{f_0}{f'}). \tag{38}$$

   # This removes the dependence of the fundamental matrix on the focal lengths [17].

2. Compute the unit eigenvector $\boldsymbol{t}$ of $\boldsymbol{E}\boldsymbol{E}^\top$ for the smallest eigenvalue. The sign of $\boldsymbol{t}$ is chosen in such a way that

$$\sum_{\alpha=1}^{N} |\boldsymbol{t}, \hat{\boldsymbol{x}}_\alpha, \boldsymbol{E}\hat{\boldsymbol{x}}'_\alpha| > 0. \tag{39}$$

   # This is the constraint that the depths of the feature points have the same sign before and after the camera motion [15].

3. Apply SVD to $-\boldsymbol{t} \times \boldsymbol{E}$ as follows.

$$-\boldsymbol{t} \times \boldsymbol{E} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^\top. \tag{40}$$

   # For a vector $\boldsymbol{a}$ and a matrix $\boldsymbol{A}$, we define $\boldsymbol{a} \times \boldsymbol{A}$ to be the matrix consisting of columns that are the vector products of $\boldsymbol{a}$ and the individual columns of $\boldsymbol{A}$.

   # $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements (*singular values*) in non-increasing order; $\boldsymbol{V}$ and $\boldsymbol{U}$ are orthogonal matrices.

4. Compute the rotation $\boldsymbol{R}$ as follows:

$$\boldsymbol{R} = \boldsymbol{V}\mathrm{diag}(1, 1, \det \boldsymbol{V}\boldsymbol{U}^\top)\boldsymbol{U}^\top. \tag{41}$$

This procedure produces a least-squares solution for $\{\boldsymbol{t}, \boldsymbol{R}\}$ even if the rank constraint $\det \boldsymbol{F} = 0$ is not strictly satisfied [15].

The 3-D position of the αth point is given by

$$\hat{\boldsymbol{r}}_\alpha = \hat{Z}_\alpha \hat{\boldsymbol{x}}_\alpha, \qquad \hat{\boldsymbol{r}}'_\alpha = \hat{Z}'_\alpha \hat{\boldsymbol{x}}'_\alpha, \tag{42}$$

with respect to the first and the second camera coordinate systems, respectively, where the depths $\hat{Z}_\alpha$ and $\hat{Z}'_\alpha$ are given as follows [15]:

$$\hat{Z}_\alpha = (\boldsymbol{t} \times \boldsymbol{R}\hat{\boldsymbol{x}}'_\alpha, \boldsymbol{n}_\alpha), \qquad \hat{Z}'_\alpha = (\boldsymbol{t} \times \hat{\boldsymbol{x}}_\alpha, \boldsymbol{n}_\alpha). \tag{43}$$

Here, we have defined

$$\boldsymbol{n}_\alpha = \frac{\hat{\boldsymbol{x}}_\alpha \times \boldsymbol{R}\hat{\boldsymbol{x}}'_\alpha}{\|\hat{\boldsymbol{x}}_\alpha \times \boldsymbol{R}\hat{\boldsymbol{x}}'_\alpha\|^2}. \tag{44}$$

Finally, we need to adjust the sings of the depths because the sign of the fundamental matrix $\boldsymbol{F}$ is indeterminate. The signs of $\{\hat{Z}_\alpha\}$ and $\{\hat{Z}'_\alpha\}$ are inverted, if necessary, so that

$$\sum_{\alpha=1}^{N} (\mathrm{sgn}[\hat{Z}_\alpha] + \mathrm{sgn}[\hat{Z}_\alpha]) > 0, \tag{45}$$

where $\mathrm{sgn}[\cdot]$ is the signature function that takes 1, 0, and -1 for $x > 0$, $x = 0$, and $x < 0$, respectively. This operation is necessary because we may not select the correct solution if we simply compute $\sum_{\alpha=1}^{N}(\hat{Z}_\alpha + \hat{Z}'_\alpha)$; a very large positive depth may turn out to be close to $-\infty$ due to noise.

5.2 Optical Flow Approach

The 3-D reconstruction from optical flow goes similarly. After optimally computing the flow fundamental matrix $\boldsymbol{W}$ and $\boldsymbol{C}$ using the previously reported method [19], the flow $\dot{\boldsymbol{x}}_\alpha$ and image location $\boldsymbol{x}_\alpha$ of each feature point are corrected so as to satisfy the flow epipolar equation (4) exactly in a statistically optimal way [16]:

$$\hat{\dot{\boldsymbol{x}}}_\alpha = \dot{\boldsymbol{x}}_\alpha + \frac{E(\dot{\boldsymbol{x}}_\alpha, \boldsymbol{x}_\alpha)}{V(\dot{\boldsymbol{x}}_\alpha, \boldsymbol{x}_\alpha)} V_0[\dot{\boldsymbol{x}}_\alpha] \boldsymbol{W} \boldsymbol{x}_\alpha, \qquad \hat{\boldsymbol{x}}_\alpha = \boldsymbol{x}_\alpha - \frac{E(\dot{\boldsymbol{x}}_\alpha, \boldsymbol{x}_\alpha)}{V(\dot{\boldsymbol{x}}_\alpha, \boldsymbol{x}_\alpha)} V_0[\boldsymbol{x}_\alpha](\boldsymbol{W}\dot{\boldsymbol{x}}_\alpha + 2\boldsymbol{C}\boldsymbol{x}_\alpha). \tag{46}$$

Here, we have defined

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

$$V(\dot{\boldsymbol{x}}_\alpha, \boldsymbol{x}_\alpha) = (\boldsymbol{W}\boldsymbol{x}_\alpha, V_0[\dot{\boldsymbol{x}}_\alpha]\boldsymbol{W}\boldsymbol{x}_\alpha) + (\boldsymbol{W}\dot{\boldsymbol{x}}_\alpha + 2\boldsymbol{C}\boldsymbol{x}_\alpha, V_0[\boldsymbol{x}_\alpha](\boldsymbol{W}\dot{\boldsymbol{x}}_\alpha + 2\boldsymbol{C}\boldsymbol{x}_\alpha)). \tag{47}$$

Letting $\dot{\boldsymbol{x}}_\alpha \leftarrow \hat{\dot{\boldsymbol{x}}}_\alpha$ and $\boldsymbol{x}_\alpha \leftarrow \hat{\boldsymbol{x}}_\alpha$, we repeat this procedure until the flow epipolar equation $E(\hat{\dot{\boldsymbol{x}}}_\alpha, \hat{\boldsymbol{x}}_\alpha) = 0$ is sufficiently satisfied. As in the finite motion approach, the convergence has quadratic speed, so one iteration is almost sufficient.

In reconstructing the 3-D structure from optical flow taken by an uncalibrated camera, all information is encoded in the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ [3, 19, 33]. They are determined up to scale and constrained by the decomposability condition (6). Since $\boldsymbol{W}$ and $\boldsymbol{C}$ are antisymmetric and symmetric matrices, respectively, they have *seven* degrees of freedom in total. The instantaneous motion of the camera is specified by the translation velocity $\boldsymbol{v}$ and the rotation velocity $\boldsymbol{\omega}$, but the absolute scale of the translational motion is indeterminate. So, the motion parameters $\{\boldsymbol{v}, \boldsymbol{\omega}\}$ have *five* degrees of freedom. It follows that, as in the case of finite motion, only *two* camera parameters can be recovered.

A practical choice for them is, again as in the case of finite motion, the focal length f and its change rate $\dot{f}$, since other parameters can be pre-calibrated and fixed while zooming usually changes freely as the camera moves. Brooks et al. [3] presented a complicated procedure for computing f, $\dot{f}$, $\boldsymbol{v}$ and $\boldsymbol{\omega}$ from the flow fundamental matrices $\boldsymbol{W} = (W_{ij})$ and $\boldsymbol{C} = (C_{ij})$. Here, we present an elegant *group-theoretical procedure* using *irreducible representations* of the group of 2-D rotations $SO(2)$ [14]. The indeterminate scale of the translation velocity $\boldsymbol{v}$ is normalized to unit length: $\|\boldsymbol{v}\| = 1$.

Let w_i be the ith component of the vector $\boldsymbol{w}$ defined in eq. (5), and do the following computation:

$$A = C_{11} + C_{22}, \qquad \tilde{B} = (C_{11} - C_{22}) + 2iC_{12}, \qquad \tilde{C} = 2(C_{13} + iC_{23}), \qquad D = C_{33}, \tag{48}$$

$$\tilde{w} = w_1 + iw_2, \qquad \tilde{\omega}' = \frac{\tilde{B}}{\tilde{w}}, \qquad \omega_1' = \Re[\tilde{\omega}'], \qquad \omega_2' = \Im[\tilde{\omega}'], \qquad \omega_3 = -\frac{A + (\tilde{w}, \tilde{\omega}')}{2w_3}, \tag{49}$$

$$f' = \sqrt{-\frac{D}{(\tilde{w}, \tilde{\omega}')}}, \qquad \tilde{\phi} = \frac{\tilde{C} - f'^2 w_3 \tilde{\omega}'}{\tilde{w}}, \qquad \omega_3 = \Re[\tilde{\phi}], \qquad \dot{f}' = -f'\Im[\tilde{\phi}], \tag{50}$$

$$\omega_1 = f'\omega_1', \qquad \omega_2 = f'\omega_2', \qquad f = f'f_0, \qquad \dot{f} = \dot{f}'f_0, \qquad \boldsymbol{v} = N[\begin{pmatrix} w_1 \\ w_2 \\ (f/f_0)w_3 \end{pmatrix}]. \tag{51}$$

Here, i is the imaginary unit. The quantities with tildes are complex numbers: $\Re[\cdot]$ and $\Im[\cdot]$ denote the real and imaginary parts, respectively. We define the "inner product" of complex numbers $z = x + iy$ and $z' = x' + iy'$ by $(z, z') = xx' + yy'$. The operation $N[\cdot]$ designates normalization into a unit vector: $N[\boldsymbol{a}] = \boldsymbol{a}/\|\boldsymbol{a}\|$. Note that ω_3 is computed in two ways by the fifth of eqs. (49) and the third of eqs. (50). The decomposability condition (6) requires that the two values coincide.

The 3-D positions of the feature points are reconstructed as follows. First, we recompute $\{\hat{\dot{\boldsymbol{x}}}_\alpha\}$ and $\{\hat{\boldsymbol{x}}_\alpha\}$ by replacing f_0 by its true value f and incorporating its change rate $\dot{f}$. This is done as follows:

$$\hat{\dot{\boldsymbol{x}}}_\alpha \leftarrow \frac{f_0}{f}\Big(\hat{\dot{\boldsymbol{x}}}_\alpha - \frac{\dot{f}}{f}\mathrm{diag}(1, 1, 0)\hat{\boldsymbol{x}}_\alpha\Big), \qquad \hat{\boldsymbol{x}}_\alpha \leftarrow \mathrm{diag}\Big(\frac{f_0}{f}, \frac{f_0}{f}, 1\Big)\hat{\boldsymbol{x}}_\alpha. \tag{52}$$

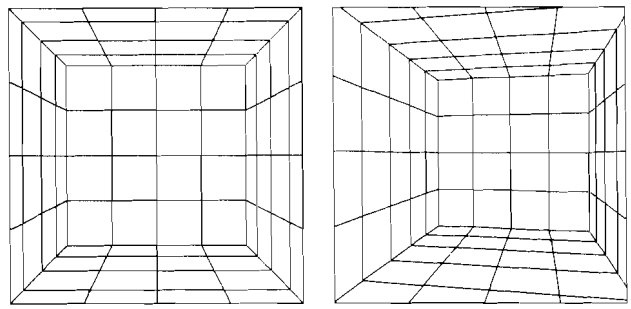


Figure 3: Simulated images of a 3-D scene.

The 3-D position of the αth point is given by

$$\hat{\boldsymbol{r}}_\alpha = \hat{Z}_\alpha \hat{\boldsymbol{x}}_\alpha, \tag{53}$$

where the depth $\hat{Z}_\alpha$ is given as follows [15]:

$$\hat{Z}_\alpha = -\frac{(\boldsymbol{v}, \boldsymbol{S}_\alpha \boldsymbol{v})}{(\boldsymbol{v}, \boldsymbol{S}_\alpha(\dot{\boldsymbol{x}}_\alpha + \boldsymbol{\omega} \times \boldsymbol{x}_\alpha))}. \tag{54}$$

Here, we have defined

$$\boldsymbol{S}_\alpha = (\boldsymbol{I} - \boldsymbol{x}_\alpha \boldsymbol{k}^\top)^\top (\boldsymbol{I} - \boldsymbol{x}_\alpha \boldsymbol{k}^\top), \tag{55}$$

where $\boldsymbol{k} = (0, 0, 1)^\top$. In view of the sign indeterminacy of the flow fundamental matrices $\boldsymbol{W}$ and $\boldsymbol{C}$, the signs of $\{\hat{Z}_\alpha\}$ are inverted, if necessary, so that the following condition is satisfied for the same reason as in the finite motion approach:

$$\sum_{\alpha=1}^{N} \mathrm{sgn}[\hat{Z}_\alpha] > 0. \tag{56}$$

6. Performance Comparison

We now compare the performance of the finite motion algorithm and the optical flow algorithm by simulation and real-image experiments using the same data.

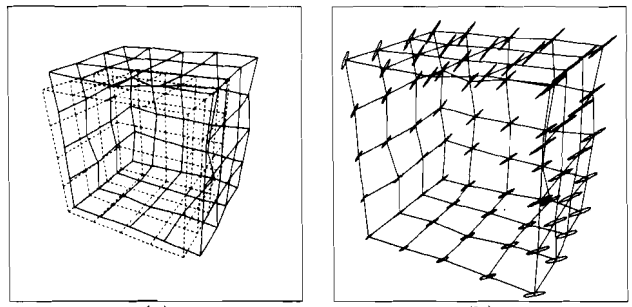
6.1 Simulation Experiments

Fig. 3 shows simulated 512 × 512-pixel images of a 3-D scene. We added independent random Gaussian noise of mean 0 and standard deviation 3 (pixels) to each of the x and y coordinates of the grid points and reconstructed the 3-D shape using the default noise model of eq. (9).

Fig. 4(a) shows the 3-D shape (solid lines) reconstructed by the finite motion algorithm superimposed on the true shape (dotted lines) rescaled to $\|\boldsymbol{t}\| = 1$. Fig. 4(b) shows uncertainty ellipsoids centered on the reconstructed vertices. We can evaluate the uncertainty of the computed fundamental matrix $\boldsymbol{F}$ in the form of the covariance tensor, so we can propagate it to compute the uncertainty of the 3-D reconstruction (we omit the details). The ellipsoids in Fig. 4(b) indicate three times the standard deviation in each orientation. As we can observe, they are very thin having their major axes approximately in the depth orientation. They are larger for points further away from the cameras[5].

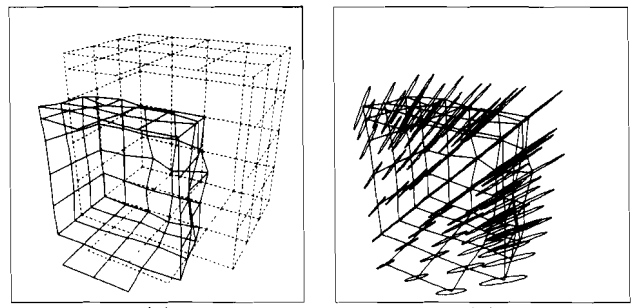
Fig. 5 shows the corresponding 3-D shape reconstructed by the optical flow algorithm using the default noise model of eq. (10). Here, the random Gaussian noise is reduced to standard deviation 0.5 (pixel), because adding more noise would deteriorate the results intolerably. From this, we can observe the poor performance of the optical flow approach even in this low noise level although the computation is optimal.

[5]The uncertainty shown here is relative to the first camera coordinate system with the camera translation normalized to unit length, so it does not have an absolute meaning since the first camera coordinate system and the camera translation also have their uncertainty. To extract an absolute meaning, we need the *gauge theory* of uncertainty description [18].



(a) (b)

Figure 4: The finite motion approach. (a) Reconstructed shape (solid lines) and the true shape (dotted lines). (b) The uncertainty ellipsoids of the grid points.



(a) (b)

Figure 5: The optical flow approach. (a) Reconstructed shape (solid lines) and the true shape (dotted lines). (b) The uncertainty ellipsoids of the grid points.

6.2 Real Image Experiments

Fig. 6 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 3-D shape. Fig. 7 shows the 3-D shape computed in two ways: (a) the finite motion approach with the noise model (9); (b) the optical flow approach with the noise model (10). Wireframes consisting of some of the reconstructed points are shown for visual aide.

On each reconstructed point is centered the uncertainty ellipsoid that indicates the standard deviation in each orientation (this time not magnified by three times). They are like thin needles, showing that the uncertainty is very large along the depth orientation. Although the reconstructed shape itself looks natural for both (a) and (b), we can clearly see that the optical flow solution has far larger uncertainty than the finite motion solution.

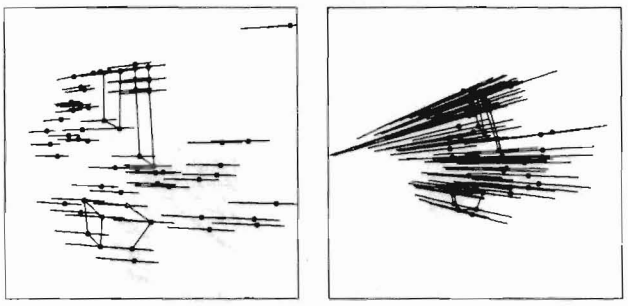
Fig. 8 shows real images (512×768 pixels) of a car. We manually selected feature points as marked in the images and reconstructed the 3-D shape. Fig. 9 shows some new views generated by creating a wireframe model from the reconstructed points and mapping the texture to it: the upper row is obtained by the finite motion approach with the noise model (9); the lower row is obtained by the optical flow approach with the noise model (10).

Although it is difficult to grasp the exact 3-D shape from static views, we are given a fairly realistic impression of the 3-D shape by continuously changing the viewpoint. After careful observations, however, we find that the 3-D shape is unnaturally deformed in the part far away from the viewer as compared with the front part, which is fairy accurate. We also find that the deformation is larger for the optical flow solution than the finite motion solution.

7. CONCLUDING REMARKS

We have presented two linear algorithms for 3-D reconstruction from point correspondences over two views: one is for finite motion; the other is for optical flow. Both are optimal in the sense that no other algorithms could possibly outperform it. They first compute the fundamental matrix and the flow fundamental matrices by renormalization

Figure 6: Real images of an indoor scene.



(a) (b)

Figure 7: Reconstructed points and their uncertainty ellipsoids: (a) the finite motion approach; (b) the optical flow approach.

followed by optimal correction. For each approach, we have derived a theoretical accuracy bound based on a static uncertainty model of feature locations and confirmed by experiments that the solution indeed falls in the vicinity of the accuracy bound. Since the major part of the optical flow algorithm is described in our previously presented paper [19], we have mainly focused on the finite motion algorithm.

Next, we presented for each approach a procedure for reconstructing individual 3-D positions in a statistically optimal way. For optical flow, in particular, we introduced elegant formulae based on the group representation theory [14]. We then compared the performance of the two algorithms by simulation and real-image experiments using the same data. This is the first impartial comparison ever done in the sense that the two algorithms are both optimal, extracting the information contained in the data to a maximum possible degree.

Our experiments have shown that the finite motion solution is always superior to the optical flow solution. Since optical flow is a first order approximation of finite motion, the computation could be stabler for optical flow when the disparity is very small. We tested this by simulation. Evidently, the 3-D information cannot be obtained if the disparity between the two images is too small whatever method we use. Gradually reducing the disparity, we ran the two algorithms for the same data and found that the optical flow algorithm always collapsed first. In all the experiments we did, we were unable to find any advantage of the optical flow algorithm in accuracy, efficiency, or stability.

The reason why so many studies of 3-D reconstruction from optical flow have been done in the past lies perhaps in the usefulness of optical flow and the ease of its detection. Optical flow has sufficient 3-D information. Indeed, humans can easily perceive the 3-D structure of the scene by simply looking at it. So, it is natural that people have sought algorithms for 3-D reconstruction from optical flow[6]. Our investigation has revealed, however, that optical flow should be regarded as an auxiliary tool for correspondence detection. It may be useful in many image processing applications including motion segmentation, but when it comes to 3-D reconstruction, one should use the finite motion algorithm as we did here; nothing is gained by using first order approximations.

[6]Some acknowledged that the purpose of studying 3-D reconstruction algorithms from optical flow was to understand the workings of the human brain [22, 32].

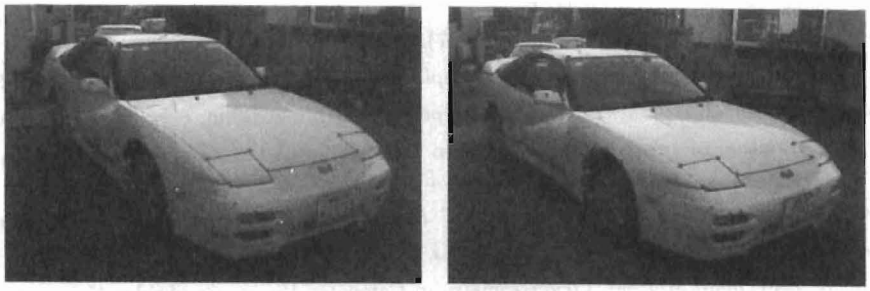


Figure 8: Real images of a car.

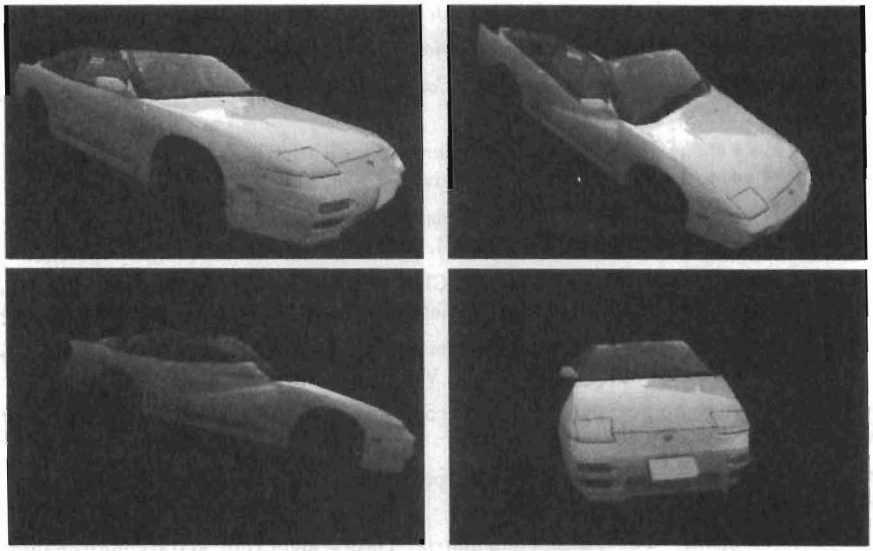


Figure 9: The reconstructed and texture-mapped 3-D shape: the finite motion approach (above); the optical flow approach (below).

Acknowledgments

The authors thank Hitoshi Mishima of 3D, Inc., Japan and Yoshiyuki Shimizu of Sharp, Ltd, Japan, for doing various experiments for us. They also thank Chikara Matsunaga of FOR-A, Co. Ltd, Japan, Toshio Ueshiba of AIST, Japan, Long Quan and Bill Triggs of INRIA Rhône Alpes, France, Mike Brooks and his colleagues of the University of Adelaide, Australia, Du Huynh of Murdoch University, Australia, Luis Baumela of Madrid Technical University, Spain, and Lourdes de Agapito of the University of Oxford, U.K., for 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).

References

[1] M. Bober, N. Geogis and J. Kittler, "On accurate and robust estimation of fundamental matrix," *Comput. Vision Image Understanding*, 72(1), 39–53 (1998).

[2] S. Bougnoux, "From projective to Euclidean space under any practical situation, a criticism of self calibration," *Proc. 6th Int. Conf. Comput. Vision.*, January 1998, Bombay, India, pp. 790–796.

[3] M. J. Brooks, W. Chojnacki and L. Baumela, "Determining the egomotion of an uncalibrated camera from instantaneous optical flow," *J. Opt. Soc. Am. A*, 14(10), 2670–2677 (1997).

[4] A. R. Bruss and B. K. P. Horn, "Passive navigation," *Comput. Vision Graphics Image Process.*, 21(1), 3–20 (1983).

[5] W Chojnacki, M. J. Brooks, A. van den Hengel and D. Gawley, "On the fitting of surfaces to data with covariances," *IEEE Trans. Patt. Anal. Mach. Intell.*, 22(11), 1294–1303 (2000).

[6] W. Chojnacki, M. J. Brooks and A. van den Hengel, "Rationalising the renormalisation method of Kanatani," *J. Math. Imaging Vision*, 14(1), 21–38 (2001).

[7] G. Csurka, C. Zeller, Z. Zhang and O. D. Faugeras, "Characterizing the uncertainty of the fundamental matrix," *Comput Vision Image Understand.*, 68(1), 18–36 (1997).

[8] O. Faugeras and Q.-T. Luong, *The Geometry of Multiple Images*, MIT Press, Cambridge, MA, U.S.A., 2001.

[9] J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA, U.S.A., 1979.

[10] R. I. Hartley, "Estimation of relative camera position for uncalibrated cameras," *Proc. 2nd Euro. Conf. Comput. Vision*, May 1992, Santa Margherita Ligure, Italy, pp. 579–587.

[11] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Patt. Anal. Mach. Intell.*, 97(6), 580–593 (1997).

[12] R. I. Hartley, "Minimizing algebraic error," *Phil. Trans. R. Soc. Lond.*, A, 3569, 1175–1192. (1998).

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.

[14] K. Kanatani, *Group-Theoretical Methods in Image Understanding*, Springer, Berlin, Germany, 1990,

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

[16] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, The Netherlands (1996).

[17] K. Kanatani and C. Matsunaga, "Closed-form expression for focal lengths from the fundamental matrix," *Proc. 4th Asian Conf. Comput. Vision*, January 2000, Taipei, Taiwan, pp. 128–133.

[18] K. Kanatani and D. D. Morris, "Gauges and gauge transformations for uncertainty description of geometric structure with indeterminacy," *IEEE Trans. Inform. Theo.*, 47(5), 2017–2028 (2001).

[19] K. Kanatani, Y. Shimizu, N. Ohta, M. J. Brooks, W. Chojnacki and A. van den Hengel, "Fundamental matrix from optical flow: Optimal computation and reliability evaluation," *J. Eletrconic Imaging*, 9(2), 194–202 (2000).

[20] Y. Kanazawa and K. Kanatani. "Do we really have to consider covariance matrices for image features?", *Proc. 8th Int. Conf. Comput. Vision*, July 2001, Vancouver, Canada, Vol. 2, pp. 586–591.

[21] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, 293(10), 133–135 (1981).

[22] H. C. Longuet-Higgins and K. Prazdny, "The interaction of a moving retinal images," *Proc. Roy. Soc. Lond.*, B208, 385–397 (1980).

[23] Q.-T. Luong and O. D. Faugeras, "The fundamental matrix: Theory, algorithm, and stability analysis," *Int. J. Comput. Vision*, 17(3), 43–7 (1996).

[24] M. Mühlich and R. Mester, "The role of total least squares in motion analysis," *Proc. 5th Euro. Conf. Comput. Vision*, June 1998, Freiburg, Germany, pp. 305–321.

[25] G. N. Newsam, D. Q. Huynh, M. J. Brooks and H.-P. Pan, "Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations," *Int. Arch. Photogram. Remote Sensing*, 31, B3-III, July 1996, Vienna, Austria, pp. 575–580.

[26] H.-P. Pan, M. J. Brooks and G. Newsam, "Image resituation: initial theory," *Proc. SPIE: Videometrics IV*, October 1995, Philadelphia, USA, pp. 162–173.

[27] H.-P. Pan, D. Q. Huynh and G. Hamlyn, "Two-image resituation: Practical algorithm," *Proc. SPIE: Videometrics IV*, October 1995, Philadelphia, USA. pp. 174–190.

[28] P. H. S. Torr and A. Zissermann, "Performance characterization of fundamental matrix estimation under image degradation," *Mach. Vision Appl.*, 9, 321–333 (1997).

[29] P. H. S. Torr and A. Zisserman, "Robust detection of degenerate configurations while estimating the fundamental matrix," *Comput. Vision Image Understanding*, 71(3), 312–333 (1998).

[30] B. Triggs, P. F. Mclauchlan, R. I. Hartley and A. W. Fitzgibbon, "Bundle adjustment–A modern synthesis," in B. Triggs, A. Zisserman and R. Szeliski, Eds., in *Vision Algorithms: Theory and Practice*, Springer, Berlin, 2000, pp. 298–375.

[31] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," *IEEE Trans. Patt. Anal. Mach. Intell.*, 6(1), 12–27 (1984)

[32] S. Ullman, *The Interpretation of Visual Motion*, MIT Press, Cambridge, MA, U.S.A., 1979.

[33] T. Viéville and O. D. Faugeras, "The first order expansion of motion equations in the uncalibrated case," *Comput. Vision Image Understanding*, 64(1), 128–146 (1996).

[34] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vision*, 27(2), 161–195 (1998).

[35] Z. Zhang, "On the optimization criteria used in two-view motion analysis," *IEEE Trans. Patt. Anal. Mach. Intell.*, 20(7), 717–729 (1998).