Fundamental Matrix from Optical Flow: Optimal Computation and Reliability Evaluation

Kenichi Kanatani*, Yoshiyuki Shimizu†, Naoya Ohta*,
Michael J. Brooks‡, Wojciech Chojnacki‡, and Anton van den Hengel‡

*Department of Computer Science, Gunma University, Kiryu, Gunma 376-8515, Japan
{kanatani,ohta}@cs.gunma-u.ac.jp

†Sharp, Ltd., Osaka, 545-0013, Japan

‡Department of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia
{mjb,wojtek,hengel}@cs.adelaide.edu.au

Abstract

The optical flow observed by a moving camera satisfies, in the absence of noise, a special equation analogous to the epipolar constraint arising in stereo vision. Computing the "flow fundamental matrix" of this equation is an essential prerequisite to undertaking 3-D analysis of the flow. This paper presents an optimal formulation of the problem of estimating this matrix under an assumed noise model. This model admits independent Gaussian noise that is not necessarily isotropic or homogeneous. A theoretical bound is derived for the accuracy of the estimate. An algorithm is then devised that employs a technique called *renormalization* to deliver an estimate and then corrects the estimate so as to satisfy a particular decomposability condition. The algorithm also provides an evaluation of the reliability of the estimate. Epipoles and their associated reliabilities are computed in both simulated and real-image experiments. Experiments indicate that the algorithm delivers results in the vicinity of the theoretical accuracy bound.

Keywords: fundamental matrix, optical flow, epipole, reliability evaluation, theoretical accuracy bound, renormalization.

1 Introduction

Two distinct approaches are available for extracting 3-D information from motion images. One utilizes data in the form of *point correspondences* [6, 9, 10, 23], while the other exploits data in the form of *optical flow*, which is theoretically an instantaneous velocity field over the image frame. Under this latter approach, we can compute the 3-D motion of the camera and the focal length, and hence the 3-D structure of the scene, in an analytically closed form by regarding the image motion as infinitesimal [5, 22]. A similar analysis can be carried out using the fundamental matrix [8, 12, 15], but it requires many stages of complicated analysis and computation.

Optical flow analysis relies heavily upon the accuracy of the optical flow estimation. Usually, optical flow is determined by applying a differential equation called the *gradient constraint* to the grey levels of the images [1, 3, 4, 11, 13, 17, 18, 24]. This has the advantage that the flow can be detected over the image frame via a single operation. However, this process involves many stages of approximation such as replacing differentials by finite differences, imposing smoothness

constraints, and applying regularization methods. As a result, the accuracy of the detected flow is not high enough for 3-D reconstruction of the scene. In order to obtain accurate flow, we need to trace individual feature points directly by the same means as those used for finite motion, such as template matching and spatio-temporal analysis. In this paper, we assume that optical flow has been obtained with high accuracy for a limited number of salient feature points.

If the images are captured by a camera moving through a stationary scene, the associated optical flow satisfies, in the absence of noise, a motion equation of special form. This equation is analogous to the epipolar equation arising in stereo vision. The coefficients of the motion equation can be expressed in terms of a matrix, which plays the same role as the fundamental matrix in stereo analysis. For this reason, we call it the *flow fundamental matrix*. The main purpose of this paper is to compute the matrix as accurately as possible from noisy data.

Our first step is to study the behavior of the noise, which is the cause of deterioration in accuracy. Having presented a mathematical model of noise, we can work towards an optimal solution under that model by employing statistical analysis. If a specific statistical model of noise is assumed, an optimal computational method for that model can be derived; if another model is assumed, another method should be derived. The assumed model should be clearly stated from the outset, and the computational method should be derived with respect to the model. It is in following this methodology that our approach goes beyond the original works in this domain [5, 22]. A further advance is achieved in that our performance evaluation takes into account a newly derived *theoretical accuracy bound*. If the method does not attain that bound, there is room for improvement; if the bound is attained, no further improvement is possible.

In this paper, we describe a procedure for computing the flow fundamental matrix under the model that noise is small, independent, and Gaussian but not necessarily isotropic or homogeneous; the noise behavior may be different from point to point and dependent on orientation. We derive a theoretical accuracy bound under the assumed model and demonstrate by experiments that our algorithm delivers results in the vicinity of the bound. In this sense, the algorithm is optimal.

The algorithm has the added advantage that it can evaluate the reliability of the solution automatically. This implies that if we carry out 3-D reconstruction based on the computed flow fundamental matrix, we can also evaluate the reliability of the recovered 3-D structure.

If the camera is calibrated and the imaging geometry can be modeled as perspective projection with a known focal length, extra constraints are imposed on the flow fundamental matrix [9]. A statistically optimal algorithm for computing the flow fundamental matrix in this case has already been proposed, and the reliability of the resulting 3-D reconstruction was evaluated in quantitative terms [19, 10]. The algorithm presented here is obtained from it by relaxing the decomposability condition.

2 Flow Fundamental Matrix

In this section, we give basic definitions and terminologies used throughout this paper.

Let $\{(x_\alpha, y_\alpha)\}$ and $\{(x'_\alpha, y'_\alpha)\}$, $\alpha = 1, ..., N$, be image coordinates (in pixels) of two sets of N points on two different images. Here, separate rectangular coordinate systems are assumed for each camera. We define the "flow" and the "midpoint" of the αth point as

$$\dot{\boldsymbol{x}}_\alpha = \begin{pmatrix} (x'_\alpha - x_\alpha)/f_0 \\ (y'_\alpha - y_\alpha)/f_0 \\ 0 \end{pmatrix}, \qquad \boldsymbol{x}_\alpha = \begin{pmatrix} (x_\alpha + x'_\alpha)/2f_0 \\ (y_\alpha + y'_\alpha)/2f_0 \\ 1 \end{pmatrix}, \tag{1}$$

where f_0 is a scale factor[1] (in pixels) chosen so that x_α/f_0, y_α/f_0, x'_α/f_0, and y'_α/f_0 have order 1. The two sets of points satisfy the following *epipolar equation* [5, 10, 19, 22]:

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

where $\boldsymbol{W} = (W_{ij})$ is an antisymmetric matrix and $\boldsymbol{C} = (C_{ij})$ is a symmetric matrix. Throughout this paper, we denote by $(\boldsymbol{a}, \boldsymbol{b})$ the inner product of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

The matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ are not independent of each other. If the non-diagonal elements of the matrix $\boldsymbol{W}$ are rearranged in the form

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

the following relationship holds [5]:

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

We call this the *decomposability condition*.

The epipolar equation (2) and the decomposability condition (4) together constitute a necessary and sufficient condition for the flow to be induced on the image plane by an infinitesimal motion of a moving camera with varying zoom relative to a stationary scene [5, 10, 19]. It can be shown that if we know the matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ to within a common scale factor, we can compute the 3-D motion of the camera and the focus, hence the 3-D structure of the scene, in an analytically closed form [5]. Our main concern here is the question of how these matrices may best be estimated in the first place.

Estimating the matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ is equivalent to estimating their sum

$$\boldsymbol{F} = \boldsymbol{W} + \boldsymbol{C}, \tag{5}$$

since $\boldsymbol{W}$ and $\boldsymbol{C}$ are given by

$$\boldsymbol{W} = A[\boldsymbol{F}], \quad \boldsymbol{C} = S[\boldsymbol{F}], \tag{6}$$

where $A[\,\cdot\,]$ and $S[\,\cdot\,]$ denote, respectively, the antisymmetrization and symmetrization operations defined by $A[\boldsymbol{A}] = (\boldsymbol{A} - \boldsymbol{A}^\top)/2$ and $S[\boldsymbol{A}] = (\boldsymbol{A} + \boldsymbol{A}^\top)/2$.

The matrix $\boldsymbol{F}$ plays a similar role to the fundamental matrix for finite motion images [6, 9, 10, 23]. In fact, the epipolar equation (2) can be derived from the epipolar equation for finite motion images in the limit of infinitesimal motion [10]. For this reason, we call $\boldsymbol{F}$ the *flow fundamental matrix*. The decomposability condition (4) corresponds to the well known constraint on the finite motion fundamental matrix that it should have rank 2.

From the epipolar equation (2), we observe that the matrices $\boldsymbol{W}$ and $\boldsymbol{C}$ have scale indeterminacy and hence the scale of $\boldsymbol{F}$ is also indeterminate. So, we adopt the normalization $\|\boldsymbol{F}\| = 1$, where the norm of a matrix $\boldsymbol{A} = (A_{ij})$ is defined by $\|\boldsymbol{A}\| = \sqrt{\sum_{i,j=1}^{3} A_{ij}^2}$. In terms of the flow fundamental matrix $\boldsymbol{F} = (F_{ij})$, the decomposability condition (4) is expressed in the following form:

$$\sum_{i,j,k,l,m,n=1}^{3} \varepsilon_{ikl}\varepsilon_{jmn} F_{ij} F_{kl} F_{mn} = 0. \tag{7}$$

Here, ε_{ijk} is the Eddington epsilon, taking values 1 and -1 if (ijk) is an even and odd permutation of (123), respectively, and 0 otherwise.

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

3 Statistical Model of Noise

In this section, we state our assumptions about the nature of the image noise, on which the subsequent statistical analysis is based.

Let $\bar{\dot{\boldsymbol{x}}}_\alpha$ and $\bar{\boldsymbol{x}}_\alpha$ be the true values of $\dot{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}_\alpha$, respectively, in the absence of noise. We write

$$\dot{\boldsymbol{x}}_\alpha = \bar{\dot{\boldsymbol{x}}}_\alpha + \Delta\dot{\boldsymbol{x}}_\alpha, \qquad \boldsymbol{x}_\alpha = \bar{\boldsymbol{x}}_\alpha + \Delta\boldsymbol{x}_\alpha \tag{8}$$

and regard $\Delta\dot{\boldsymbol{x}}_\alpha$ and $\Delta\boldsymbol{x}_\alpha$ as independent Gaussian random variables of mean $\boldsymbol{0}$ but not necessarily isotropic or homogeneous. Let $V[\dot{\boldsymbol{x}}_\alpha]$ and $V[\boldsymbol{x}_\alpha]$ be the covariance matrices of vectors $\dot{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}_\alpha$, respectively. In practice, we need not know their absolute values; it suffices to know them *up to a common scale factor*. So, we write

$$V[\dot{\boldsymbol{x}}_\alpha] = \epsilon^2 V_0[\dot{\boldsymbol{x}}_\alpha], \qquad V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha], \tag{9}$$

and assume that $V_0[\dot{\boldsymbol{x}}_\alpha]$ 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. Since the third components of $\dot{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}_\alpha$ are constants, the matrices $V_0[\dot{\boldsymbol{x}}_\alpha]$ and $V_0[\boldsymbol{x}_\alpha]$ are singular with third columns and third rows filled with zeros.

The normalized covariance matrices can be estimated from the residual of template matching or optical flow detection [7, 14, 20, 21]. If the noise can be assumed to have the same isotropic distribution for all (x_α, y_α) and (x'_α, y'_α), $\alpha = 1, ..., N$, we have

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

where $\mathrm{diag}(\cdots)$ signifies the usual diagonal matrix. We use these as the default values when no information is available about the noise behavior.

Our goal is to estimate the flow fundamental matrix $\boldsymbol{F}$ that satisfies the epipolar equation for the true values $\{\bar{\dot{\boldsymbol{x}}}_\alpha,\ \bar{\boldsymbol{x}}_\alpha\}$ from their noisy observations $\{\dot{\boldsymbol{x}}_\alpha,\ \boldsymbol{x}_\alpha\}$.

4 Theoretical Accuracy Bound

In this section, we present a theoretical bound on the accuracy with which the flow fundamental matrix $\boldsymbol{F}$ may be estimated, under the statistical model of image noise described in the preceding section.

Let $\hat{\boldsymbol{F}}$ be an estimate of the flow fundamental matrix. Regarded as a function of the random variables $\dot{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}_\alpha$, the estimate $\hat{\boldsymbol{F}}$ is also a random variable based on the same probability space as that upon which $\dot{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}_\alpha$ are based. Let $\bar{\boldsymbol{F}}$ be the true value of the flow fundamental matrix. 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 $E[\,\cdot\,]$ denotes expectation. The operator $\otimes$ denotes the tensor product: for matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$, the $(ijkl)$ element of $\boldsymbol{A} \otimes \boldsymbol{B}$ is $A_{ij}B_{kl}$. For tensors $\mathcal{P} = (P_{ijkl})$ and $\mathcal{T} = (T_{ijkl})$, the matrix 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

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

Here, δ_{ij} is the Kronecker delta, taking 1 for $i = j$ and 0 otherwise. Define the *moment tensor* $\bar{\mathcal{M}} = (\bar{M}_{ijkl})$ by

$$\bar{M}_{ijkl} = \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(\frac{\bar{x}_{\alpha(i)}\dot{\bar{x}}_{\alpha(j)} - \bar{x}_{\alpha(j)}\dot{\bar{x}}_{\alpha(i)}}{2} + \bar{x}_{\alpha(i)}\bar{x}_{\alpha(j)}\Big)\Big(\frac{\bar{x}_{\alpha(k)}\dot{\bar{x}}_{\alpha(l)} - \bar{x}_{\alpha(l)}\dot{\bar{x}}_{\alpha(k)}}{2} + \bar{x}_{\alpha(k)}\bar{x}_{\alpha(l)}\Big), \tag{13}$$

where $\dot{\bar{x}}_{\alpha(i)}$ and $\bar{x}_{\alpha(i)}$ are the ith components of $\dot{\bar{\boldsymbol{x}}}_\alpha$ and $\bar{\boldsymbol{x}}_\alpha$, respectively. The scalar $\bar{W}_\alpha$ is defined by

$$\bar{W}_\alpha = \frac{1}{(\bar{\boldsymbol{W}}\bar{\boldsymbol{x}}_\alpha, V_0[\dot{\boldsymbol{x}}_\alpha]\bar{\boldsymbol{W}}\bar{\boldsymbol{x}}_\alpha) + (\bar{\boldsymbol{W}}\dot{\bar{\boldsymbol{x}}}_\alpha + 2\bar{\boldsymbol{C}}\bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha](\bar{\boldsymbol{W}}\dot{\bar{\boldsymbol{x}}}_\alpha + 2\bar{\boldsymbol{C}}\bar{\boldsymbol{x}}_\alpha))}\ , \tag{14}$$

where $\bar{\boldsymbol{W}}$ and $\bar{\boldsymbol{C}}$ are the true values of the matrices $\boldsymbol{W}$ and $\boldsymbol{C}$. According to the general theory of Kanatani [10], the lower bound on the covariance tensor $\mathcal{V}[\hat{\boldsymbol{F}}]$ is given by

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

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 as follows:

$$P^{\mathcal{S}}_{ijkl} = \delta_{ik}\delta_{jl} - \frac{\bar{K}_{ij}\bar{K}_{kl}}{\|\bar{\boldsymbol{K}}\|^2}. \tag{16}$$

Here $\bar{\boldsymbol{K}} = (\bar{K}_{ij})$ is defined by

$$\bar{K}_{ij} = \sum_{k,l,m,n=1}^{3} (\varepsilon_{ikl}\varepsilon_{jmn}\bar{F}_{kl} + \varepsilon_{ijk}\varepsilon_{lmn}(\bar{F}_{kl} + \bar{F}_{lk}))\bar{F}_{mn}. \tag{17}$$

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-vector [10].

A tensor $\mathcal{T} = (T_{ijkl})$ is said to be *symmetric* if $T_{ijkl} = T_{klij}$. A symmetric $3\times3\times3\times3$ 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:

$$\mathcal{T}_r^- = \sum_{i=1}^{r} \frac{\boldsymbol{U}_i \otimes \boldsymbol{U}_i}{\lambda_i}. \tag{18}$$

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

$$\mathrm{rms}(\hat{\boldsymbol{F}}) = \sqrt{E[\|\mathcal{P}(\hat{\boldsymbol{F}} - \bar{\boldsymbol{F}})\|^2]}. \tag{19}$$

Since $\mathcal{P}$ is a projection, it follows that $0 \leq \mathrm{rms}(\hat{\boldsymbol{F}}) \leq 1$. From eq. (15), we may infer the lower bound

$$\mathrm{rms}(\hat{\boldsymbol{F}}) \geq \sqrt{\mathrm{tr}\mathcal{V}[\hat{\boldsymbol{F}}]}, \tag{20}$$

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

$$\mathrm{tr}\mathcal{T} = \sum_{k,l=1}^{3} T_{klkl}. \tag{21}$$

5 Algorithm

In this section, we describe our algorithm for computing the flow fundamental matrix. The algorithm has two stages. First, we compute the matrix $\boldsymbol{F}$ without considering the decomposability condition by applying a technique called *renormalization* [10]. In the second stage, we impose the decomposability condition on the resulting $\boldsymbol{F}$ in a statistically optimal manner.

5.1 Renormalization

The renormalization computation algorithm proceeds as follows:

1. Let $c = 0$, $W_\alpha = 1$, $\alpha = 1, ..., N$, and $J = \infty$.
   % In real computation, the symbol ∞ means a very large number, e.g., 10^{10}.

2. Compute the tensors $\mathcal{M} = (M_{ijkl})$ and $\mathcal{N} = (N_{ijkl})$ defined by

$$\begin{aligned} M_{ijkl} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \Big(\frac{x_{\alpha(i)}\dot{x}_{\alpha(j)} - x_{\alpha(j)}\dot{x}_{\alpha(i)}}{2} + x_{\alpha(i)}x_{\alpha(j)}\Big)\Big(\frac{x_{\alpha(k)}\dot{x}_{\alpha(l)} - x_{\alpha(l)}\dot{x}_{\alpha(k)}}{2} \\ + x_{\alpha(k)}x_{\alpha(l)}\Big), \end{aligned} \tag{22}$$

$$\begin{aligned} N_{ijkl} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \Big(\frac{1}{4}(V_0[\boldsymbol{x}_\alpha]_{ik}\dot{x}_{\alpha(j)}\dot{x}_{\alpha(l)} - V_0[\boldsymbol{x}_\alpha]_{il}\dot{x}_{\alpha(j)}\dot{x}_{\alpha(k)} - V_0[\boldsymbol{x}_\alpha]_{jk}\dot{x}_{\alpha(i)}\dot{x}_{\alpha(l)} \\ + V_0[\boldsymbol{x}_\alpha]_{jl}\dot{x}_{\alpha(i)}\dot{x}_{\alpha(k)}) + \frac{1}{2}(V_0[\boldsymbol{x}_\alpha]_{ik}x_{\alpha(l)}\dot{x}_{\alpha(j)} + V_0[\boldsymbol{x}_\alpha]_{il}x_{\alpha(k)}\dot{x}_{\alpha(j)} \\ - V_0[\boldsymbol{x}_\alpha]_{jk}x_{\alpha(l)}\dot{x}_{\alpha(i)} - V_0[\boldsymbol{x}_\alpha]_{jl}x_{\alpha(k)}\dot{x}_{\alpha(i)} + V_0[\boldsymbol{x}_\alpha]_{ik}x_{\alpha(j)}\dot{x}_{\alpha(l)} \\ - V_0[\boldsymbol{x}_\alpha]_{il}x_{\alpha(j)}\dot{x}_{\alpha(k)} + V_0[\boldsymbol{x}_\alpha]_{jk}x_{\alpha(i)}\dot{x}_{\alpha(l)} - V_0[\boldsymbol{x}_\alpha]_{jl}x_{\alpha(i)}\dot{x}_{\alpha(k)}) \\ + \frac{1}{4}(V_0[\dot{\boldsymbol{x}}_\alpha]_{jl}x_{\alpha(i)}x_{\alpha(k)} - V_0[\dot{\boldsymbol{x}}_\alpha]_{jk}x_{\alpha(i)}x_{\alpha(l)} - V_0[\dot{\boldsymbol{x}}_\alpha]_{il}x_{\alpha(j)}x_{\alpha(k)} \\ + V_0[\dot{\boldsymbol{x}}_\alpha]_{ik}x_{\alpha(j)}x_{\alpha(l)}) + V_0[\boldsymbol{x}_\alpha]_{ik}x_{\alpha(j)}x_{\alpha(l)} + V_0[\boldsymbol{x}_\alpha]_{il}x_{\alpha(j)}x_{\alpha(k)} \\ + V_0[\boldsymbol{x}_\alpha]_{jk}x_{\alpha(i)}x_{\alpha(l)} + V_0[\boldsymbol{x}_\alpha]_{jl}x_{\alpha(i)}x_{\alpha(k)}\Big). \end{aligned} \tag{23}$$

   % $\dot{x}_{\alpha(i)}$ and $x_{\alpha(i)}$ are the ith components of $\dot{\boldsymbol{x}}_\alpha$ and $\boldsymbol{x}_\alpha$, respectively, and $V_0[\dot{\boldsymbol{x}}_\alpha]_{ij}$ and $V_0[\boldsymbol{x}_\alpha]_{ij}$ are the (ij) elements of $V_0[\dot{\boldsymbol{x}}_\alpha]$ and $V_0[\boldsymbol{x}_\alpha]$, 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. Perform the following computation:

- Update c as follows:

$$c \leftarrow c + \frac{\lambda_9}{(\boldsymbol{F}_9, \mathcal{N}\boldsymbol{F}_9)}. \tag{24}$$

- Compute W_α, $\alpha = 1, ..., N$, such that

$$W_\alpha = \frac{1}{(\boldsymbol{W}\boldsymbol{x}_\alpha, V_0[\dot{\boldsymbol{x}}_\alpha]\boldsymbol{W}\boldsymbol{x}_\alpha) + (\boldsymbol{W}\dot{\boldsymbol{x}}_\alpha + 2\boldsymbol{C}\boldsymbol{x}_\alpha, V_0[\boldsymbol{x}_\alpha](\boldsymbol{W}\dot{\boldsymbol{x}}_\alpha + 2\boldsymbol{C}\boldsymbol{x}_\alpha))}. \tag{25}$$

- Compute the tensors $\mathcal{M}$ and $\mathcal{N}$ given in eqs. (22) and (23).
- Let

$$J' \leftarrow J, \quad J \leftarrow (\boldsymbol{F}_9, \mathcal{M}\boldsymbol{F}_9). \tag{26}$$

- If $J' < J$, then let $J \leftarrow J'$. Else, compute the nine eigenvalues $\lambda_1 \geq \cdots \geq \lambda_9$ of the tensor

$$\hat{\mathcal{M}} = \mathcal{M} - c\mathcal{N} \tag{27}$$

  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 convergence.

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

The above procedure is based on the fact that, in the absence of noise, $\bar{\boldsymbol{F}}$ is the eigenmatrix of the moment tensor $\mathcal{M}$ defined by eq. (22) with eigenvalue 0. It can be shown that, in the presence of noise, the moment tensor $\mathcal{M}$ is statistically biased from its true value, to a first-order approximation, by a constant times the tensor $\mathcal{N}$ defined by eq. (23). So, using eq. (27), we iteratively remove the bias in $\mathcal{M}$ in such a way that the smallest eigenvalue of $\mathcal{M}$ converges to zero.

5.2 Optimal correction

We now apply a correction to $\boldsymbol{F}$, aiming to shift it, via an iterative scheme, to the "nearest" value that satisfies the decomposability condition. Our procedure will require as inputs the eigenvalues and eigenmatrices emerging from the previous scheme. The steps of the method are as follows:

1. Estimate the squared noise level $\hat{\epsilon}^2$ as follows:

$$\hat{\epsilon}^2 = \frac{(\boldsymbol{F}, \mathcal{M}\boldsymbol{F})}{1 - 8/N}. \tag{28}$$

2. Compute the normalized covariance tensor of $\boldsymbol{F}$ given by

$$\mathcal{V}_0[\boldsymbol{F}] = \frac{1}{N}\sum_{i=1}^{8} \frac{\boldsymbol{F}_i \otimes \boldsymbol{F}_i}{\lambda_i}. \tag{29}$$

3. Set

$$D(\boldsymbol{F}) = \sum_{i,j,k,l,m,n=1}^{3} \varepsilon_{ikl}\varepsilon_{jmn}F_{ij}F_{kl}F_{mn}. \tag{30}$$

4. Repeat the following computation until $D(\boldsymbol{F}) \approx 0$.

   - Compute the matrix $\boldsymbol{K} = (K_{ij})$ defined by

$$K_{ij} = \sum_{k,l,m,n=1}^{3} \Big(\varepsilon_{ikl}\varepsilon_{jmn} F_{kl} + \varepsilon_{ijk}\varepsilon_{lmn}(F_{kl} + F_{lk}) \Big) F_{mn}. \tag{31}$$

   - Update $\boldsymbol{F}$ as follows:

$$\boldsymbol{F} \leftarrow N[\boldsymbol{F} - \frac{D(\boldsymbol{F})\mathcal{V}_0[\boldsymbol{F}]\boldsymbol{K}}{(\boldsymbol{K}, \mathcal{V}_0[\boldsymbol{F}]\boldsymbol{K})}]. \tag{32}$$

     % Here $N[\,\cdot\,]$ denotes the normalization operator given by $N[\boldsymbol{F}] = \boldsymbol{F}/\|\boldsymbol{F}\|$.

   - Compute the projection tensor $\mathcal{P} = (P_{ijkl})$ such that

$$P_{ijkl} = \delta_{ik}\delta_{jl} - F_{ij}F_{kl}. \tag{33}$$

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

$$\mathcal{V}_0[\boldsymbol{F}]_{ijkl} \leftarrow \sum_{m,n,p,q=1}^{3} P_{ijmn}P_{klpq}\mathcal{V}_0[\boldsymbol{F}]_{mnpq}. \tag{34}$$

     % Since errors in $\boldsymbol{F}$ should be orthogonal to $\boldsymbol{F}$, their domain should be altered as $\boldsymbol{F}$ is updated.

Note that if we define the function $D(\boldsymbol{F})$ by eq. (30), the decomposability condition (7) is written as $D(\boldsymbol{F}) = 0$. It can be shown [10] that the estimate $\boldsymbol{F}$ resulting from the renormalization procedure has the normalized covariance tensor given in eq. (29). The above procedure enforces the decomposability condition $D(\boldsymbol{F}) = 0$ 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 shown that renormalization coupled with this type of correction produces a solution that attains the theoretical accuracy bound in the first order [10].

6 Epipoles and Epipolars

In this section, we give a theoretical bound on the accuracy of computing epipoles and epipolars. The *epipole* of the flow is defined to be the point represented by the vector $\boldsymbol{w}$ given by eq. (3):

$$\boldsymbol{x}_e = Z[\boldsymbol{w}]. \tag{35}$$

Here $Z[\,\cdot\,]$ denotes normalization to make the third component 1. This epipole is the position in the image pointed to by the camera translation velocity, but its physical orientation cannot be determined unless the internal camera parameters are given. The *epipolars* are defined to be the lines of the form

$$(\boldsymbol{n}_\alpha, \boldsymbol{x}) = 0, \qquad \boldsymbol{n}_\alpha = N[\boldsymbol{x}_\alpha \times \boldsymbol{x}_e]. \tag{36}$$

By definition, all the epipolars meet at the epipole $\boldsymbol{x}_e$. If the camera does not rotate, the flow vectors diverge or converge along the epipolars. If the camera rotates, the flow has a spiral pattern.

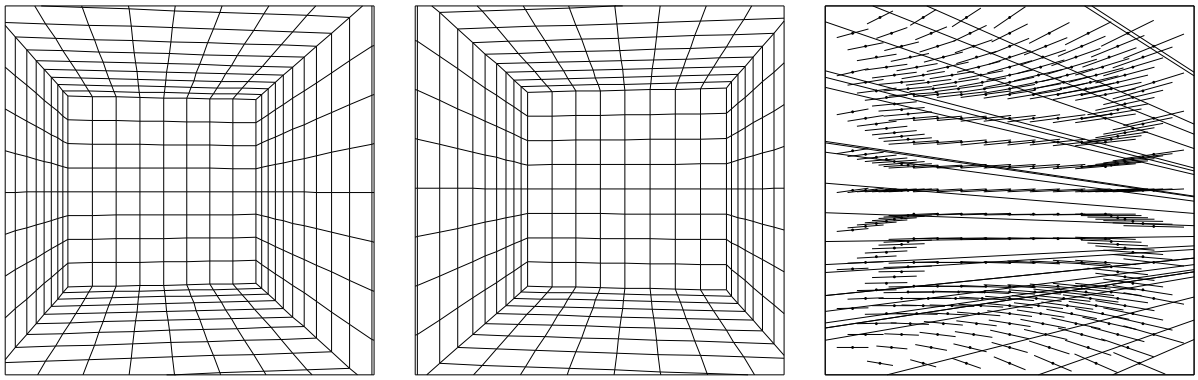


Figure 1: Simulated images of a 3-D scene and their flow and epipolars.

Since we know the lower bound on the covariance tensor $\mathcal{V}[\boldsymbol{F}]$, we have the lower bound on the covariance tensor of $\boldsymbol{W}$ $(= A[\boldsymbol{F}])$. Recalling the definition of $\boldsymbol{w}$ given in eq. (3), we see that the lower bound on the covariance matrix $V[\boldsymbol{w}]$ is given by

$$V[\boldsymbol{w}]_{ij} = \frac{1}{4} \sum_{k,l,m,n=1}^{3} \varepsilon_{ikl}\varepsilon_{jmn}\mathcal{V}[\boldsymbol{F}]_{klmn}. \tag{37}$$

Hence, the lower bound on the covariance matrix of the epipole $\boldsymbol{x}_e$ is given by

$$V[\boldsymbol{x}_e] = \frac{\boldsymbol{Q}_{\bar{\mathbf{x}}_e} V[\boldsymbol{w}] \boldsymbol{Q}_{\bar{\mathbf{x}}_e}^{\top}}{(\boldsymbol{k}, \bar{\boldsymbol{w}})^2}, \tag{38}$$

where the bars denote the true values and the projection matrix $\boldsymbol{Q}$ is defined by

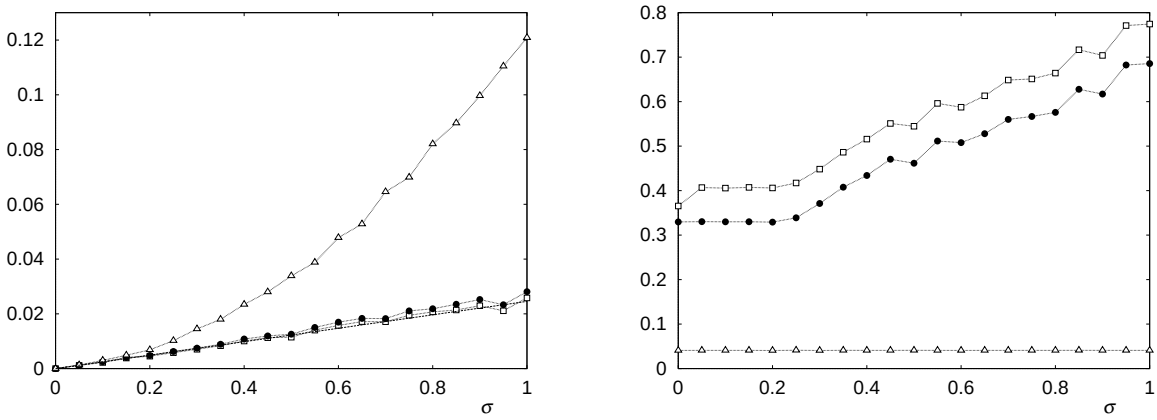
$$\boldsymbol{Q}_{\mathbf{x}} = \boldsymbol{I} - \boldsymbol{x}\boldsymbol{k}^{\top}. \tag{39}$$

Here $\boldsymbol{k} = (0, 0, 1)^{\top}$ is a unit vector in the direction of the optical axis of the camera.

7 Simulated Image Examples

We now present results of simulations to demonstrate that our algorithm indeed delivers a solution in the vicinity of the theoretical accuracy bound.

Fig. 1 shows two synthetic images (512 × 512 pixels) of a 3-D environment made up of grid patterns. These images are supposedly captured by a moving camera with a varying focal length. The figure also shows the induced optical flow vectors that pass through the midpoints of the corresponding points. Some of the epipolars 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 flow fundamental matrix was computed by using the default noise model of eqs. (10). The renormalization process 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. (33). The symbol □ denotes solutions obtained via the method presented in this paper, and the dotted line indicates the theoretical lower bound



(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).

derived from eq. (20). The symbol ● denotes renormalization solutions without applying the optimal correction of Section 5.2. The symbol △ denotes solutions computed by the least-squares method (*algebraic distance minimization*) that minimizes the expression

$$\frac{1}{N}\sum_{\alpha=1}^{N}\Big((\boldsymbol{x}_\alpha, \boldsymbol{W}\dot{\boldsymbol{x}}_\alpha) + (\boldsymbol{x}_\alpha, \boldsymbol{C}\boldsymbol{x}_\alpha)\Big)^2, \tag{40}$$

obtained directly from the epipolar equation (2). Applying this method, rather than our own, amounts to removing the iterations of Step 4 in the renormalization procedure of Section 5.1.

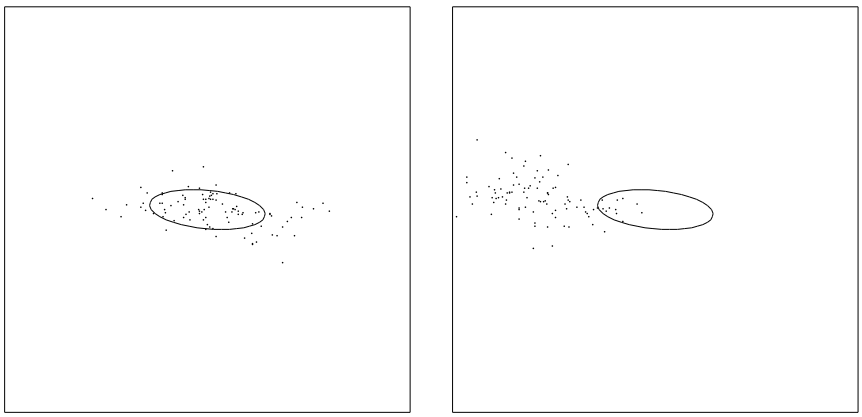
As we can see from Fig. 2(a), the errors in our estimates practically fall on the theoretical lower bound. Fig. 2(b) shows the average computation time on a Sun Ultra-30 workstation (SunOS 5.6). Naturally, the method takes more time than the least-squares method, but the theoretical accuracy bound is attained only at this higher computational cost.

Fig. 3(a) shows an enlargement near the epipole (outside the image frame), in which we plot the epipoles computed 100 times using different noise each time for $\sigma = 0.5$ (pixels). The ellipse depicts the lower bound on the standard deviation in each orientation. It is computed from the covariance matrix $V[\boldsymbol{x}_\epsilon]$ given in eq. (38) and is centered on the true epipole $\bar{\boldsymbol{x}}_\epsilon$. Again, the indications are that the errors in our estimates conform to the theoretical lower bound. Fig. 3(b) shows the corresponding result for the least-squares solutions; it is evident that the distribution has a large statistical bias.

The optimality of our algorithm implies that we can evaluate the reliability of our solution by simply computing the right-hand side of eq. (15). Since it is expressed in terms of the true values $\{\dot{\bar{\boldsymbol{x}}}_\alpha, \bar{\boldsymbol{x}}_\alpha\}$, $\bar{\boldsymbol{W}}$, $\bar{\boldsymbol{C}}$, and ϵ^2, we replace them with the data $\{\dot{\boldsymbol{x}}_\alpha, \boldsymbol{x}_\alpha\}$ and the computed estimates $\hat{\boldsymbol{W}}$ and $\hat{\boldsymbol{C}}$. The square noise level ϵ^2 is estimated by eq. (28).

8 Real Image Examples

Here, we show two real image examples of contrasting natures. One uses large displacements of salient feature points; the other uses optical flow detected by the conventional method based on



(a) (b)

Figure 3: Reliability of the epipole. The ellipses indicate the theoretical lower bound on the standard deviation in each orientation. (a) Our solutions. (b) Least-squares solutions.

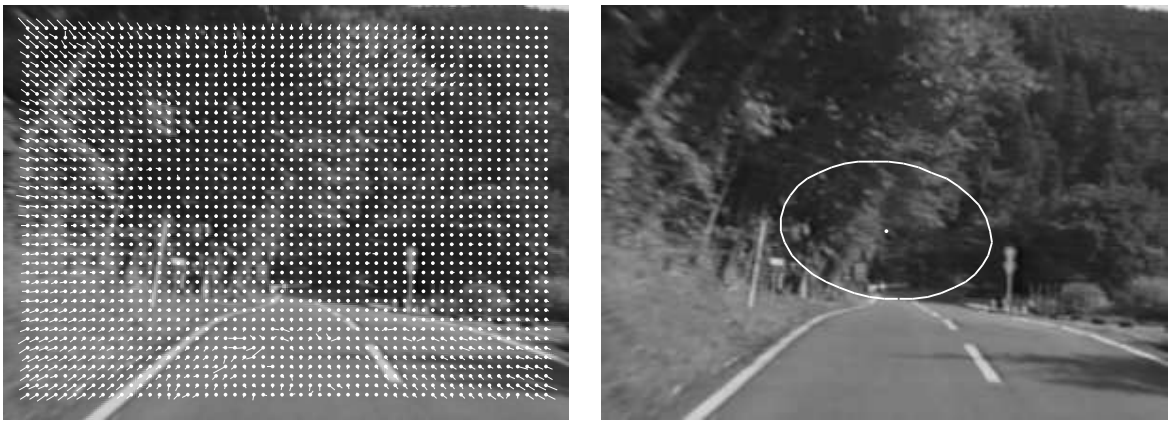
Figure 4: Real images of an indoor scene, and the epipoles and epipolars.

the gradient constraint.

Fig. 4 shows two images (768 × 512 pixels) of an indoor scene taken by a moving camera with varying zoom. Feature points were located manually, and the flow fundamental matrix was computed by using the default noise model of eqs. (10). The computed epipoles and epipolars are superimposed over the extracted flow. The inherent reliability of the epipole is 12.8 pixels, which is obtained by computing $\sqrt{\mathrm{tr}V[\boldsymbol{x}_e]}$ after substituting the data and the computed solution into eqs. (38).

As mentioned earlier, optical flow computed by the conventional gradient constraint method is unlikely to be sufficiently accurate for 3-D reconstruction if the inter-frame difference is very small or the scene consists of patches of almost uniform gray levels. Nevertheless, our technique can be applied to flow of this kind if more readily obtained information, such as the locations of epipoles, is to be extracted.

Fig. 5(a) shows optical flow computed from two consecutive images (300 × 220 pixels) of a road scene viewed from a vehicle moving forward. The flow was detected using the standard method based on the gradient constraint [16]; the flow components are displayed every 5 pixels. The covariance matrix of each flow component was evaluated from the residual of fitting the gradient constraint to the gray levels [16, 19]. The flow fundamental matrix was computed from



(a) (b)

Figure 5: (a) Optical flow computed from two consecutive images of a road scene. (b) The estimated epipole and its uncertainty ellipse.

this flow by using the covariance matrices thus evaluated. Fig. 5(b) shows the computed epipole and its uncertainty ellipse superimposed in the original image.

The detected flow is very small and not very accurate because a large portion of the scene, including the road surface, is of almost uniform gray levels. Hence, the computed epipole has a relatively large uncertainty as compared with that in Fig. 4, which was determined by large displacements of salient feature points. The important factor is, however, that our method enables us to evaluate the uncertainty of our computation in quantitative terms: the ellipse in Fig. 5(a) indicates the theoretical lower bound on the covariance, so the epipole estimation cannot be improved any further for this particular flow. This type of evaluation is crucial in practical applications involving optical flow because the accuracy of optical flow varies dramatically from situation to situation.

9 Program Package

The algorithm described in this paper is implemented in C++ and is placed on our Web page[2] as a public-domain program. It outputs a solution $\hat{\boldsymbol{F}}$ along with its *standard deviation pair* $\{\boldsymbol{F}^{(+)}, \boldsymbol{F}^{(-)}\}$. These two matrices 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 are most likely to occur. Let $\lambda_{\max}$ be the maximum eigenvalue of the tensor on the right-hand side of eq. (15) computed after substituting the data and the estimate $\hat{\boldsymbol{F}}$ for their true values, and let $\boldsymbol{U}_{\max}$ be the corresponding eigenmatrix of unit norm. The primary deviation pair is defined by

$$\boldsymbol{F}^{(+)} = N[\hat{\boldsymbol{F}} + \sqrt{\lambda_{\max}}\boldsymbol{U}_{\max}], \quad \boldsymbol{F}^{(-)} = N[\hat{\boldsymbol{F}} - \sqrt{\lambda_{\max}}\boldsymbol{U}_{\max}]. \tag{41}$$

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. Thus, the reliability of the computed solution is evaluated in quantitative terms.

The flow 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

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

feature points are on a special quadric called a *critical surface*, a typical instance of which is a planar surface [9]. If $\lambda_{\max}$ in eqs. (41) is predicted to be approximately 1 in the course of computation, the program judges that degeneracy has occurred and stops the computation after issuing a warning message.

10 Concluding Remarks

An algorithm is presented for computing the flow fundamental matrix from a set of corresponding points over two images in the presence of independent Gaussian noise that may be anisotropic and inhomogeneous. The problem formulation is stated optimally relative to the assumed noise model. The computation is divided into two stages: estimation of the flow fundamental matrix $\boldsymbol{F}$ by renormalization without considering the decomposability condition and correction of $\boldsymbol{F}$ in order to impose the condition. Experiments indicate that the obtained estimates are in the vicinity of the theoretical accuracy bound. Also, the reliability of the computed solution is evaluated in quantitative terms. This evaluation process is illustrated by computing the epipoles and epipolars from simulated and real images. Our algorithm is implemented as a public-domain program in C++.

Acknowledgments

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

References

[1] A. Bab-Hadiashar and D. Suter, "Robust optic flow computation," *Int. J. Comput. Vision*, 29(1), 59–77 (1998).

[2] J. L. Barron, D. J. Fleet and S. S. Beachemin, "Performance of optical flow techniques," *Int. J. Comput. Vision*, 12(1), 43–77 (1994).

[3] M. J. Black and A. D. Jepson, "Estimating optical flow in segmented images using variable-order parametric models with local deformations," *IEEE Trans. Patt. Analysis Intell.*, 18(10), 972–786 (1996).

[4] J. W. Brandt, "Improved accuracy in gradient-based optical flow estimation," *Int. J. Comput. Vision*, 25(1), 1–22 (1997).

[5] 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).

[6] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, MA, U.S.A., 1993.

[7] W. Förstner, "Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision," *Comput. Vision Graphics Image Process.*, 40, 273–310 (1987).

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

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

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

[11] S.-H. Lai and B. C. Vemuri, "Reliable and efficient computation of optical flow," *Int. J. Comput. Vision*, 29(2), 87–105 (1998).

[12] Q.-T. Luong and O. D. Faugeras, "Self-calibration of a moving camera from point correspondences and fundamental matrices," *Int. J. Comput. Vision*, 23(3), 261–289 (1997).

[13] A. Mitiche and P. Bouthemy, "Computation and analysis of image motion: A synopsis of current problems and methods," *Int. J. Comput. Vision*, 19(1), 29–55 (1996).

[14] D. D. Morris and T. Kanade, "A unified factorization algorithm for points, line segments and planes with uncertainty models," in *Proc. Int. Conf. Comput. Vision*, January 1998, Bombay, India, pp. 696–702.

[15] 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, 575–580 (1996).

[16] N. Ohta, "Image movement detection with reliability indices," *IEICE Trans.*, E79(10), 3379–3388 (1991).

[17] N. Ohta, "Optical flow detection using a general noise model," *IEICE Trans. Inf. & Syst.*, E79-D(7), 951–957 (1996).

[18] N. Ohta, "Uncertainty models of the gradient constraint for optical flow computation," *IEICE Trans. Inf. & Syst.*, E79-D(7), 958–964 (1996).

[19] N. Ohta and K. Kanatani, "Optimal structure-from-motion algorithm for optical flow," *IEICE Trans. Inf. & Syst.*, E78-D(12), 1559–1566 (1995).

[20] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vision Patt. Recogn.*, June 1994, Seattle, WA, U.S.A., pp. 593–600.

[21] A. Singh, "An estimation-theoretic framework for image-flow computation," in *Proc. 3rd Int. Conf. Comput. Vision*, December, 1990, Osaka, Japan, pp. 168–177.

[22] 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).

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

[24] L. Zhang, T. Sakurai and H. Miike, "Detection of motion fields under spatio-temporal non-uniform illumination," *Image Vision Comput.*, 17(3/4), 309–320 (1999).