International Journal of Computer Vision 66(2), 109–122, 2006

DOI: 10.1007/s11263-005-3952-y

Stabilizing the Focal Length Computation for 3-D Reconstruction from Two Uncalibrated Views

KENICHI KANATANI
Department of Computer Science, Okayama University, Okayama 700-8530 Japan,
kanatani@suri.it.okayama-u.ac.jp

ATSUTADA NAKATSUJI
Internet Terminals Division, NEC Engineering, Ltd., Yokohama, Kanagawa 224-0053 Japan,
a-nakatsuji@pb.jp.nec.com

YASUYUKI SUGAYA
Department of Computer Science, Okayama University, Okayama 700-8530 Japan,
sugaya@suri.it.okayama-u.ac.jp

Received March 16, 2004; Revised November 22, 2004; Accepted December 7, 2004

Abstract. In order to reconstruct 3-D shape from two uncalibrated views, one needs to resolve two problems: (i) the computed focal lengths can be imaginary; (ii) the computation fails for fixated images. We present a practical remedy for these by subsampling feature points and fixing the focal length. We first summarize theoretical backgrounds and then do simulations, which reveal a rather surprising fact that when the focal length is actually fixed, not using that knowledge yields better results for non-fixated images. We give an explanation to this seeming paradox and derive a hybrid method switching the computation by judging whether or not the images are fixated. Doing simulations and real image experiments, we demonstrate the effectiveness of our method.

Keywords: structure from motion, two-view analysis, focal length estimation, fundamental matrix, fixated images

1. Introduction

Many techniques have been proposed for reconstructing 3-D shape from images (Hartley and Zisserman, 2000). They are classified into two types: using separate images and using a continuous video stream. Among the former, the two-view method using two uncalibrated images (Hartley and Silpa-Anan, 2002; Kanatani and Ohta, 2003) is the simplest. Using three or more images may improve the accuracy, but a large amount of computation is necessary for matching multiple images and estimating the camera positions and their internal parameters for all the frames (Pollefeys et al., 1999). In contrast, the two-view method merely requires one to match feature points between the two images and compute the fundamental matrix. Today, effective algorithms are available for robustly matching two images (Kanazawa and Kanatani, 2002; Zhang et al., 1995) and for accurately computing the fundamental matrix (Chojnacki et al., 2000; Kanatani, 2000; Leedan and Meer, 2000), making the two-view method very practical for real applications.

However, this method has a serious drawback: since all the computations are based on the feature point matches over two images, the result is very sensitive

to the quality of the matches. In particular, the focal lengths for the two images are often computed to be imaginary (Hartley and Silpa-Anan, 2002) due to matching inaccuracies; wrong points may be matched, or the matched points may not exactly correspond to identical points in the scene.

On top of that, there is a problem of degeneracy: the computation fails if the two images are such that a point in the scene is fixated at their principal points (Brooks et al., 1998; Kanatani and Matsunaga, 2000). We call such an image pair *fixated images*. In order to do 3-D reconstruction, therefore, one must avert the camera from the object in a different way for each image. This is a great obstacle in practice, since for humans it is most natural to take images of something by fixating it.

This paper analyzes these problems in detail and presents a practical remedy. We first summarize the theoretical backgrounds. Then, we go on to practical issues. To avoid imaginary focal lengths, we subsample feature points. To cope with fixated images, we fix the focal length for the two images. It is known that 3-D reconstruction is possible even from fixated images if the two focal lengths are the same (Brooks et al., 1998; Kanatani and Matsunaga, 2000). Assuming a fixed focal length is not a serious constraint, since the focus and zooming are usually fixed in the course of taking pictures for 3-D reconstruction.

However, we reveal a rather surprising fact: when the focal length is actually fixed, *not using that knowledge yields better results if the images are not fixated*. We give an explanation to this seeming paradox. At the same time, we exploite this fact and derive a hybrid method switching the computation by judging whether or not the images are fixated. Doing simulations and real image experiments, we demonstrate the effectiveness of our method.

In Section 2, we introduce our assumptions and terminologies. Sections 3 and 4 summarize the theories for computing the focal lengths from the fundamental matrix. In Sections 5 and 6, we compare the variable and fixed focal length methods by simulations. We test our subsampling technique and conclude that when the focal length is fixed, not using that knowledge yields better results if the images are not fixated. In Section 7, we present a hybrid method switching the computation by testing fixation. Section 8 shows real image experiments to demonstrate the effectiveness of our method. Section 9 is our conclusion.

2. Geometry of Fixated Images

We assume that the camera skew angle is 0° and the aspect ratio is 1. Most digital cameras today seem to satisfy these conditions. If not, appropriate geometric correction is necessary, but we do not go into this issue here.

Heyden and Åström (1997) showed that if such a camera is used, the 3-D reconstruction is possible without knowing the focal length and the principal point location, but generally we need three or more images. Hartley (1992) showed that two images are sufficient if the principal point is given. We assume that the principle point is known (typically at the center of the image frame) and take it as the image coordinate origin. However, the focal length is assumed to be unknown.

If a point (x, y) in the first image corresponds to a point (x', y') in the second, we have the following constraint (Hartley and Zisserman, 2000):

$$(\boldsymbol{x}, \boldsymbol{F}\boldsymbol{x}') = 0. \tag{1}$$

Here, the points (x, y) and (x', y') are represented by 3-D vectors

$$\boldsymbol{x} = \begin{pmatrix} x/f_0 \\ y/f_0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{x}' = \begin{pmatrix} x'/f_0 \\ y'/f_0 \\ 1 \end{pmatrix}, \tag{2}$$

where f_0 (in pixels) is a scale factor[1] for stabilizing numerical computation. Throughout this paper, we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$. Equation (1) is called the *epipolar equation*; the matrix $\boldsymbol{F}$ is of rank 2 and called the *fundamental matrix* (Hartley and Zisserman, 2000).

We say that two images are *fixated* if the optical axes of the cameras that took these images intersect in the scene (Fig. 1). It follows that the origin of one image corresponds to the origin of the other. In the vector representation of eqs. (2), the origin $(0, 0)$ is represented by $\boldsymbol{k} = (0, 0, 1)^\top$. So, the condition for fixation is

$$(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k}) = 0, \tag{3}$$

or equivalently $F_{33} = 0$.

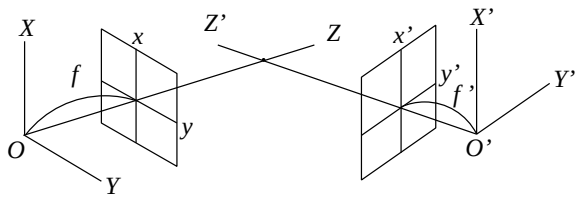


Figure 1. Two images are fixated if the optical axes intersect.

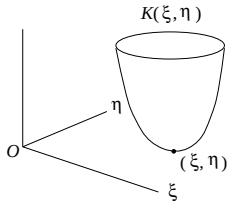


Figure 3. The focal lengths $\{f,\ f'\}$ are determined by the tangent point of the surface $K(\xi, \eta)$ to the $\xi\eta$-plane, at which $K(\xi, \eta)$ takes its minimum 0.

3. Variable Focal Length Method

We now summarize the procedures for computing the focal lengths $\{f,\ f'\}$ of the two cameras from the fundamental matrix $\boldsymbol{F}$.

It was Hartley (1992) who first presented an analytic procedure. He applied the singular value decomposition (SVD) and solved linear equations in four unknowns. Pan et al. (1995a, b) reduced this problem to solving cubic equations. Newsam et al. (1996) refined these algorithms into a combination of SVD and linear equations in three unknowns. They also derived the degeneracy condition for the solution to be indeterminate. Bougnoux (1998) presented a closed-form formula for f in terms of the fundamental matrix $\boldsymbol{F}$ and the *epipoles* $\boldsymbol{e}$ and $\boldsymbol{e}'$. Kanatani and Matsunaga (2000) and Ueshiba and Tomita (2003) presented alternative formulations. These analyses are mathematical equivalent, all indicating that the computation fails for fixated images. Here, we describe the procedure presented in Kanatani and Matsunaga (2000).

We first change the variables from $\{f,\ f'\}$ into $\{\xi,\ \eta\}$ by

$$\xi = \left(\frac{f_0}{f}\right)^2 - 1, \qquad \eta = \left(\frac{f_0}{f'}\right)^2 - 1. \tag{4}$$

Define the following polynomial $K(\xi, \eta)$ of order 4 in $\{\xi,\ \eta\}$ (order 2 in each):

$$\begin{aligned} K(\xi,\eta) = {} & (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^4\xi^2\eta^2 + 2(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2\|\boldsymbol{F}^\top\boldsymbol{k}\|^2\xi^2\eta \\ & +2(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2\|\boldsymbol{F}\boldsymbol{k}\|^2\xi\eta^2 + \|\boldsymbol{F}^\top\boldsymbol{k}\|^4\xi^2 \\ & +\|\boldsymbol{F}\boldsymbol{k}\|^4\eta^2 + 4(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k})\xi\eta \\ & +2\|\boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{k}\|^2\xi + 2\|\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k}\|^2\eta + \|\boldsymbol{F}\boldsymbol{F}^\top\|^2 \\ & -\frac{1}{2}\Big((\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2\xi\eta + \|\boldsymbol{F}^\top\boldsymbol{k}\|^2\xi \\ & +\|\boldsymbol{F}\boldsymbol{k}\|^2\eta + \|\boldsymbol{F}\|^2\Big)^2. \end{aligned} \tag{5}$$

Throughout this paper, we define the vector norm by $\|\boldsymbol{a}\| = \sqrt{\sum_{i=1}^3 a_i^2}$ for $\boldsymbol{a} = (a_i)$ and the matrix norm by $\|\boldsymbol{A}\| = \sqrt{\sum_{i,j=1}^3 A_{ij}^2}$ for $\boldsymbol{A} = (A_{ij})$.

The unknowns $\{\xi,\ \eta\}$ are determined from the following condition (see Appendix A for the proof):

$$K = \frac{\partial K}{\partial \xi} = \frac{\partial K}{\partial \eta} = 0. \tag{6}$$

This appears to be overspecification, providing three equations for two unknowns. It turns out, however, that the three equations are algebraically dependent, only two among them being independent (see Appendix B for the proof). Geometrically, the function $K(\xi, \eta)$ defines a locally nonnegative concave surface that is tangent to the $\xi\eta$-plane with minimum 0 (Fig. 2).

Since $K(\xi, \eta)$ is a polynomial of order 4, its minimum can be easily computed by Newton iterations. It is also possible to solve eq. (6) analytically (Kanatani and Matsunaga, 2000). Alternatively, we can derive various types of analytical formula via different approaches (Bougnoux, 1998; Hartley, 1992; Kanatani and Matsunaga, 2000; Ueshiba and Tomita, 2003). Among them, the simplest may be the following form (see Appendix C for the derivation):

$$\begin{aligned} \xi &= \frac{\|\boldsymbol{F}\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k})\|\boldsymbol{e}' \times \boldsymbol{k}\|^2/(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})}{\|\boldsymbol{e}' \times \boldsymbol{k}\|^2\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2}, \\ \eta &= \frac{\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k})\|\boldsymbol{e} \times \boldsymbol{k}\|^2/(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})}{\|\boldsymbol{e} \times \boldsymbol{k}\|^2\|\boldsymbol{F}\boldsymbol{k}\|^2 - (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2}. \end{aligned} \tag{7}$$

Here, $\boldsymbol{e}$ and $\boldsymbol{e}'$ are, respectively, the unit eigenvectors of $\boldsymbol{F}^\top$ and $\boldsymbol{F}$ for eigenvalue[2] 0; they represent the *epipoles* (Hartley and Zisserman, 2000), pointing from

the respective centers of projection to the centers of projection of the other images.

From eqs. (7), it is immediately seen that the computation fails for fixated images, for which $(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})$ vanishes, causing zero division. For non-fixated images, the focal lengths $\{f,\, f'\}$ are given from eqs. (4) in the form

$$f = \frac{f_0}{\sqrt{1+\xi}}, \qquad f' = \frac{f_0}{\sqrt{1+\eta}}. \tag{8}$$

However, if the computed fundamental matrix $\boldsymbol{F}$ is not accurate enough, the inside of one or both of the square roots can be negative, resulting in imaginary focal lengths (Hartley and Silpa-Anan, 2002).

4. Fixed Focal Length Method

We next consider the case when the focal lengths $\{f,\, f'\}$ are known to be equal. Mathematical analysis of this case was done in various different forms by Brooks et al. (1998), Kanatani and Matsunaga (2000), Sturm (2001), and Ueshiba and Tomita (2003), all indicating that the focal lengths can be computed from fixated images but that degeneracy occurs if the optical axes of the two cameras are parallel or if they form an isosceles triangle with the baseline as its base. Here, we describe the procedure presented in Kanatani and Matsunaga (2002).

Letting $\xi = \eta$ in eq. (5), we obtain the following polynomial $K(\xi)$ of order 4 in ξ:

$$K(\xi) = a_1\xi^4 + a_2\xi^3 + a_3\xi^2 + a_4\xi + a_5, \tag{9}$$

$$\begin{aligned}
a_1 &= \frac{1}{2}(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^4,\\
a_2 &= (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2(\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 + \|\boldsymbol{F}\boldsymbol{k}\|^2),\\
a_3 &= \frac{1}{2}(\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 - \|\boldsymbol{F}\boldsymbol{k}\|^2)^2 + (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})(4(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k})\\
&\quad -(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})\|\boldsymbol{F}\|^2),\\
a_4 &= 2(\|\boldsymbol{F}\boldsymbol{F}^\top\boldsymbol{k}\|^2 + \|\boldsymbol{F}^\top\boldsymbol{F}\boldsymbol{k}\|^2)\\
&\quad -(\|\boldsymbol{F}^\top\boldsymbol{k}\|^2 + \|\boldsymbol{F}\boldsymbol{k}\|^2)\|\boldsymbol{F}\|^2,\\
a_5 &= \|\boldsymbol{F}\boldsymbol{F}^\top\|^2 - \frac{1}{2}\|\boldsymbol{F}\|^4.
\end{aligned} \tag{10}$$

Equation (6) reduces to

$$K(\xi) = K'(\xi) = 0. \tag{11}$$

The solution is analytically obtained as follows (see Appendix D for the proof):

- If $a_1 \neq 0$,
  - if $3a_2^2 - 8a_1a_3 \neq 0$, compute the two solutions of the quadratic equation

$$(3a_2^2 - 8a_1a_3)x^2 + 2(a_2a_3 - 6a_1a_4)x + (a_2a_4 - 16a_1a_5) = 0. \tag{12}$$

    Let ξ be the one for which $|K(x)|$ is smaller;
  - if $3a_2^2 - 8a_1a_3 = 0$, let

$$\xi = -\frac{a_2a_4 - 16a_1a_5}{2(a_2a_3 - 6a_1a_4)}. \tag{13}$$

- If $a_1 = 0$ and $a_2 \neq 0$, let

$$\xi = -\frac{a_3a_4 - 9a_2a_5}{2(a_3^2 - 3a_2a_4)}. \tag{14}$$

- If $a_1 = a_2 = 0$ and $a_3 \neq 0$, let

$$\xi = -\frac{a_4}{2a_3}. \tag{15}$$

- If $a_1 = a_2 = a_3 = 0$, no solution exists.

The last is the degeneracy case: parallel optical axes or an isosceles triangle configuration (Kanatani and Matsunaga, 2000).

However, this analysis is based on the assumption that the fundamental matrix $\boldsymbol{F}$ is exact. Equation (11) gives two constraints on one variable ξ, but they are in general inconsistent if $\boldsymbol{F}$ is computed from noisy data. Geometrically, eq. (11) states that the solution is given by the position on the ξ-axis at which the curve $K(\xi)$ takes its minimum 0 (in the degeneracy case, the curve is "flattened" into a line (Kanatani and Matsunaga, 2000)). However, the minimum is in general positive (Fig. 3), because $K(\xi)$ is the cross section of the surface $K(\xi, \eta)$ in Fig. 2 with a plane perpendicular to the $\xi\eta$-plane passing through the line $\xi = \eta$. It follows that the minimum of $K(\xi)$ is 0 when and only when the minimum of $K(\xi, \eta)$ is on the line $\xi = \eta$. This condition is generally violated if $\boldsymbol{F}$ is not exact.

Ueshiba and Tomita (2003) analytically obtained a unique solution by regarding the two principal points as extra unknowns, assuming that the images are fixated. However, the camera must be rotated around

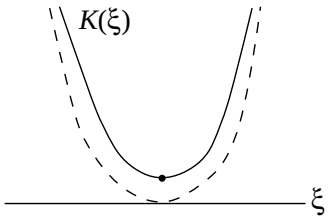


Figure 2. The focal length can be determined by the position at which the curve $K(\xi)$ takes its minimum 0. However, the minimum is generally positive.

the optical axis for the solution to exist. Also, their method cannot be applied to non-fixated images.

In order to avoid this difficulty, we compute the value ξ at which the curve $K(\xi)$ takes its minimum, i.e., we solve $K'(\xi) = 0$, although this does not produce an optimal solution (in the sense of maximum likelihood). In order to obtain an optimal solution, we need to recompute the fundamental matrix $\boldsymbol{F}$ subject to the constraint that the resulting $\boldsymbol{F}$ yield identical focal lengths by eqs. (7). At present, however, no such systematic computational scheme is known other than brute force search, and no criteria are available to avoid local minima. The purpose of this study is to find a practical compromise, so we adopt the above simple scheme.

Since $K'(\xi)$ is a cubic polynomial, the solution may be analytically obtained in theory. However, the computation branches depending on whether $a_1 \sim a_3$ are zero or not, and there is no good way to set a suitable threshold. This difficulty can be avoided by numerically computing the solution of $K'(\xi) = 0$ by Newton iterations:

$$\xi \leftarrow \xi - \frac{K'(\xi)}{K''(\xi)}. \tag{16}$$

We use eq. (15) as the initial value. Usually, two or three iterations are sufficient. From the computed ξ, the focal lengths $\{f, f'\}$ are given by eqs. (8), namely,

$$f = f' = \frac{f_0}{\sqrt{1+\xi}}. \tag{17}$$

In this case, too, the solution can be imaginary.

5. Variable vs. Fixed Focal Lengths

The focal lengths can be imaginary in the presence of noise whether we use the variable focal length method (which we hereafter abbreviate to the *v.f. method*) or the fixed focal length method (which we abbreviate to the *f.f. method*). But which is better *if the ground truth is* $f = f'$*?* We examined this by simulation.

The left of Fig. 4 shows two simulated images of a cylindrical grid surface. The image size is supposedly 600×800 pixels; the focal lengths are $\bar{f} = \bar{f}' = 1000$ (pixels). The right of Fig. 4 is the top view of the camera motion. We need to stay away from isosceles triangle camera configurations to avoide degeneracy. Yet, it is very natural for humans to keep the same depth while moving the camera. So, we assumed a nearly isosceles triangle configuration by setting the ratio of the depths to the object to 1 : 0.94. The baseline makes 70° with the optical axis for the first camera and 86.5° for the second camera. Then, the center of the second frame is displaced from its fixated position by d pixels, which we varied over $0 \leq d \leq 40$ (the left of Fig. 4 is for $d = 20$).

We added Gaussian noise of mean 0 and standard deviation σ (pixels) to the x and y coordinates of the 117 vertices independently. In order to simulate realistic situations, we randomly chose 10% of the vertices and increased the noise magnitude five times there[3]. From these noisy vertices, we computed the fundamental matrix; we used an algorithm called *renormalization*[4] (Kanatani, 2000; Kanatani and Ohta, 2003), which is known to be statistically optimal (Kanatani, 1996).

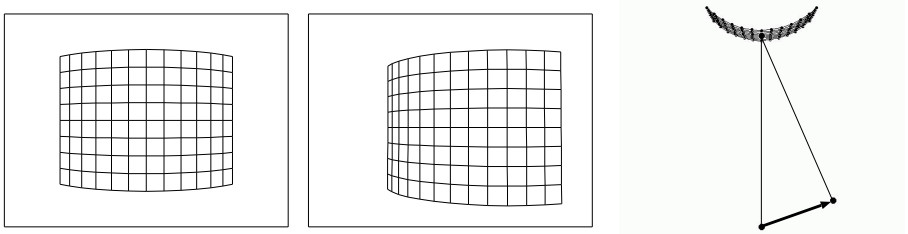


Figure 4. *Left*: Simulated images of a cylindrical grid surface (600×800 pixels). The focal lengths are $\bar{f} = \bar{f}' = 1000$ (pixels). The center of the second frame is displaced by d (= 20 for the images shown here) pixels from its fixated position. *Right*: Top view of the camera motion.

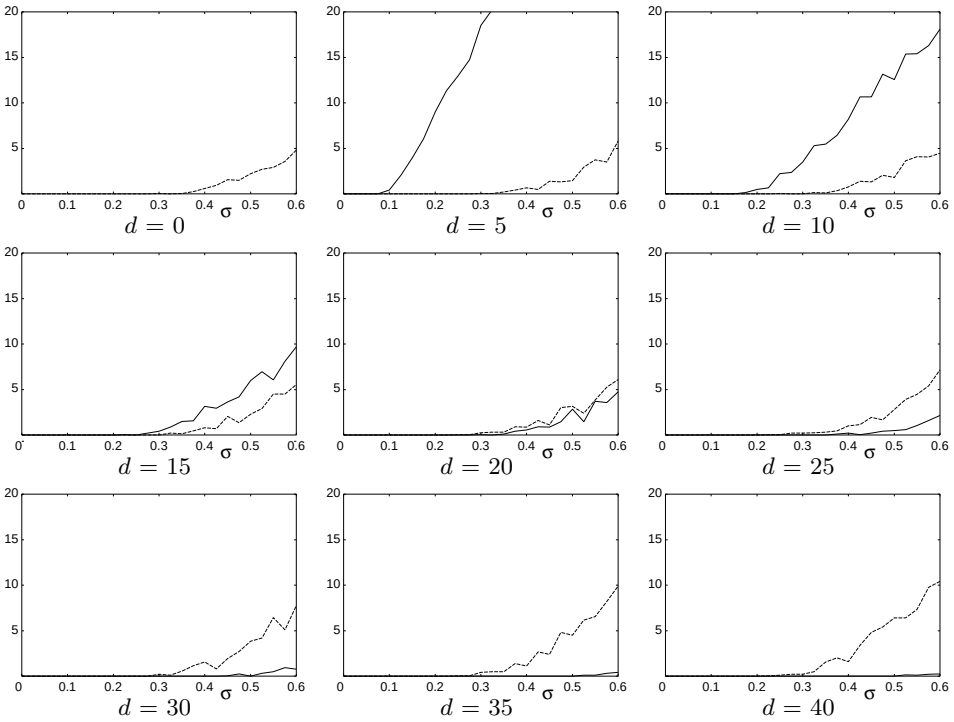


Figure 5. The horizontal axis is for the noise standard deviation σ (pixels). The vertical axis shows the percentage of the occurrences of imaginary focal lengths. The solid and dashed lines are for the v.f. and f.f. methods, respectively. The value d (pixels) measures the deviation from fixation.

Figure 5 plots the percentage of the occurrences of imaginary focal lengths over 2000 trials for each σ. The solid and dashed lines are for the v.f. and f.f. methods, respectively. At $d = 0$ (fixated images), the values for the v.f. method are not plotted, because they are out of the range; they are about 60% for all σ. As d increases, the percentage for the v.f. method drops, while it stays almost the same for the f.f. method. As a result, the relative order is reversed near $d = 20$.

6. Accuracy of the Focal Lengths

In order to see the comparative accuracy of the focal lengths, we need to avoid the occurrences of imaginary focal lengths.

To this end, Hartley and Silpa-Anan (2002) used the knowledge about the approximate focal length and its minimum value: they optimized the fundamental matrix and the principal points so that the computed focal lengths are close to each other, close to their predictions, and close to their minimum values and at the same time the principal point is close to the frame center. The result depends on the predictions we make and the measure of closeness we use.

Here, we adopt subsampling of feature points. If the computed focal lengths are not both real, we randomly remove one pair of corresponding feature points and recompute the fundamental matrix. If we fail to obtain real focal lengths for $N/10$ consecutive repetitions (N is the number of correspondences), we randomly remove two pairs and do the same. If this fails, we go on removing more pairs until real focal lengths are obtained. According to our experience, real focal lengths are usually obtained after a few such trials if not at the first trial. Of course, we may not be able to obtain any solution if the matches are very poor or completely wrong.

Many other strategies can be conceivable. For example, we may prefer those feature points that are close to their epipolar lines predicted by the fundamental matrix $\boldsymbol{F}$ computed in the preceding step. We tried such methods in many forms, but we were unable to find any method better than the above straightforward one.

Imaginary focal lengths occur if the values (ξ, η) computed by eqs. (7) are perturbed across the line ξ

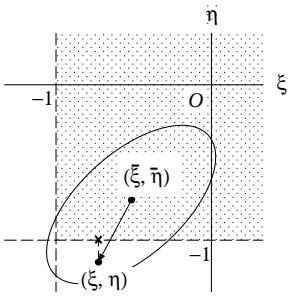


Figure 6. The dotted region is the admissible domain for (ξ, η). Imaginary focal lengths occur if the true values $(\bar{\xi}, \bar{\eta})$ are perturbed into (ξ, η) across the line $\xi = -1$ or the line $\eta = -1$. The ellipse schematically illustrates the uncertainty region.

$= -1$ or the line $\eta = -1$ due to the inaccuracy of the fundamental matrix $\boldsymbol{F}$. The difficulty is that the admissible domain $\xi > -1$ and $\eta > -1$ (the dotted region in Fig. 6) is topologically an *open* set: if it were closed, we could minimally correct (ξ, η) back onto the boundary line, which is not admissible, resulting in $f = \infty$ or $f' = \infty$.

We may be able to avoid this by empirically specifying a closed region within which we expect the solution to exist or, more generally, by introducing an *a priori* distribution of the solution (i.e., the Bayesian approach) as done by Hartley and Silpa-Anan (2002). However, it is very difficult to introduce such empirical constraints, on which the resulting solution strongly depends. So, we take the approach of generating another solution inside the admissible domain by subsampling feature points.

We evaluated the accuracy of the focal lengths computed by our method by the root-mean-square error

$$E = \sqrt{\frac{1}{2000}\sum_{a=1}^{1000}\Big((f_a - \bar{f})^2 + (f'_a - \bar{f}')^2\Big)} \qquad (18)$$

over 1000 trials, where f_a and f'_a are the values in the ath trial, and $\bar{f}$ and $\bar{f}'$ are their true values. We computed this for different σ and d, using the simulated images in Fig 4. Figure 7 shows the results corresponding to Fig. 5.

From this, we see that the focal lengths computed by the v.f. method from fixated images ($d = 0$) are meaningless, while the f.f. method can successfully compute fairly accurate values, as expected. How-

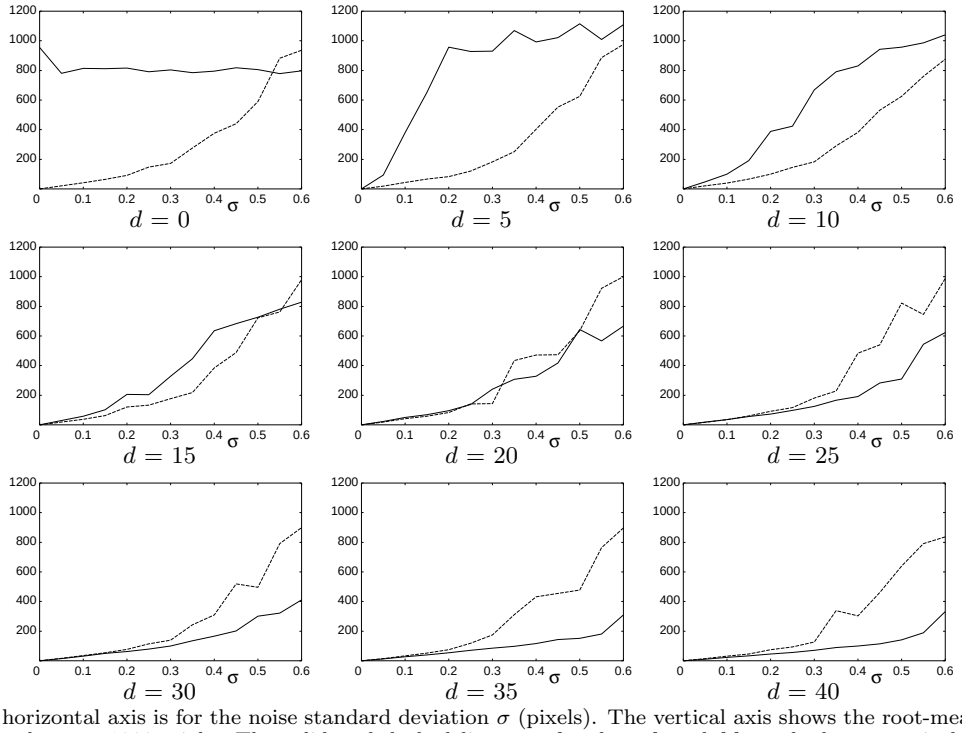


Figure 7. The horizontal axis is for the noise standard deviation σ (pixels). The vertical axis shows the root-mean-square error in the focal lengths over 1000 trials. The solid and dashed lines are for the v.f. and f.f. methods, respectively. The value d (pixels) measures the deviation from fixation.

ever, the v.f. method gradually gains in accuracy as d increases, while the f.f. method has almost the same accuracy. As a result, the relative accuracy is reversed around $d = 20$.

This is a rather surprising fact, since this means that when the focal lengths are known to be equal, *not using that knowledge yields better results* if d is large. This can be explained as follows. Recall that we first compute the fundamental matrix $\boldsymbol{F}$ by a statistically optimal method (in our case the renormalization method (Kanatani, 2000; Kanatani and Ohta, 2003)). However, the optimality is guaranteed only when no constraint is imposed other than $\det \boldsymbol{F} = 0$. If we *know* that the two focal lengths $\{f,\ f'\}$ are equal, the computed solution $\boldsymbol{F}$ is *no longer optimal*, because an optimal solution $\boldsymbol{F}$ should be computed subject to the constraint that it yield real and equal focal lengths. However, such a computation is very difficult in practice, as described in Section 4. So, we introduced a practical compromise, which led to a seemingly surprising result.

7. Hybrid Method

From the observation in the preceding section, we can expect high accuracy if we use the f.f. method when the images are nearly fixated and the v.f. method when they are not. Here, we adopt the following strategy.

In the first and second images are defined epipolar lines $(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{x}') = 0$ and $(\boldsymbol{x}, \boldsymbol{F}\boldsymbol{k}) = 0$ by the image origin of the other images. In terms of the image coordinates, they are

$$\begin{aligned} F_{13}x + F_{23}y + F_{33}f_0 &= 0, \\ F_{31}x' + F_{32}y' + F_{33}f_0 &= 0. \end{aligned} \tag{19}$$

If the images are fixated, the origins should be on these epipolar lines. So, the degree of fixation can be measured by the distances h and h' (pixels) of these lines from the origins:

$$h = \frac{|F_{33}|f_0}{\sqrt{F_{13}^2 + F_{23}^2}}, \qquad h' = \frac{|F_{33}|f_0}{\sqrt{F_{31}^2 + F_{32}^2}}. \tag{20}$$

We judge that the images are fixated if $h \leq h_c$ and $h' \leq h_c$ for a threshold h_c (pixels). This judgment is independent of the scale of $\boldsymbol{F}$ or the average magnitude of the errors in $\boldsymbol{F}$.

Many other switching schemes are conceivable. For example, we may conduct statistical hypothesis testing based on the covariance tensor of the computed fundamental matrix, which can be obtained as a byproduct of the renormalization computation (Kanatani, 2000; Kanatani and Ohta, 2003)), or introduce model selection using the geometric AIC or the geometric MDL (Kanatani, 1996, 1998, 2004). However, it is very difficult to compute these criteria precisely. If we introduce approximations or use estimates, the result is greatly influenced by the accuracy of the approximations and estimates we use. After trying many alternatives, we have concluded that the above simple criterion works the best.

In our experiments, we used the threshold $h_c = 20$ (pixels), partly because the relative accuracy of the v.f. and f.f. methods is reversed around $d = 20$ (pixels) in the simulation and partly because the deviation of about 20 pixels is inevitable if humans try to take fixated images manually. This threshold should be adjusted according to the image size, the image resolution, and the focal length. After many simulations using various image sizes, image resolutions, and focal lengths, we have found that in all cases the critical value is approximately 0.02 radians if measured in the angle of view .

Figure 8 is the simulation result using the data of Fig. 4. We incremented d from 0 (fixated images) to 40 for $\sigma = 0.3$ (pixels). The vertical axis is for the root-mean-square error E (pixels) in eq. (17). The solid and dashed lines are for the v.f. and f.f. methods, respectively; the dotted line is for the hybrid

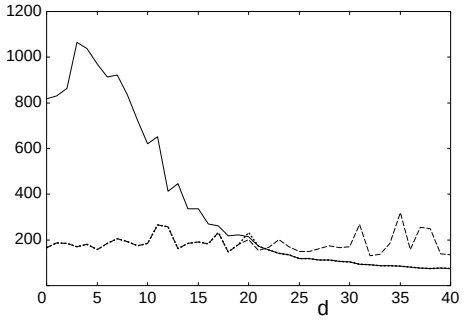
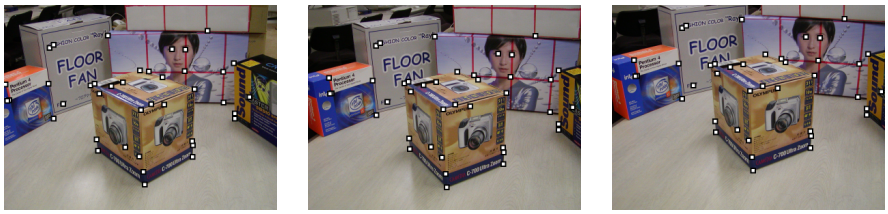


Figure 8. The horizontal axis is for the deviation d (pixels) from fixation. The vertical axis shows the root-mean square error in the computed focal lengths for the noise standard deviation $\sigma = 0.5$ (pixels). The solid and dashed lines are for the v.f. and f.f. methods, respectively; the dotted line is for the hybrid method.



(a) (b) (c)

Figure 9. Input images and selected feature points. The image pair (a) and (b) is fixated, while the image (c) is taken by slightly averting the optical axis.

method. We can see that the hybrid method adopts the f.f. method when d is small and switches to the v.f. method when d is large. The transition occurs around the value $d = 20$, to which the threshold for h and h' is set. As a result, the method with higher accuracy is automatically chosen irrespective of the value d.

8. Real Image Examples

In Fig. 9, the image pair (a) and (b) is fixated, while the image (c) is taken by slightly averting the optical axis. We chose 39 corresponding feature points as marked in the images. Algorithms for automatically detecting feature points and matching them are available (Kanazawa and Kanatani, 2002; Zhang et al., 1995), but mismatches are inevitable to some extent. Since our aim here is not to study the matching performance, we chose the feature points by hand.

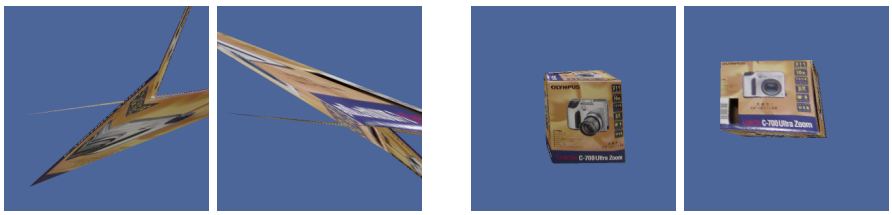
From the feature point matches marked in the images, we computed the fundamental matrix by renormalization and tested if the image pair (a) and (b) and the image pair (a) and (c) are fixated. In both cases, real focal lengths were obtained without subsampling. The computed values of h and h' are listed in Table 1; the image pair (a) and (b) is judged to be fixated, while the image pair (a) and (c) is not.

Table 1 also lists the focal lengths $\{f, f'\}$ computed using the two method. According to a simple calibration using a reference pattern, the true focal length is estimated to be $f = f' \approx 1000$ (pixels), from which the v.f. values computed from the fixated image pair (a) and (b) are wide apart, while the f.f. value seems reasonably good. For the non-fixated image pair (a) and (c), both methods estimate reasonable values, but the v.f. values are slightly better, in agreement with the simulation results.

Figure 10 shows the 3-D shape reconstructed from the fixated pair. As expected, the v.f. method produces a meaningless shape (Fig. 10(a)) because of the unrealistic focal lengths, while the f.f. method produces a fairly accurate shape (Fig. 10(b)). The two images in Fig. 10(b) are the front view (left figure) and the upper view (right figure) of the box; ide-

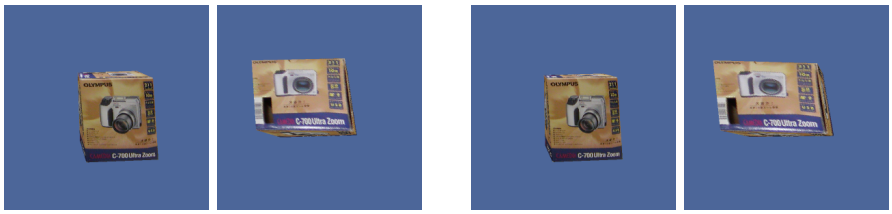
Table 1: Fixation test and focal length estimation for the image pair (a) and (b) and for the image pair (a) and (c) in Fig. 9. The unit is pixels.

	Fixation test		v.f.		f.f.
	h	h'	f	f'	$f = f'$
(a), (b)	1.03	1.03	436	443	811
(a), (c)	54.22	54.22	929	906	855



(a) (b)

Figure 10. 3-D reconstruction from fixated images: (a) v.f. method; (b) f.f. method.



(a) (b)

Figure 11. 3-D reconstruction from non-fixated images: (a) v.f. method; (b) f.f. method.

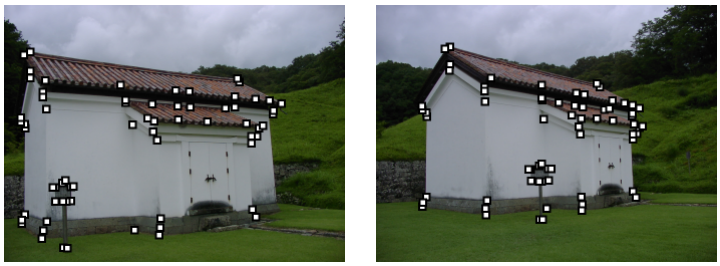


Figure 12. Input images and selected feature points.

(a) (b)

Figure 13. 3-D reconstruction: (a) v.f. method; (b) f.f. method.

ally they should be both rectangles. Figure 11 shows the 3-D shape reconstructed from the non-fixated pair corresponding to Fig. 10(b). Although both methods produce fairly accurate shapes, we can see that the v.f. method gives a more accurate result because of the better estimates of the focal lengths.

Figure 12 shows two images (680×512 pixels) of a store house taken by a third person without paying any attention to the fixation issue. The zooming condition was not recorded except that the focal length was not changed in the course of shooting. From the feature point matches marked in the images (we selected them manually), we computed the fundamental matrix by renormalization; real focal lengths were obtained without subsampling. We tested fixation and found that $h = 13.46$ (pixels) and $h' = 11.74$ (pixels), so we judged them to be fixated. This evidences the fact that humans are very likely to take fixated images unconsciously; particularly so, if we note that there is nothing to fixate in the centers of the two images.

The focal lengths are estimated to be $f = 973$ (pixels) and $f' = 778$ (pixels) by the v.f. method and $f = f' = 799$ (pixels) by the f.f. method. Figure 13 shows front and top views of the 3-D shape reconstructed by the two methods. Since the store house should have rectangular corners, we conclude that the f.f. method produces a better result, as expected, meaning that the focal lengths estimated by the f.f. method are closer to their true values.

9. Concluding Remarks

In this paper, we studied the problem of the occurrences of imaginary focal lengths and the degen-

eracy of fixated images that plagues 3-D reconstruction from two uncalibrated views. We first summarized the theoretical backgrounds. Then, we presented a practical remedy scheme of subsampling feature points and fixing the focal length. Our simulation experiments revealed a rather surprising fact that when the focal length is actually fixed, not using that knowledge yields better results for non-fixated images. We gave an explanation to this seeming paradox and derived a hybrid method switching the computation by judging fixation. Doing simulations and real image experiments, we have demonstrated the effectiveness of our method.

Appendix A: Decomposability Condition

Let $\{f, f'\}$ be the true focal lengths for the first and the second cameras, and let the corresponding points (x, y) and (x', y') be represented, instead of eqs. (2), by the following vectors.

$$\bar{\boldsymbol{x}} = \begin{pmatrix} x/f \\ y/f \\ 1 \end{pmatrix}, \qquad \bar{\boldsymbol{x}}' = \begin{pmatrix} x'/f' \\ y'/f' \\ 1 \end{pmatrix}. \tag{21}$$

The epipolar equation (1) now has the form

$$(\bar{\boldsymbol{x}}, \boldsymbol{E}\bar{\boldsymbol{x}}') = 0, \tag{22}$$

where $\boldsymbol{E}$ is a rank 2 matrix, called the *essential matrix* (Hartley and Zisserman, 2000; Kanatani, 1996). Suppose the second camera is translated by $\boldsymbol{t}$ and rotated by $\boldsymbol{R}$ (rotation matrix) relative to the first camera; we call $\{\boldsymbol{t}, \boldsymbol{R}\}$ the *motion parameters*. It can be easily shown (Hartley and Zisserman, 2000; Kanatani, 1996) that

$$\boldsymbol{E} \simeq \boldsymbol{t} \times \boldsymbol{R}, \tag{23}$$

where $\simeq$ means that both sides are equal up to a nonzero constant. 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 $\boldsymbol{A}$.

Huang and Faugeras (1989) proved that a matrix $\boldsymbol{E}$ has the form of eq. (23) if and only if one of its singular values is zero and the other two are equal (*decomposability condition*). Kanatani (1993) proved that this is equivalent to the following condition:

$$\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = \frac{1}{2}\|\boldsymbol{E}\|^4. \tag{24}$$

In fact, if $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$ are the singular values of $\boldsymbol{E}$, the constraint $\det \boldsymbol{E} = 0$ implies $\sigma_3 = 0$. It is easy to see that eq. (24) is equivalent to $(\sigma_1^2 - \sigma_2^2)^2 = 0$ (Kanatani, 1993).

Comparing eqs. (1) and (22), we can see that the essential matrix $\boldsymbol{E}$ is related to the fundamental matrix $\boldsymbol{F}$ in the form

$$\boldsymbol{E} \simeq \mathrm{diag}\Big(1, 1, \frac{f_0}{f}\Big)\boldsymbol{F}\mathrm{diag}\Big(1, 1, \frac{f_0}{f'}\Big), \tag{25}$$

where $\mathrm{diag}(\cdots)$ denotes the diagonal matrix that has $\cdots$ as its diagonal elements in that order. Thus, $\boldsymbol{E}$ can be regarded as a function of $\{f, f'\}$. Define the following function of $\{f, f'\}$:

$$K(f, f') = \|\boldsymbol{E}\boldsymbol{E}^\top\|^2 - \frac{1}{2}\|\boldsymbol{E}\|^4. \tag{26}$$

Substituting eq. (25), we can see that $K(f, f')$ is a polynomial of order 8 in $\{f, f'\}$ (order 4 in each).

Equation (24) implies that the focal lengths $\{f, f'\}$ are obtained by solving $K(f, f') = 0$. It appears that this equation alone is insufficient to solve for the two unknowns $\{f, f'\}$. It turns out, however, that the partial derivatives of $K(f, f')$ with respect to $\{f, f'\}$ are both 0 (Kanatani, 1996) (see Appendix B for the proof). Thus, $\{f, f'\}$ are the solution of

$$K = \frac{\partial K}{\partial f} = \frac{\partial K}{\partial f'} = 0, \tag{27}$$

which provides three constraints, but only the two of them are independent.

If we change the variables from $\{f, f'\}$ into $\{\xi, \eta\}$ defined by eqs. (4), the function $K(f, f')$ is rewritten in the form of eq. (5) after lengthy calculations. Thus, eq. (27) is equivalent to eq. (6).

Appendix B: Singularity of the Essential Matrix

Since the scale of the essential matrix $\boldsymbol{E}$ is indeterminate, we can assume that $\|\boldsymbol{E}\| = 1$ without loosing generality. This means that $\boldsymbol{E}$ is a point on an 8-dimensional unit sphere S^8 centered at the origin in the 9-dimensional parameter space. Equation (24) states that the true value $\boldsymbol{E}$ is on the intersection of the sphere S^8 with the manifold defined by $\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = 1/2$. Let $\boldsymbol{E} + \Delta\boldsymbol{E}$ be a neighboring point to $\boldsymbol{E}$ on

the manifold $\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = 1/2$. To a first approximation, we obtain

$$\begin{aligned}\frac{1}{2} &= \|(\boldsymbol{E}+\Delta\boldsymbol{E})(\boldsymbol{E}+\Delta\boldsymbol{E})^\top\|^2 = \|\boldsymbol{E}\boldsymbol{E}^\top\|^2 \\ &\quad +(\boldsymbol{E}\boldsymbol{E}^\top; \boldsymbol{E}\Delta\boldsymbol{E}^\top) + (\boldsymbol{E}\boldsymbol{E}^\top; \Delta\boldsymbol{E}\boldsymbol{E}^\top) \\ &= \frac{1}{2} + 2(\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{E}; \Delta\boldsymbol{E}),\end{aligned} \tag{28}$$

where we define the inner product of matrices $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$ by $(\boldsymbol{A}; \boldsymbol{B}) = \sum_{i,j=1}^3 A_{ij}B_{ij}$. Equation (28) implies that a first approximation

$$(\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{E}; \Delta\boldsymbol{E}) = 0, \tag{29}$$

which means that $\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{E}$ is the "surface normal" to the manifold $\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = 1/2$ at $\boldsymbol{E}$. Substituting eq. (23), we obtain

$$\begin{aligned}\boldsymbol{E}\boldsymbol{E}^\top\boldsymbol{E} &\simeq (\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R}\boldsymbol{R}^\top(\boldsymbol{t}\times\boldsymbol{I})^\top(\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R} \\ &= (\boldsymbol{t}\times\boldsymbol{I})(\boldsymbol{t}\times\boldsymbol{I})^\top(\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R} \\ &= (\|\boldsymbol{t}\|^2\boldsymbol{I} - \boldsymbol{t}\boldsymbol{t}^\top)(\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R} = \|\boldsymbol{t}\|^2\bar{\boldsymbol{E}} \\ &\quad +\boldsymbol{t}((\boldsymbol{t}\times\boldsymbol{t})^\top)\boldsymbol{R} = \|\boldsymbol{t}\|^2\boldsymbol{E},\end{aligned} \tag{30}$$

where we have used the identity $\boldsymbol{t}\times\boldsymbol{R} = (\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R}$ and noted that $\boldsymbol{t}\times\boldsymbol{I}$ is the skew-symmetric matrix consisting of the components of $\boldsymbol{t}$ (Kanatani, 1996).

Because $\boldsymbol{E}$ itself is the surface normal to the sphere S^8 at $\boldsymbol{E}$, eq. (30) implies that the sphere S^8 is *tangent* to the manifold $\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = 1/2$ at $\boldsymbol{E}$, sharing a common tangent space there. If $\boldsymbol{E}$ is constrained to be in a two-parameter subset of the sphere S^8 parameterized by $\{f,\ f'\}$, the tangency relation also holds for arbitrary perturbations of $\{f,\ f'\}$. Hence holds eq. (27).

Since $\|\boldsymbol{E}\boldsymbol{E}^\top\|^2 = 1/2$ and S^8 are 8-dimensional as manifolds, their intersection in the 9-dimensional space should be a 7-dimensional manifolds in general. Because of the above *non-transversality*, however, the intersection degenerates into a 6-dimensional manifold. The solution $\boldsymbol{E}$ is on the intersection of this 6-dimensional manifold with the 7-dimensional manifold defined by $\det\boldsymbol{E} = 0$. It follows that the intersection is a 5-dimensional manifold. Thus, the essential matrix has five degrees of freedom corresponding to the camera rotation and the normalized translation.

Appendix C: Analytical Expression for the Focal Lengths

From eqs. (23) and (25), the fundamental matrix $\boldsymbol{F}$ is related to the motion parameters $\{\boldsymbol{t},\ \boldsymbol{R}\}$ by

$$\boldsymbol{F} \simeq \mathrm{diag}\Big(1, 1, \frac{f}{f_0}\Big)(\boldsymbol{t}\times\boldsymbol{R})\mathrm{diag}\Big(1, 1, \frac{f'}{f_0}\Big), \tag{31}$$

from which we see that the eigenvector $\boldsymbol{e}$ of $\boldsymbol{F}^\top$ and the eigenvector $\boldsymbol{e}'$ of $\boldsymbol{F}$ are related to the motion parameters$\{\boldsymbol{t},\ \boldsymbol{R}\}$ in the from

$$\boldsymbol{e} \simeq \mathrm{diag}\Big(1, 1, \frac{f_0}{f}\Big)\boldsymbol{t}, \qquad \boldsymbol{e}' \simeq \mathrm{diag}\Big(1, 1, \frac{f_0}{f'}\Big)\boldsymbol{R}^\top\boldsymbol{t}. \tag{32}$$

Hence, we obtain

$$\boldsymbol{t} \simeq \mathrm{diag}\Big(1, 1, \frac{f}{f_0}\Big)\boldsymbol{e} \simeq \boldsymbol{R}\mathrm{diag}\Big(1, 1, \frac{f'}{f_0}\Big)\boldsymbol{e}'. \tag{33}$$

Substituting this into eq. (31), we obtain

$$\begin{aligned}\boldsymbol{F} &\simeq \boldsymbol{e}\times\mathrm{diag}\Big(1, 1, \frac{f_0}{f}\Big)\boldsymbol{R}\mathrm{diag}\Big(1, 1, \frac{f'}{f_0}\Big) \\ &\simeq \mathrm{diag}\Big(1, 1, \frac{f}{f_0}\Big)\boldsymbol{R}\mathrm{diag}\Big(1, 1, \frac{f_0}{f'}\Big)\times\boldsymbol{e}'.\end{aligned} \tag{34}$$

Here, we have used the identity $(\boldsymbol{T}\boldsymbol{u})\times\boldsymbol{I} = (\boldsymbol{T}^{-1})^\top(\boldsymbol{u}\times\boldsymbol{I})\boldsymbol{T}^{-1}$ for an arbitrary vector $\boldsymbol{u}$ and an arbitrary nonsingular matrix $\boldsymbol{T}$. We define $\boldsymbol{T}\times\boldsymbol{u}$ to be $\boldsymbol{T}(\boldsymbol{u}\times\boldsymbol{I})^\top$.

From eq. (34), we have

$$\begin{aligned}\boldsymbol{F}\mathrm{diag}\Big(1, 1, \frac{f_0}{f'}\Big) &\simeq \boldsymbol{e}\times\mathrm{diag}\Big(1, 1, \frac{f_0}{f}\Big)\boldsymbol{R}, \\ \mathrm{diag}\Big(1, 1, \frac{f_0}{f}\Big)\boldsymbol{F} &\simeq \boldsymbol{R}\mathrm{diag}\Big(1, 1, \frac{f_0}{f'}\Big)\times\boldsymbol{e}'.\end{aligned} \tag{35}$$

Eliminating $\boldsymbol{R}$ by using the orthogonality relation $\boldsymbol{R}^\top\boldsymbol{R} = \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{I}$, we obtain the following *Kruppa equations* (Bougnoux, 1998; Hartley and Zisserman, 2000):

$$\boldsymbol{F}\mathrm{diag}\Big(1, 1, \frac{f_0^2}{f'^2}\Big)\boldsymbol{F}^\top \simeq \boldsymbol{e}\times\mathrm{diag}\Big(1, 1, \frac{f_0^2}{f^2}\Big)\times\boldsymbol{e},$$

$$\boldsymbol{F}^\top \mathrm{diag}\Big(1, 1, \frac{f_0^2}{f^2}\Big)\boldsymbol{F} \simeq \boldsymbol{e}' \times \mathrm{diag}\Big(1, 1, \frac{f_0^2}{f'^2}\Big) \times \boldsymbol{e}'. \tag{36}$$

In terms of $\{\xi, \eta\}$ defined by eqs. (4), these equations are rewritten as

$$\begin{aligned} \boldsymbol{F}(\boldsymbol{I} + \eta \boldsymbol{k}\boldsymbol{k}^\top)\boldsymbol{F}^\top &\simeq \boldsymbol{e} \times (\boldsymbol{I} + \xi \boldsymbol{k}\boldsymbol{k}^\top) \times \boldsymbol{e}, \\ \boldsymbol{F}^\top(\boldsymbol{I} + \xi \boldsymbol{k}\boldsymbol{k}^\top)\boldsymbol{F} &\simeq \boldsymbol{e}' \times (\boldsymbol{I} + \eta \boldsymbol{k}\boldsymbol{k}^\top) \times \boldsymbol{e}', \end{aligned} \tag{37}$$

which reduce to

$$\boldsymbol{F}\boldsymbol{F}^\top + \eta(\boldsymbol{F}\boldsymbol{k})(\boldsymbol{F}\boldsymbol{k})^\top \simeq \boldsymbol{P}_{\mathbf{e}} + \xi(\boldsymbol{e} \times \boldsymbol{k})(\boldsymbol{e} \times \boldsymbol{k})^\top, \tag{38}$$

$$\begin{aligned} \boldsymbol{F}^\top \boldsymbol{F} + \xi(\boldsymbol{F}^\top \boldsymbol{k})(\boldsymbol{F}^\top \boldsymbol{k})^\top &\simeq \boldsymbol{P}_{\mathbf{e}'} \\ + \eta(\boldsymbol{e}' \times \boldsymbol{k})(\boldsymbol{e}' \times \boldsymbol{k})^\top, & \end{aligned} \tag{39}$$

where we define the projection matrices

$$\boldsymbol{P}_{\mathbf{e}} = \boldsymbol{I} - \boldsymbol{e}\boldsymbol{e}^\top, \qquad \boldsymbol{P}_{\mathbf{e}'} = \boldsymbol{I} - \boldsymbol{e}'\boldsymbol{e}'^\top, \tag{40}$$

onto the planes perpendicular to $\boldsymbol{e}$ and $\boldsymbol{e}'$, respectively. Multiplying $\boldsymbol{k}$ from the right on both sides of eqs. (38) and (39), we obtain

$$\begin{aligned} \boldsymbol{F}\boldsymbol{F}^\top \boldsymbol{k} + \eta(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})\boldsymbol{F}\boldsymbol{k} &= c\boldsymbol{P}_{\mathbf{e}}\boldsymbol{k}, \\ \boldsymbol{F}^\top \boldsymbol{F}\boldsymbol{k} + \xi(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})\boldsymbol{F}^\top \boldsymbol{k} &= c'\boldsymbol{P}_{\mathbf{e}'}\boldsymbol{k}, \end{aligned} \tag{41}$$

where c and c' are unknown constants.

Computing the inner product of $\boldsymbol{k}$ and both sides of the second of eqs. (41) and the inner product of $\boldsymbol{F}^\top \boldsymbol{k}$ and both sides of the second of eqs. (41), we obtain

$$\|\boldsymbol{F}\boldsymbol{k}\|^2 + (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2 \xi = c'\|\boldsymbol{e}' \times \boldsymbol{k}\|^2, \tag{42}$$

$$\begin{aligned} &(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{F}^\top \boldsymbol{F}\boldsymbol{k}) + (\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})\|\boldsymbol{F}^\top \boldsymbol{k}\|^2 \xi \\ &\quad = c'(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k}), \end{aligned} \tag{43}$$

which can be solved for $\{\xi, c'\}$, resulting in the first of eqs. (7).

Computing the inner product of $\boldsymbol{k}$ and both sides of the first of eqs. (41) and the inner product of $\boldsymbol{F}^\top \boldsymbol{k}$ and both sides of the first of eqs. (41), we obtain

$$\|\boldsymbol{F}^\top \boldsymbol{k}\|^2 + \eta(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})^2 = c\|\boldsymbol{e} \times \boldsymbol{k}\|^2, \tag{44}$$

$$(\boldsymbol{k}, \boldsymbol{F}^\top \boldsymbol{F}\boldsymbol{F}^\top \boldsymbol{k}) + \eta(\boldsymbol{k}, \boldsymbol{F}\boldsymbol{k})\|\boldsymbol{F}\boldsymbol{k}\|^2 = c(\boldsymbol{k}, \boldsymbol{F}^\top \boldsymbol{k}), \tag{45}$$

which can be solved for $\{\eta, c\}$, resulting in the second of eqs. (7).

Our derivation is essentially the same as the formula for f given by Bougnoux (1998), but the result has a slightly different appearance.

Appendix D: Analytical Fixed Focal Length

In the fixed focal length case, the solution ξ is given by a common root of $K(\xi) = 0$ and $K'(\xi) = 0$. The solution ξ should also satisfy $\xi K'(\xi) = 0$. If $a_1 \neq 0$, these three equations are written as

$$\begin{pmatrix} a_1 & a_2 & a_3\xi^2 + a_4\xi + a_5 \\ & 4a_1 & 3a_2\xi^2 + 2a_3\xi + a_4 \\ 4a_1 & 3a_2 & 2a_3\xi^2 + a_4\xi \end{pmatrix} \begin{pmatrix} \xi^4 \\ \xi^3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{46}$$

Since a nontrivial solution should exist, its determinant should vanish:

$$\begin{aligned} &\begin{vmatrix} a_1 & a_2 & a_3\xi^2 + a_4\xi + a_5 \\ & 4a_1 & 3a_2\xi^2 + 2a_3\xi + a_4 \\ 4a_1 & 3a_2 & 2a_3\xi^2 + a_4\xi \end{vmatrix} \\ &\quad = a_1(3a_2^2 - 8a_1a_3)\xi^2 + 2a_1(a_2a_3 - 6a_1a_4)\xi \\ &\qquad + a_1(a_2a_4 - 16a_1a_5) = 0. \end{aligned} \tag{47}$$

From this, we obtain eq. (12). If $a_1 = 0$ and $a_2 \neq 0$, the three equations $K(\xi) = 0$, $K'(\xi) = 0$, and $\xi K'(\xi) = 0$ are rewritten as

$$\begin{pmatrix} a_2 & a_3 & a_4\xi + a_5 \\ & 3a_2 & 2a_3\xi + a_4 \\ 3a_2 & 2a_3 & a_4\xi \end{pmatrix} \begin{pmatrix} \xi^3 \\ \xi^2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{48}$$

Since the determinant should vanish, we have

$$\begin{aligned} &\begin{vmatrix} a_2 & a_3 & a_4\xi + a_5 \\ & 3a_2 & 2a_3\xi + a_4 \\ 3a_2 & 2a_3 & a_4\xi \end{vmatrix} \\ &\quad = 2a_2(a_3^2 - 3a_2a_4)\xi \\ &\qquad + a_2(a_3a_4 - 9a_2a_5) = 0, \end{aligned} \tag{49}$$

and obtain eq. (13). If $a_1 = a_2 = 0$ and $a_3 \neq 0$, we simply solve the linear equation $K'(\xi) = 0$, so we obtain eq. (15). If $a_1 = a_2 = a_3 = 0$, the solution is indeterminate.

Acknowledgments

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. 17500112).

Notes

1. We used the value $f_0 = 600$ in our experiment, but no practical difference should result by letting $f_0 = 1$.
2. Even in the presence of noise, the fundamental matrix $\boldsymbol{F}$ is computed to be $\det \boldsymbol{F} = 0$, so $\boldsymbol{F}^\top$ and $\boldsymbol{F}$ both have eigenvalue 0.
3. We have confirmed by experiments that the dependence on d is the same for homogeneous noise, too, but the same phenomena emerge for a smaller noise level σ for inhomogeneous noise.
4. The C++ source code is available at: http://www.img.tutkie.tut.ac.jp

References

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

Brooks, M. J., de Agapito, L., Huynh, D. Q., and Baumela, L. 1998. Towards robust metric reconstruction via a dynamic uncalibrated stereo head. *Image Vision Comput.*/ 16(14):989–1002.

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

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

Hartley, R., and Silpa-Anan, C. 2002. Reconstruction from two views using approximate calibration. In *Proc. 5th Asian Conf. Comput. Vision*, Melbourne, Australia, vol. 1, pp. 338–343.

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

Heyden, A. and Åström, K. 1997. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point In *Proc. IEEE Conf. Comput. Vision Pattern Recog.* Puerto Rico, pp. 438–443.

Huang, T. S. and Faugeras, O. D. 1989. Some properties of the E matrix in two-view motion estimation. *IEEE Trans. Patt. Anal. Mach. Intell.* 11(12):1310–1312.

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

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

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

Kanatani, K. 2000. Optimal fundamental matrix computation: Algorithm and reliability analysis. In *Proc. 6th Symp. Sensing via Image Inf.*/ Yokohama, Japan, pp. 291–298.

Kanatani, K. 2004. Uncertainty modeling and model selection for geometric inference. *IEEE Trans. Patt. Anal. Mach. Intell.*/ 26(10):1307–1319.

Kanatani K. and Matsunaga, C. 2000. Closed-form expression for focal lengths from the fundamental matrix. In *Proc. 4th Asian Conf. Comput. Vision*, Taipei, Taiwan, vol. 1, pp. 128–133.

Kanatani, K. and Ohta, N. 2003. Comparing optimal three-dimensional reconstruction for finite motion and optical flow. *J. Electronic Imaging*, 12(3):478–488.

Kanazawa, Y. and Kanatani, K. 2004. Robust image matching under a large disparity. In *Proc. 6th Asian Conf. Computer Vision*, Jeju, Korea, vol. 2, pp. 1128–1133.

Leedan, Y. and Meer, P. 2000. Heteroscedastic regression in computer vision: Problems with bilinear constraint. *Int. J. Comput. Vision.*/, 37(2):127–150.

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

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

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

Pollefeys, M., Koch, R., and Van Gool, L. 1999. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Int. J. Comput. Vision*, 32(1):7–26.

Sturm, P. 2001. On focal length calibration from two views. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*/ Kauai, HI, USA, vol. 2, pp. 145–150.

Ueshiba, T. and Tomita, F. 2003. Self-calibration from two perspective views under various conditions: Closed-form solutions and degenerate configurations. In *Proc. Australia-Japan Advanced Workshop on Computer Vision*, Adelaide, Australia, pp. 118–125.

Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.-T. 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell.* 78:87–119.