IPSJ Transactions on Computer Vision and Applications Vol.6 120–131 (Nov. 2014)

[DOI: 10.2197/ipsjtcva.6.120]

Research Paper

Decomposing Three Fundamental Matrices for Initializing 3-D Reconstruction from Three Views

Yasushi Kanazawa[1,a)] Yasuyuki Sugaya[1,b)] Kenichi Kanatani[2,c)]

Received: December 16, 2013, Accepted: August 9, 2014, Released: November 11, 2014

Abstract: This paper focuses on initializing 3-D reconstruction from scratch without any prior scene information. Traditionally, this has been done from two-view matching, which is prone to the degeneracy called "imaginary focal lengths." We overcome this difficulty by using three images, but we do not require three-view matching; all we need is three fundamental matrices separately computed from pair-wise image matching. We exploit the redundancy of the three fundamental matrices to optimize the camera parameters and the 3-D structure. The main theme of this paper is to give an analytical procedure for computing the positions and orientations of the three cameras and their internal parameters from three fundamental matrices. The emphasis is on resolving the ambiguity of the solution resulting from the sign indeterminacy of the fundamental matrices. We do numerical simulation to show that imaginary focal lengths are less likely for our three view methods, resulting in higher accuracy than the conventional two-view method. We also test the degeneracy tolerance capability of our method by using endoscopic intestine tract images, for which the camera configuration is almost always nearly degenerate. We demonstrate that our method allows us to obtain more detailed intestine structures than two-view reconstruction and observe how our three-view reconstruction is refined by bundle adjustment. Our method is expected to broaden medical applications of endoscopic images.

Keywords: three-view 3-D reconstruction, fundamental matrix, focal length computation, motion parameter computation, endoscopic image analysis

1. Introduction

Today, 3-D reconstruction from images is a common technique of computer vision thanks to various reconstruction tools available on the Web. The basic principle is what is known as *bundle adjustment*, computing from point correspondences over multiple images all 3-D point positions and all camera parameters by searching the high-dimensional parameter space. The search is done so as to minimize the discrepancy, or the *reprojection error*, between the observed images and the projections of the 3-D points predicted using the estimated camera parameters. The best known bundle adjustment software is *SBA* of Lourakis and Argyros [15]. Snavely et al. [17], [18] combined it with feature point detection and matching as a package called *bundler*. Bundle adjustment is an iterative process, requiring an initial solution, which is usually computed by choosing from among the input image pairs of well matched views. This is because the 3-D shape and the camera parameters are easily computed from two views, and various practically high-accuracy techniques have been presented [12].

However, it is well known that two-view reconstruction fails if the two cameras are in a "fixating" configuration [3], [8], i.e., their optical axes intersect in the scene (**Fig. 1** (a)). This includes the case of parallell optical axes (Fig. 1 (b)), which are regarded as intersecting at infinity. This configuration is very natural when one takes images of the same object from different positions. In fact, almost all casual tourist images are taken in this way. Another problem is that irrespective of the camera configuration, the information obtained from two views is minimal, resulting in the same number of equations as the number of unknowns. This may be an advantage in that the solution can be obtained analytically, but often the solution that satisfies all equations does not exist for noisy data. Typically, the square of some expressions containing the focal lengths become negative; this problem is known as the "imaginary focal length degeneracy."

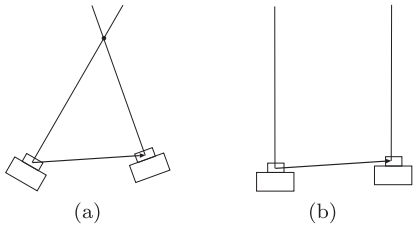

Fig. 1 Fixating camera configuration. (a) The optical axes intersect. (b) The optical axes are parallel.

The purpose of this paper is not so much to achieve yet higher reconstruction accuracy. Rather, we focus on preventing degeneracy. Namely, we want to initialize 3-D reconstruction stably from scratch, i.e., without requiring any prior information about the scene structure or the camera positions. There have already been some such attempts. Observing that fixating configurations occur when the principal point of one image matches to that of

[1] Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441–8580, Japan
[2] Professor Emeritus, Okayama University, Okayama 700–8535, Japan
[a)] kanazawa@cs.tut.ac.jp
[b)] sugaya@iim.cs.tut.ac.jp
[c)] kanatani2013@yahoo.co.jp

the other image, Hartley and Silpa-Anan [4] used the regularization approach to minimally move the assumed principal points so that the imaginary focal lengths do not arise, but the solution depends on the regularization parameter. Kanatani et al. [9] proposed random resampling of matching points to avoid imaginary focal lengths, but a sufficient number of correspondences are necessary. Goldberger [2] adopted the projective reconstruction framework, computing the camera matrices up to projectivity from fundamental matrices and epipoles computed from image pairs. For Euclidean reconstruction, however, more information is required [16]. In this paper, we impose a strict constraint on the cameras so that the Euclidean structure results from minimum information, yet extra degrees of freedom remain to be adjusted to suppress imaginary focal lengths. This is made possible by using three images, but we do not require three-view matching; all we need is three fundamental matrices separately computed from pair-wise image matching.

In this paper, we provide an analytical procedure for computing the positions and orientations of the three cameras and their internal parameters from three fundamental matrices. This was already given in our earlier version [14], but here a special emphasis is placed on resolving the ambiguity of the solution resulting from the sign indeterminacy of the fundamental matrices. We also discuss an optimal process for computing the 3-D position of corresponding image points. If correspondence is obtained between two images, this is done by the optimal triangulation scheme of Kanatani et al. [13]. We extend this to the case where correspondence is given over three images.

In Section 2, we state the objective of this paper. In Section 3, we present our strategy for computing focal lengths from three fundamental matrices, and in Sections 4 and 5, we describe our iterative scheme for simultaneously computing the camera translations and rotations. Section 6 deals with the issue of sign ambiguity of the solution. We explain the origin of the ambiguity and introduce our rule for resolving it. Section 7 summarizes our parameter computation algorithm. In Section 8 and Section 9, we show how observed image positions are optimally corrected for computing their 3-D positions for those points that are matched over the three views; for points matched between two images, this can be done using the standard procedure of two-view reconstruction [5], [12]. In Section 10, we do numerical simulation and observe that imaginary focal lengths are less likely to occur, resulting in higher accuracy than two-view reconstruction. In Section 11, we apply our method to endoscopic intestine tract images. This provides a good testbed for the degeneracy tolerance capability of our method, because the camera configuration is very pathological: the camera moves almost in one direction in intestine tracts and hence always in a near fixation configuration, which is very likely to cause imaginary focal lengths. We observe that our method produces a more detailed structure in a wider range than pairwise two-view reconstructions combined. We also observe how our three-view reconstruction is refined by bundle adjustment. In Section 12, we conclude.

2. The Task

For two-view reconstruction, the cameras must satisfy the following conditions [5], [12]:

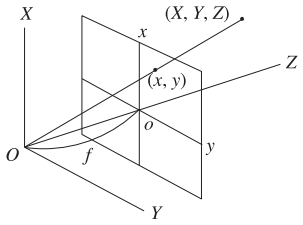

Fig. 2 The XYZ camera coordinate system and the xy image coordinate system.

(1) The principal point is known.
(2) The aspect ratio is 1.
(3) No image skew exists.

This constraint stems from the fact that the available information from two views is limited. We could relax this for three views [2], [4], [16], but since our intention is to exploit the redundancy of three-view information to do optimization, we adopt the same constraint. This is no big restriction in practice, because today's cameras mostly satisfy the requirements or can easily be so calibrated beforehand.

We define a camera-based XYZ coordinate system with the origin at the lens center and the Z-axis along its optical axis. On the image plane, we define an xy image coordinate system such that the origin o is at the principle point (at the frame center by default) with the x-axis upward and the y-axis rightward (**Fig. 2**). This setting is for the x- and y-axes to constitute, together with the optical axis, a right-handed system, which is necessary for 3-D rotation computation (for this purpose, we could instead take the x-axis rightward and the y-axis downward).

We capture three images of the same scene by three cameras (or equivalently by moving one camera). We call these images the 0th, 1st, and 2nd views, and the corresponding cameras the 0th, 1st, and 2nd cameras, respectively. Suppose a point (x, y) in the 0th view corresponds to (x', y') in the 1st view. We write the epipolar equation [5] between them in the form

$$(\boldsymbol{x}, \boldsymbol{F}_{01}\boldsymbol{x}') = 0, \qquad \boldsymbol{x} = \begin{pmatrix} x/f_0 \\ y/f_0 \\ 1 \end{pmatrix}, \qquad \boldsymbol{x}' = \begin{pmatrix} x'/f_0 \\ y'/f_0 \\ 1 \end{pmatrix}, \tag{1}$$

where $\boldsymbol{F}_{01}$ is the fundamental matrix between the 0th and 1st views. We write $(\boldsymbol{a}, \boldsymbol{b})$ for the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The scaling constant f_0 is for stabilizing numerical computation; we take it to be an approximate focal length of the cameras and call it the *default focal length* (we set it to 600 pixels in our experiment). The fundamental matrices $\boldsymbol{F}_{02}$ between the 0th and 2nd views and $\boldsymbol{F}_{12}$ between the 1st and 2nd views are similarly defined. Fundamental matrices are uniquely computed from eight or more point correspondence pairs (theoretically seven points are sufficient, but the solution may not be unique). In our experiment, we use the EFNS (Extended Fundamental Numerical Scheme) of Kanatani and Sugaya [10], which can compute an exact reprojection error minimization solution.

We regard the XYZ coordinate system of the 0th camera as the world coordinate system. Let $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ be the lens centers of the 1st and the 2nd cameras, respectively, and $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ their rotations relative to the 0th camera (**Fig. 3**). Let f, f', and f'' be the focal lengths of the 0th, 1st, and 2nd cameras, respectively. The

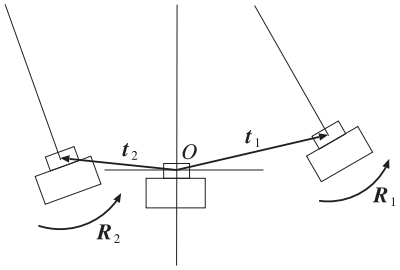

Fig. 3 The translations $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ and the rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ of the 1st and 2nd cameras relative to the 0th camera.

fundamental matrices $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$ ideally (i.e., if they are exact) satisfy the following identities [5], [7]:

$$\begin{aligned}
\boldsymbol{F}_{01} &\simeq \mathrm{diag}(1,1,\frac{f}{f_0})\Big(\boldsymbol{t}_1 \times \boldsymbol{R}_1\Big)\mathrm{diag}(1,1,\frac{f'}{f_0}),\\
\boldsymbol{F}_{02} &\simeq \mathrm{diag}(1,1,\frac{f}{f_0})\Big(\boldsymbol{t}_2 \times \boldsymbol{R}_2\Big)\mathrm{diag}(1,1,\frac{f''}{f_0}),\\
\boldsymbol{F}_{12} &\simeq \mathrm{diag}(1,1,\frac{f'}{f_0})\Big((\boldsymbol{R}_1^\top(\boldsymbol{t}_2-\boldsymbol{t}_1)) \times (\boldsymbol{R}_1^\top\boldsymbol{R}_2)\Big)\mathrm{diag}(1,1,\frac{f''}{f_0}).
\end{aligned} \tag{2}$$

Here, the symbol $\simeq$ denotes equality up to a nonzero constant and $\mathrm{diag}(a,b,c)$ denotes the diagonal matrix with a, b, and c as the diagonal elements in that order. For a vector $\boldsymbol{v}$ and a matrix $\boldsymbol{A}$, we define $\boldsymbol{v} \times \boldsymbol{A}$ to be the matrix whose columns are the vector products of $\boldsymbol{v}$ and the corresponding columns of $\boldsymbol{A}$. The task of this paper is to compute f, f', f'', $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, $\boldsymbol{R}_1$, and $\boldsymbol{R}_2$ from given fundamental matrices $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$, considering the fact that the computed $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$ may not be exact. Since the fundamental matrices are determined only up to scale, we assume that $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$ are all normalized to unit Frobenius norm, as is customary for most available fundamental matrix computation tools.

3. Focal Length Computation

Instead of computing f, f', and f'', we compute the following x, y, and z:

$$x \equiv \left(\frac{f_0}{f}\right)^2 - 1, \qquad y \equiv \left(\frac{f_0}{f'}\right)^2 - 1, \qquad z \equiv \left(\frac{f_0}{f''}\right)^2 - 1. \tag{3}$$

This definition is in this section only (these x and y are different from those in Eq. (1)) . It is known [9] that x and y ideally minimize, in the neighborhood of the solution, the quadratic polynomial in x and y

$$\begin{aligned}
K_{01}(x,y) &= (\boldsymbol{k}, \boldsymbol{F}_{01}\boldsymbol{k})^4x^2y^2 + 2(\boldsymbol{k}, \boldsymbol{F}_{01}\boldsymbol{k})^2\|\boldsymbol{F}_{01}^\top\boldsymbol{k}\|^2x^2y\\
&+2(\boldsymbol{k}, \boldsymbol{F}_{01}\boldsymbol{k})^2\|\boldsymbol{F}_{01}\boldsymbol{k}\|^2xy^2 + \|\boldsymbol{F}_{01}^\top\boldsymbol{k}\|^4x^2 + \|\boldsymbol{F}_{01}\boldsymbol{k}\|^4y^2\\
&+4(\boldsymbol{k}, \boldsymbol{F}_{01}\boldsymbol{k})(\boldsymbol{k}, \boldsymbol{F}_{01}\boldsymbol{F}_{01}^\top\boldsymbol{F}_{01}\boldsymbol{k})xy + 2\|\boldsymbol{F}_{01}\boldsymbol{F}_{01}^\top\boldsymbol{k}\|^2x\\
&+2\|\boldsymbol{F}_{01}^\top\boldsymbol{F}_{01}\boldsymbol{k}\|^2y + \|\boldsymbol{F}_{01}\boldsymbol{F}_{01}^\top\|^2 - \frac{1}{2}\Big((\boldsymbol{k}, \boldsymbol{F}_{01}\boldsymbol{k})^2xy\\
&+\|\boldsymbol{F}_{01}^\top\boldsymbol{k}\|^2x + \|\boldsymbol{F}_{01}\boldsymbol{k}\|^2y + \|\boldsymbol{F}_{01}\|^2\Big)^2,
\end{aligned} \tag{4}$$

where $\boldsymbol{k} \equiv (0,0,1)^\top$, and that the minimum is 0 (**Fig. 4** (a)). If quadric polynomials $K_{02}(x,z)$ and $K_{12}(y,z)$ are similarly defined, x and z minimize $K_{02}(x,z)$, and y and z minimize $K_{12}(y,z)$; their minimums are 0. Hence, we can determine x and y from

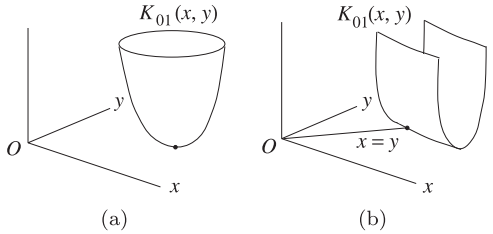

Fig. 4 (a) The values of x and y minimizes $K_{01}(x,y)$, and the minimum is 0. (b) For a fixating camera configuration, the minimum of $K_{0,1}(x,y)$ is not uniquely determined. If $x = y$, the solution is at the intersection of the minimum contour of $K_{0,1}(x,y)$ with the line $x = y$.

$K_{01}(x,y)$, y and z from $K_{12}(y,z)$, and x and z from $K_{02}(x,z)$. Moreover, the solution is analytically computed by the *Bougnoux formula* [5], [9]. In the presence of noise, however, the analytically obtained solutions are in general inconsistent to each other. This is because three fundamental matrices are not mutually independent but should satisfy some compatibility constraint [5], which does not hold for noisy data (the constraint given in Ref. [5] is a weak form; their fundamental matrices are formulated at the projective framework. Here, we are further assuming the conditions (1), (2), and (3) listed in the beginning of Section 2). Theoretically, we could first modify $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$ to make them compatible, but that would involve mathematical difficulties. Here, we adopt a much easier strategy of finding the three focal lengths by a simple averaging scheme: we compute the solution x, y, and z that minimize

$$F(x,y,z) = K_{01}(x,y) + K_{02}(x,z) + K_{12}(y,z). \tag{5}$$

Note that although this $F(x,y,z)$ is identically 0 for ideal data, its minimization for noisy data does not necessarily compute a theoretically optimal solution. Theoretically, the three terms should be given appropriate weights that reflect the scale and the uncertainty of each term. We omit such weights, partly because their theoretical evaluation is difficult and partly because our aim is to "initialize" 3-D reconstruction to be refined later by other means such as bundle adjustment, depending on the application. Here, we simply regard the three terms as having approximately the same scales and equal degrees of uncertainty.

In our experiment, we minimized Eq. (5) by using Newton iterations starting from $x = y = z = 0$, which is equivalent to $f = f' = f'' = f_0$. Then, f, f', and f'' are given from Eq. (3) in the form

$$f = \frac{f_0}{\sqrt{1+x}}, \qquad f' = \frac{f_0}{\sqrt{1+y}}, \qquad f'' = \frac{f_0}{\sqrt{1+z}}. \tag{6}$$

Note that if any of x, y, and z are equal to or less than -1, the computation fails. This is the so called "imaginary focal length problem," which frequently occurs in two-view reconstruction. One of the causes of this phenomenon is that the analytical solution relies on the fact that the solution not only minimizes $K_{01}(x,y)$, $K_{02}(x,z)$, and $K_{12}(x,z)$ but also their minimums are exactly 0, which does not hold for real data. Here, we are not assuming that their minimums are 0, so we expect that the imaginary focal length problem will be alleviated, if not completely avoided.

If the three focal lengths f, f', and f'' are known to be the same, e.g., when the same camera is used to take the three images, and if we want to use this information, two approaches are

conceivable. One is to minimize the function

$$F(x) = K_{01}(x, x) + K_{02}(x, x) + K_{12}(x, x), \tag{7}$$

for a single variable x. Another approach is to correct the computed x, y, and z to a single value $\bar{x}$ in such a way that the increase in the value of Eq. (5) is minimum. If Eq. (5) is approximated by a quadratic function in the neighborhood of its minimum, we obtain

$$\bar{x} = \Big((F_{xx} + F_{xy} + F_{xz})x + (F_{yx} + F_{yy} + F_{yz})y + (F_{zx} + F_{zy} + F_{zz})z\Big)\Big/\Big(F_{xx} + F_{yy} + F_{zz} + 2(F_{xy} + F_{yz} + F_{zx}),\Big) \tag{8}$$

where F_{xx}, F_{xy}, etc. are the second derivatives of Eq. (5) at its minimum. The role of the equal focal length constraint in two-view reconstruction was studied experimentally in Ref. [9]. We also repeated similar experiments, but the results are inconclusive. We found no definitive evidence that the equal focal length constraint improves the accuracy: better results were observed in some but not all cases. The reason that treating the focal length as varying, even though an identical camera is used, sometimes produces better 3-D shapes may be that some data inconsistency in the presence of noise is absorbed by multiple degrees of freedom of the focal lengths. However, this is not always the case. So, we leave this issue to future research.

It is known [9] that if two cameras, say the 0th and the 1st, are in a fixating configuration, the minimum of $K_{01}(x, y)$ in Eq. (4) degenerates to a curve in the xy plane so it does not have a unique minimum. If we assume that $f = f'$, the solution is uniquely determined as the intersection of that curve with the line $x = y$ (Fig. 4 (b)). However, if the two cameras are in an "isosceles" configuration (fixating with equal distance; **Fig. 5** (a)), the minimum curve of $K_{01}(x, y)$ is "tangent" to the line $x = y$ and hence no clear intersection is defined (Fig. 5 (b)). The same holds for the other pairs of cameras. However, our three-view formulation can uniquely determine the solution even when fixating camera configurations are included, unless the three cameras are in a simultaneous fixating configuration, in which case the Hessian of $F(x, y, z)$ in Eq. (5) becomes singular at the minimum, making numerical minimization unstable.

According to our experience, we observed that in situations where noise is very large and the camera configuration is nearly degenerate, as in the case of endoscopic image analysis to be shown later, the occurrence of imaginary focal length is less likely when the equal focal length constraint is imposed.

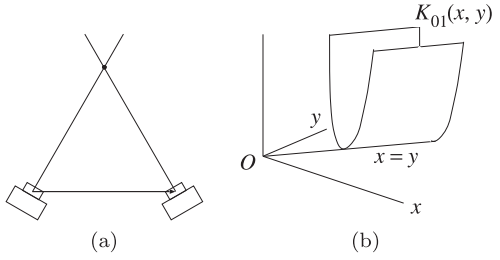

Fig. 5 (a) Isosceles camera configuran. (b) The minimum contour of $K_{01}(x, y)$ is tangent to the line $x = y$.

As mentioned earlier, the aim of our study is to obtain initial 3-D reconstruction to be refined later. For this purpose, it seem that we may simply use the nominal focal length provided the camera manufacturer or the value obtained by prior camera calibration. Today, however, more and more applications require 3-D reconstruction from images of unknown origin, such as on the Web. Even if the camera information is provided, the zoom setting at the time of shooting is often not recorded. Hence, estimating the focal lengths from images alone is an important issue today. The endoscopic image analysis to be shown later is one such example.

4. Translation Computation

The relative camera translation can be computed from the fundamental matrix between two views [12]. Hence, the three fundamental matrices $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$ can determine the translations between all the camera pairs. However, their signs and scales are indeterminate. Although we cannot fix the absolute scale as long as images are used, we can fix their relative scales from the "triangle condition," requiring that the three translations form a closed triangle. However, as we show shortly, the triangle condition involves camera rotations, so, unlike two-view reconstruction, translations cannot be determined separately. Here, we introduce a procedure for computing the translations and rotations at the same time.

Using the computed focal lengths f, f', and f'', we define the *essential matrices* $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$ by

$$\begin{aligned}
\boldsymbol{E}_{01} &\equiv \mathrm{diag}(1, 1, \frac{f_0}{f})\boldsymbol{F}_{01}\mathrm{diag}(1, 1, \frac{f_0}{f'}),\\
\boldsymbol{E}_{02} &\equiv \mathrm{diag}(1, 1, \frac{f_0}{f})\boldsymbol{F}_{02}\mathrm{diag}(1, 1, \frac{f_0}{f''}),\\
\boldsymbol{E}_{12} &\equiv \mathrm{diag}(1, 1, \frac{f_0}{f'})\boldsymbol{F}_{12}\mathrm{diag}(1, 1, \frac{f_0}{f''}),
\end{aligned} \tag{9}$$

From Eq. (2), they ideally satisfy

$$\boldsymbol{E}_{01} \simeq \boldsymbol{t}_1 \times \boldsymbol{R}_1, \quad \boldsymbol{E}_{02} \simeq \boldsymbol{t}_2 \times \boldsymbol{R}_2, \quad \boldsymbol{E}_{12} \simeq \boldsymbol{t}_{12} \times \boldsymbol{R}_1^\top \boldsymbol{R}_2, \tag{10}$$

where $\boldsymbol{t}_{12}$ is the translation of the 2nd camera relative to the 1st camera with respect to the 1st camera coordinate system. Since the 1st camera coordinate system is rotated relatieve to the 0th camera coordinate system by $\boldsymbol{R}_1$, the translation of the 2nd camera relative to the 1st camera with respect to the world coordinate system is $\boldsymbol{R}_1\boldsymbol{t}_{12}$, which should equal $\boldsymbol{t}_2 - \boldsymbol{t}_1$. Hence, the triangle condition is written as follows (**Fig. 6**):

$$\boldsymbol{t}_{12} = \boldsymbol{R}_1^\top(\boldsymbol{t}_2 - \boldsymbol{t}_1). \tag{11}$$

This condition involves $\boldsymbol{R}_1$, which is unknown yet. We resolve this problem as follows. Since Eq. (10) implies that $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and

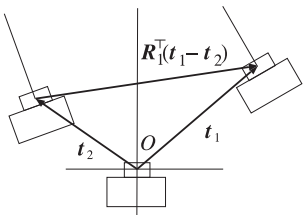

Fig. 6 Triangle condition.

$\boldsymbol{t}_{12}$ are, respectively, null vectors of $\boldsymbol{E}_{01}^\top$, $\boldsymbol{E}_{02}^\top$, and $\boldsymbol{E}_{12}^\top$ in the absence of noise, we compute those unit vectors $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$ that minimize $\|\boldsymbol{E}_{01}^\top\boldsymbol{t}_1\|^2$, $\|\boldsymbol{E}_{02}^\top\boldsymbol{t}_2\|^2$, and $\|\boldsymbol{E}_{12}^\top\boldsymbol{t}_{12}\|^2$, respectively. The solution is given by the unit eigenvectors of $\boldsymbol{E}_{01}\boldsymbol{E}_{01}^\top$, $\boldsymbol{E}_{02}\boldsymbol{E}_{02}^\top$, and $\boldsymbol{E}_{12}\boldsymbol{E}_{12}^\top$ for their smallest eigenvalues. At this stage, the scales and the signs of $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$ are indeterminate. We choose their signs so that

$$\sum_\alpha |\boldsymbol{t}_1, \boldsymbol{x}_\alpha, \boldsymbol{E}_{01}\boldsymbol{x}'_\alpha| > 0, \qquad \sum_\alpha |\boldsymbol{t}_2, \boldsymbol{x}_\alpha, \boldsymbol{E}_{02}\boldsymbol{x}''_\alpha| > 0,$$
$$\sum_\alpha |\boldsymbol{t}_{12}, \boldsymbol{x}'_\alpha, \boldsymbol{E}_{12}\boldsymbol{x}''_\alpha| > 0, \tag{12}$$

where $|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}|$ is the scalar triplet product of $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. The vectors $\boldsymbol{x}_\alpha$, $\boldsymbol{x}'_\alpha$, and $\boldsymbol{x}''_\alpha$ are the coordinates of the αth point represented by vectors as in Eq. (1) with the default focal length f_0 replaced by the computed f, f', and f''. The summations run over the image pairs from which that point is visible. In other words, three-view matching is not necessary; each inequality of Eq. (12) is tested by the matching over two views from which the fundamental matrix was computed for that image pair. The meaning of Eq. (12) will be discussed in Section 6 in detail.

Once the signs of $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$ are determined, we can determine the rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ (next section). Then, substituting the computed $\boldsymbol{R}_1$ into the triangle condition of Eq. (11), we minimize not $\|\boldsymbol{E}_{01}^\top\boldsymbol{t}_1\|^2$, $\|\boldsymbol{E}_{02}^\top\boldsymbol{t}_2\|^2$, and $\|\boldsymbol{E}_{12}^\top\boldsymbol{t}_{12}\|^2$ separately but their sum

$$\|\boldsymbol{E}_{01}^\top\boldsymbol{t}_1\|^2 + \|\boldsymbol{E}_{02}^\top\boldsymbol{t}_2\|^2 + \|\boldsymbol{E}_{12}^\top\boldsymbol{t}_{12}\|^2 = (\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}, \boldsymbol{G}\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}), \tag{13}$$

where we define the 6×6 matrix $\boldsymbol{G}$ by

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{E}_{01}\boldsymbol{E}_{01}^\top + \boldsymbol{R}_1\boldsymbol{E}_{12}\boldsymbol{E}_{12}^\top\boldsymbol{R}_1^\top & -\boldsymbol{R}_1\boldsymbol{E}_{12}\boldsymbol{E}_{12}^\top\boldsymbol{R}_1^\top \\ -\boldsymbol{R}_1\boldsymbol{E}_{12}\boldsymbol{E}_{12}^\top\boldsymbol{R}_1^\top & \boldsymbol{E}_{02}\boldsymbol{E}_{02}^\top + \boldsymbol{R}_1\boldsymbol{E}_{12}\boldsymbol{E}_{12}^\top\boldsymbol{R}_1^\top \end{pmatrix}. \tag{14}$$

As in the case of Eq. (5), Eq. (13) is identically 0 for ideal data, but for noisy observation the left side should theoretically be given weights that reflect the scale and the uncertainty of each term. Here, we ignore them for the ease of computation, assuming that the three terms have approximately the same scales and equal degrees of uncertainty.

As is well known, Eq. (13) is minimized by the unit eigenvector $\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}$ of $\boldsymbol{G}$ for the smallest eigenvalue, which is normalized to $\|\boldsymbol{t}_1\|^2 + \|\boldsymbol{t}_2\|^2 = 1$. The sign is adjusted so that the recomputed $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ align to their original orientations. After $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are thus updated, we compute $\boldsymbol{t}_{12}$ in Eq. (11). From these $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$, we update $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ (next section). Using the resulting $\boldsymbol{R}_1$, we compute the unit eigenvector of $\boldsymbol{G}$ in Eq. (14) to update $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$. We repeat this until they converge; usually, only a few iterations are sufficient.

5. Rotation Computation

Given $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$, we compute $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ that satisfy Eq. (10) by minimizing

$$\|c_1\boldsymbol{E}_{01} - \boldsymbol{t}_1\times\boldsymbol{R}_1\|^2 + \|c_2\boldsymbol{E}_{02} - \boldsymbol{t}_2\times\boldsymbol{R}_2\|^2 + \|c_3\boldsymbol{E}_{12} - \boldsymbol{t}_{12}\times\boldsymbol{R}_1^\top\boldsymbol{R}_2\|^2, \tag{15}$$

where $\|\cdots\|$ means the Frobenius norm and the constants c_1, c_2, and c_3 account for the scale indeterminacy of the essential matrices in Eq. (10). It can be shown [7] (see Appendix A.1), however, that these constants are irrelevant (we can simply let c_1, $= c_2 = c_3 = 1$), and minimizing Eq. (15) is equivalent to maximizing

$$J = \mathrm{tr}[\boldsymbol{K}_{01}^\top\boldsymbol{R}_1] + \mathrm{tr}[\boldsymbol{K}_{02}^\top\boldsymbol{R}_2] + \mathrm{tr}[\boldsymbol{K}_{12}^\top\boldsymbol{R}_1^\top\boldsymbol{R}_2], \tag{16}$$

where $\mathrm{tr}[\,\cdot\,]$ denotes the trace of a matrix and we define

$$\boldsymbol{K}_{01} = -\boldsymbol{t}_1\times\boldsymbol{E}_{01}, \quad \boldsymbol{K}_{02} = -\boldsymbol{t}_2\times\boldsymbol{E}_{02}, \quad \boldsymbol{K}_{12} = -\boldsymbol{t}_{12}\times\boldsymbol{E}_{12}. \tag{17}$$

Here again, we omit the weight that would reflect the scale and the uncertainty of each term of Eqs. (15) and (16), assuming that each term has approximately the same scale and equal degree of uncertainty.

For maximizing Eq. (16), we make use of the following proposition [7]:

Proposition: Let $\boldsymbol{K} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ be the singular value decomposition of matrix $\boldsymbol{K}$. The rotation $\boldsymbol{R}$ that maximizes $\mathrm{tr}[\boldsymbol{K}^\top\boldsymbol{R}]$ is given by

$$\boldsymbol{R} = \boldsymbol{V}\mathrm{diag}(1, 1, \det(\boldsymbol{V}\boldsymbol{U}^\top))\boldsymbol{U}^\top. \tag{18}$$

First, we compute the rotation $\boldsymbol{R}_1$ that maximizes $\mathrm{tr}[\boldsymbol{K}_{01}^\top\boldsymbol{R}_1]$. Equation (16) can be rewritten as

$$J = \mathrm{tr}[\boldsymbol{K}_{01}^\top\boldsymbol{R}_1] + \mathrm{tr}[(\boldsymbol{K}_{02} + \boldsymbol{R}_1\boldsymbol{K}_{12})^\top\boldsymbol{R}_2]. \tag{19}$$

Using the computed $\boldsymbol{R}_1$, we determine the rotation $\boldsymbol{R}_2$ that maximizes $\mathrm{tr}[(\boldsymbol{K}_{02} + \boldsymbol{R}_1\boldsymbol{K}_{12})^\top\boldsymbol{R}_2]$. Equation (16) can also be rewritten as

$$J = \mathrm{tr}[\boldsymbol{K}_{02}^\top\boldsymbol{R}_2] + \mathrm{tr}[(\boldsymbol{K}_{01} + \boldsymbol{R}_2\boldsymbol{K}_{12}^\top)^\top\boldsymbol{R}_1]. \tag{20}$$

Using the computed $\boldsymbol{R}_2$, we determine the rotation $\boldsymbol{R}_1$ that maximizes $\mathrm{tr}[(\boldsymbol{K}_{01} + \boldsymbol{R}_2\boldsymbol{K}_{12}^\top)^\top\boldsymbol{R}_1]$. We iterate this, each time J increasing, until J ceases to increase.

6. Sign Selection Rule

For the above computation of the translations and rotations, we need to resolve a critical issue: the signs of $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$ in Eq. (9) are indeterminate, because the fundamental matrices $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$ have sign indeterminacy [5], [7]. The inequalities in Eq. (12) state that almost all points are "in front" of the three camera pairs *provided* the signs of $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$ are correct [7], [12]. This stems from the fact that the epipolar equation in Eq. (1) holds even if the point is "behind" the cameras and that the essential matrices in Eq. (9) inherit the sign indeterminacy of the fundamental matrices in Eq. (2).

This ambiguity also occurs for two-view reconstruction, but no problem arises. This is because the sign ambiguity of a single fundamental matrix $\boldsymbol{F}$ results in *two* solutions, which is known to produce two 3-D shapes that are mirror images to each other [5], [7]. Hence, it suffices to choose the 3-D shape that are in front of the two cameras in the final reconstruction stage [7]. For three fundamental matrices $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$, however, the sign ambiguity of each of them results in *eight* solutions, which are difficult to disambiguate by a simple means in the final stage.

Imposition of Eq. (12) ensures that the signs of $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$ are *compatible* with the signs of $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$. This means that:

- Either $\boldsymbol{t}_1$ and $\boldsymbol{E}_{01}$ have correct signs or their signs should be reversed.
- Either $\boldsymbol{t}_2$ and $\boldsymbol{E}_{02}$ have correct signs or their signs should be reversed.

Hence, the combination of these yields *four* possibilities, but among them only two are compatible with the triangle condition of Eq. (11).

Note that $\boldsymbol{K}_{01}$ in Eq. (17) uniquely determines $\boldsymbol{R}_1$, since it is unchanged if the signs of $\boldsymbol{t}_1$ and $\boldsymbol{E}_{01}$ are simultaneously reversed. Similarly, $\boldsymbol{K}_{02}$ in Eq. (17) uniquely determines $\boldsymbol{R}_2$. In order to select a correct combination, we assume that the signs of $\boldsymbol{t}_1$ and $\boldsymbol{E}_{01}$ are correct, because one degree of sign indeterminacy, which corresponds to mirror image ambiguity, is inevitable as long as the 3-D reconstruction is based on the epipolar constraint. For selecting the correct signs of $\boldsymbol{t}_2$ and $\boldsymbol{E}_{02}$, we note that $\boldsymbol{E}_{12}^\top \boldsymbol{R}_1^\top (\boldsymbol{t}_2 - \boldsymbol{t}_1) = 0$ in the absence of noise, but this does not hold if the sign of $\boldsymbol{t}_2$ (and hence the sign of $\boldsymbol{E}_{02}$) is not correct. So, we introduce the following rule:

Rule 1: If $\|\boldsymbol{E}_{12}^\top \boldsymbol{R}_1^\top (\boldsymbol{t}_2 - \boldsymbol{t}_1)\| > \|\boldsymbol{E}_{12}^\top \boldsymbol{R}_1^\top (\boldsymbol{t}_2 + \boldsymbol{t}_1)\|$, reverse the signs of $\boldsymbol{t}_2$ and $\boldsymbol{E}_{02}$.

Now, the signs of $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are correctly selected. However, the sign of $\boldsymbol{t}_{12}$ may not be correct: if we compute $\boldsymbol{t}_{12}$ from them using the triangulation condition of Eq. (11), its sign may not be compatible with that computed from $\boldsymbol{E}_{12}$. In that case, $\boldsymbol{K}_{12}$ in Eq. (17) may not be correctly computed. As we see in Eq. (10), $\boldsymbol{E}_{12} \simeq \boldsymbol{t}_{12} \times \boldsymbol{R}_1^\top \boldsymbol{R}_2$ should hold in the absence of noise, so we add the following rule:

Rule 2: If $\|\boldsymbol{E}_{12} - \boldsymbol{t}_{12} \times \boldsymbol{R}_1^\top \boldsymbol{R}_2\| > \|\boldsymbol{E}_{12} + \boldsymbol{t}_{12} \times \boldsymbol{R}_1^\top \boldsymbol{R}_2\|$, we change the sign of $\boldsymbol{K}_{12}$.

Using these two rules, we can correctly compute $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, $\boldsymbol{R}_1$, and $\boldsymbol{R}_2$ *provided* the sign of $\boldsymbol{E}_1$ (and hence the sign of $\boldsymbol{t}_1$) is correct. The remaining ambiguity is resolved in the final stage of 3-D reconstruction by choosing, from among two mirror image solutions, the shape that are in front of all cameras, as will be discussed in Section 9.

7. Summary of the Parameter Computation

We now summarize the computation of the focal lengths f, f', and f'', the translations $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$, and the rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$. It consists of three procedures *focal lengths*, *translations*, and *rotations*:

focal lengths

(1) From the given fundamental matrices $\boldsymbol{F}_{01}$, $\boldsymbol{F}_{02}$, and $\boldsymbol{F}_{12}$, define the functions $K_{01}(x, y)$, $K_{02}(x, y)$, and $K_{12}(x, y)$ as in Eq. (4).

(2) Compute the x, y, and z that minimize Eq. (5) (the equal focal length constraint may or may not be imposed).

(3) Return the focal lengths f, f', and f'' given by Eq. (6) and stop.

translations

(1) Using the focal lengths f, f', and f'' computed by *focal lengths*, compute the essential matrices $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$ by Eq. (9).

(2) Compute the unit eigenvectors $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$ of $\boldsymbol{E}_{01}\boldsymbol{E}_{01}^\top$, $\boldsymbol{E}_{02}\boldsymbol{E}_{02}^\top$, and $\boldsymbol{E}_{12}\boldsymbol{E}_{12}^\top$, respectively, for the smallest eigenvalues.

(3) Adjust the signs of $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$ so that Eq. (12) holds.

(4) From the computed $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$, compute the rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ by *rotations*.

(5) Let $\boldsymbol{t}_1^{(0)} \leftarrow \boldsymbol{t}_1$, $\boldsymbol{t}_2^{(0)} \leftarrow \boldsymbol{t}_2$, and $\boldsymbol{t}_{12}^{(0)} \leftarrow \boldsymbol{t}_{12}$.

(6) Update $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ by computing the unit eigenvector of the matrix $\boldsymbol{G}$ in Eq. (14) for the smallest eigenvalue. If $(\boldsymbol{t}_1, \boldsymbol{t}_1^{(0)}) < 0$ and $(\boldsymbol{t}_2, \boldsymbol{t}_2^{(0)}) < 0$, reverse the signs of $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$.

(7) Let $\boldsymbol{t}_{12} \leftarrow \boldsymbol{R}_1^\top (\boldsymbol{t}_2 - \boldsymbol{t}_1)$.

(8) If $\|\boldsymbol{t}_1 \times \boldsymbol{t}_1^{(0)}\| \approx 0$, $\|\boldsymbol{t}_2 \times \boldsymbol{t}_2^{(0)}\| \approx 0$, and $\|\boldsymbol{t}_{12} \times \boldsymbol{t}_{12}^{(0)}\| \approx 0$, return $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, $\boldsymbol{R}_1$, and $\boldsymbol{R}_2$ and stop. Else, go back to Step (4).

rotations

(1) From the given $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{t}_{12}$, compute the matrices $\boldsymbol{K}_{01}$, $\boldsymbol{K}_{02}$, and $\boldsymbol{K}_{12}$ of Eq. (17).

(2) Compute the rotation $\boldsymbol{R}_1$ via the SVD of $\boldsymbol{K}_{01}$, and let $\boldsymbol{R}_1^{(0)} \leftarrow \boldsymbol{R}_1$.

(3) Compute the rotation $\boldsymbol{R}_2$ via the SVD of $\boldsymbol{K}_{02}$, and let $\boldsymbol{R}_2^{(0)} \leftarrow \boldsymbol{R}_2$.

(4) If $\|\boldsymbol{E}_{12}^\top \boldsymbol{R}_1^\top (\boldsymbol{t}_2 - \boldsymbol{t}_1)\| > \|\boldsymbol{E}_{12}^\top \boldsymbol{R}_1^\top (\boldsymbol{t}_2 + \boldsymbol{t}_1)\|$, reverse the signs of $\boldsymbol{t}_2$ and $\boldsymbol{E}_{02}$.

(5) If $\|\boldsymbol{E}_{12} - \boldsymbol{t}_{12} \times \boldsymbol{R}_1^\top \boldsymbol{R}_2\| > \|\boldsymbol{E}_{12} + \boldsymbol{t}_{12} \times \boldsymbol{R}_1^\top \boldsymbol{R}_2\|$, reverse the sign of $\boldsymbol{K}_{12}$.

(6) Compute the rotation $\boldsymbol{R}_2$ via the SVD of $\boldsymbol{K}_{02} + \boldsymbol{R}_1 \boldsymbol{K}_{12}$. Using that $\boldsymbol{R}_2$, compute the rotation $\boldsymbol{R}_1$ via the SVD of $\boldsymbol{K}_{01} + \boldsymbol{R}_2 \boldsymbol{K}_{12}^\top$.

(7) If $\boldsymbol{R}_1 \approx \boldsymbol{R}_1^{(0)}$ and $\boldsymbol{R}_2 \approx \boldsymbol{R}_2^{(0)}$, return $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ and stop. Else, let $\boldsymbol{R}_1^{(0)} \leftarrow \boldsymbol{R}_1$ and $\boldsymbol{R}_2^{(0)} \leftarrow \boldsymbol{R}_2$, and go back to Step (6).

8. Optimal Correction of Matching Points

After all the parameters (the focal lengths f, f', and f'', the translations $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$, and the rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$) are computed, we determine the 3-D positions of individual points. As pointed out earlier, our method does not require point matching over three images; all we need is three fundamental matrices separately computed from pair-wise point matching. So, point data are classified into two types: points that are matched only between a pair of images, and points that happen to be matched over all the three images. For points of the former type, we use the standard two-view reconstruction method [5], [12]. We now describe the procedure for points of the latter type.

First, we recompute the essential matrices $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$ as follows:

$$\boldsymbol{E}_{01} = \boldsymbol{t}_1 \times \boldsymbol{R}_1, \qquad \boldsymbol{E}_{02} = \boldsymbol{t}_2 \times \boldsymbol{R}_2,$$
$$\boldsymbol{E}_{12} = \Big(\boldsymbol{R}_1^\top (\boldsymbol{t}_2 - \boldsymbol{t}_1)\Big) \times \boldsymbol{R}_1^\top \boldsymbol{R}_2. \tag{21}$$

Then, we optimally correct the observed image positions, i.e., optimally correct the vectors $\boldsymbol{x}$, $\boldsymbol{x}'$, and $\boldsymbol{x}''$, constructed from the image coordinates as in Eq. (1) with the default focal length f_0 replaced by the computed f, f', and f'', to $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}'$, and $\hat{\boldsymbol{x}}''$, respectively. The correction is done so that the corrected positions *exactly* satisfies the epipolar equations for the essential matrices $\boldsymbol{E}_{01}$, $\boldsymbol{E}_{02}$, and $\boldsymbol{E}_{12}$ recomputed by Eq. (21) using the computed

translations $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ and rotations $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$. To be specific, we modify $\boldsymbol{x}$, $\boldsymbol{x}'$, and $\boldsymbol{x}''$ to $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}'$, and $\hat{\boldsymbol{x}}''$, respectively, in such a way that $\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|^2+\|\hat{\boldsymbol{x}}'-\boldsymbol{x}'\|^2+\|\hat{\boldsymbol{x}}''-\boldsymbol{x}''\|^2$ is minimized subject to $(\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') = (\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') = (\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') = 0$. For two views, this is nothing but the optimal triangulation procedure of Kanatani et al. [10], [13]. For points matched over three views, this process is extend to the following form (see Appendix A.2 for the derivation):

(1) Let $E_0 = \infty$ (a sufficiently large number), $\hat{\boldsymbol{x}} = \boldsymbol{x}$, $\hat{\boldsymbol{x}}' = \boldsymbol{x}'$, $\hat{\boldsymbol{x}}'' = \boldsymbol{x}''$, and $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}' = \tilde{\boldsymbol{x}}'' = \boldsymbol{0}$.

(2) Compute λ, λ', and λ'' by solving the following linear equation, where $\boldsymbol{P}_k = \mathrm{diag}(1,1,0)$:

$$\begin{pmatrix}
(\boldsymbol{P}_k\boldsymbol{E}_{01}\hat{\boldsymbol{x}}', \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') + (\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}}) & (\boldsymbol{P}_k\boldsymbol{E}_{01}\hat{\boldsymbol{x}}', \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') & (\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}}, \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') \\
(\boldsymbol{P}_k\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'', \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') & (\boldsymbol{P}_k\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'', \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') + (\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}}) & (\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}}, \boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}') \\
(\boldsymbol{P}_k\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'', \boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}}) & (\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}', \boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}}) & (\boldsymbol{P}_k\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') + (\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}')
\end{pmatrix}
\begin{pmatrix} \lambda \\ \lambda' \\ \lambda'' \end{pmatrix}$$
$$= \begin{pmatrix}
(\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') + (\tilde{\boldsymbol{x}}, \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') + (\tilde{\boldsymbol{x}}', \boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}}) \\
(\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') + (\tilde{\boldsymbol{x}}, \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') + (\tilde{\boldsymbol{x}}'', \boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}}) \\
(\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') + (\tilde{\boldsymbol{x}}', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') + (\tilde{\boldsymbol{x}}'', \boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}')
\end{pmatrix}. \tag{22}$$

(3) Update $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{x}}'$, and $\tilde{\boldsymbol{x}}''$ as follows:

$$\begin{aligned}
\tilde{\boldsymbol{x}} &\leftarrow \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}\hat{\boldsymbol{x}}' + \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'', \\
\tilde{\boldsymbol{x}}' &\leftarrow \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}} + \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'', \\
\tilde{\boldsymbol{x}}'' &\leftarrow \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}} + \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}'.
\end{aligned} \tag{23}$$

(4) Update $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}'$, and $\hat{\boldsymbol{x}}''$ as follows:

$$\hat{\boldsymbol{x}} \leftarrow \boldsymbol{x} - \tilde{\boldsymbol{x}}, \quad \hat{\boldsymbol{x}}' \leftarrow \boldsymbol{x} - \tilde{\boldsymbol{x}}', \quad \hat{\boldsymbol{x}}' \leftarrow \boldsymbol{x} - \tilde{\boldsymbol{x}}''. \tag{24}$$

(5) Compute the reprojection error $E = \|\tilde{\boldsymbol{x}}\|^2 + \|\tilde{\boldsymbol{x}}'\|^2 + \|\tilde{\boldsymbol{x}}''\|^2$. If $E \approx E_0$, then return $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}'$, $\hat{\boldsymbol{x}}''$ and stop. Else, let $E_0 \leftarrow E$ and go back to Step (2).

For points matched only over two images, we use the method described in Refs. [10], [13].

9. Computation of 3-D Position

After the observed image positions are optimally corrected by the procedure described above, we compute their 3-D positions. The projection matrices $\boldsymbol{P}$, $\boldsymbol{P}'$, and $\boldsymbol{P}''$ of the three cameras have the form

$$\begin{aligned}
\boldsymbol{P} &= \mathrm{diag}(1,1,\frac{f_0}{f})\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\end{pmatrix}, \\
\boldsymbol{P}' &= \mathrm{diag}(1,1,\frac{f_0}{f'})\begin{pmatrix}\boldsymbol{R}_1^\top & -\boldsymbol{R}_1^\top\boldsymbol{t}_1\end{pmatrix}, \\
\boldsymbol{P}'' &= \mathrm{diag}(1,1,\frac{f_0}{f''})\begin{pmatrix}\boldsymbol{R}_2^\top & -\boldsymbol{R}_2^\top\boldsymbol{t}_2\end{pmatrix}.
\end{aligned} \tag{25}$$

Let $\boldsymbol{X}_\alpha = (X_\alpha, Y_\alpha, Z_\alpha)^\top$ be the 3-D position of the αth point, and $\hat{\boldsymbol{x}}_\alpha$, $\hat{\boldsymbol{x}}'_\alpha$, $\hat{\boldsymbol{x}}''_\alpha$ its 2-D positions in the 0th, 1st, and 2nd views, respectively, after the optimal correction. The following projection relationships hold:

$$\hat{\boldsymbol{x}}_\alpha \simeq \boldsymbol{P}\begin{pmatrix}\boldsymbol{X}_\alpha \\ 1\end{pmatrix}, \quad \hat{\boldsymbol{x}}'_\alpha \simeq \boldsymbol{P}'\begin{pmatrix}\boldsymbol{X}_\alpha \\ 1\end{pmatrix}, \quad \hat{\boldsymbol{x}}''_\alpha \simeq \boldsymbol{P}''\begin{pmatrix}\boldsymbol{X}_\alpha \\ 1\end{pmatrix}. \tag{26}$$

By canceling the denominators, we obtain the following linear equations in $\boldsymbol{X}_\alpha$:

$$\begin{pmatrix}
\hat{x}_\alpha P_{31} - f_0 P_{11} & \hat{x}_\alpha P_{32} - f_0 P_{12} & \hat{x}_\alpha P_{33} - f_0 P_{13} \\
\hat{y}_\alpha P_{31} - f_0 P_{21} & \hat{y}_\alpha P_{32} - f_0 P_{22} & \hat{y}_\alpha P_{33} - f_0 P_{23} \\
\hat{x}'_\alpha P'_{31} - f_0 P'_{11} & \hat{x}'_\alpha P'_{32} - f_0 P'_{12} & \hat{x}'_\alpha P'_{33} - f_0 P'_{13} \\
\hat{y}'_\alpha P'_{31} - f_0 P'_{21} & \hat{y}'_\alpha P'_{32} - f_0 P'_{22} & \hat{y}'_\alpha P'_{33} - f_0 P'_{23} \\
\hat{x}''_\alpha P''_{31} - f_0 P''_{11} & \hat{x}''_\alpha P''_{32} - f_0 P''_{12} & \hat{x}''_\alpha P''_{33} - f_0 P''_{13} \\
\hat{y}''_\alpha P''_{31} - f_0 P''_{21} & \hat{y}''_\alpha P''_{32} - f_0 P''_{22} & \hat{y}''_\alpha P''_{33} - f_0 P''_{23}
\end{pmatrix}
\begin{pmatrix} X_\alpha \\ Y_\alpha \\ Z_\alpha \end{pmatrix} = -\begin{pmatrix}
\hat{x}_\alpha P_{34} - f_0 P_{14} \\
\hat{y}_\alpha P_{34} - f_0 P_{24} \\
\hat{x}'_\alpha P'_{34} - f_0 P'_{14} \\
\hat{y}'_\alpha P'_{34} - f_0 P'_{24} \\
\hat{x}''_\alpha P''_{34} - f_0 P''_{14} \\
\hat{y}''_\alpha P''_{34} - f_0 P''_{24}
\end{pmatrix}. \tag{27}$$

Since Eq. (26) *exactly* holds due to the optimal correction procedure, we can choose any three from the six equations in Eq. (27) to solve for $\boldsymbol{X}_\alpha$. Alternatively, we can use all the equations by least squares. If the point is visible only in two views, we remove the corresponding rows and columns from the above equation.

So far, we have assumed that the sign of $\boldsymbol{E}_{01}$ is correct (Section 6). If its sign is wrong (hence the signs of $\boldsymbol{E}_{02}$ and $\boldsymbol{E}_{12}$ are also wrong), the reconstructed shape is a mirror image of the true shape locating *behind* the cameras [5], [7]. Hence, if

$$\sum_\alpha^N \mathrm{sgn}(Z_\alpha) < 0, \tag{28}$$

for the visible points from the 0th camera, where $\mathrm{sgn}(x)$ returns 1, −1, and 0 according to $x > 0$, $x < 0$, and $x = 0$, respectively, we reverse the signs of all $(X_\alpha, Y_\alpha, Z_\alpha)^\top$.

10. Simulation Experiments

Figure 7 shows three simulated views (0th, 1st, and 2nd from left) of a grid surface. The frame size is assumed to be 800×800 pixels and the focal lengths $f = f' = f'' = 600$ pixels. We added independent Gaussian random noise of mean 0 and standard deviation σ pixels to the x and y coordinates of each grid point and conducted calibration and 3-D reconstruction. For a computed focal length f, we evaluated the difference $\Delta f = f - \bar{f}$ from its true value $\bar{f}$. If the computation failed ("imaginary focal lengths"), we let $f = 0$. Since the absolute scale of translation is indeterminate, we evaluated for a computed translation $\boldsymbol{t}$ the angle $\Delta\theta = \cos^{-1}(\boldsymbol{t}, \bar{\boldsymbol{t}})/\|\boldsymbol{t}\| \cdot \|\bar{\boldsymbol{t}}\|$ (in degree) it makes from its true value $\bar{\boldsymbol{t}}$. If the computation failed due to imaginary focal lengths,

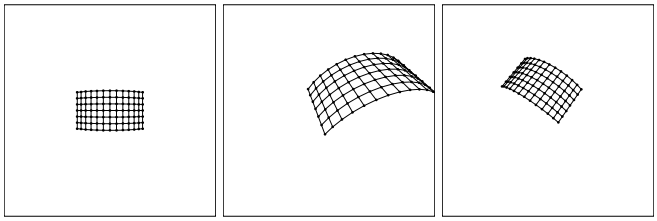

Fig. 7 The 0th, 1st, and 2nd views a simulated curved grid surface. The 0th and the 2nd cameras are nearly in a fixating configuration.

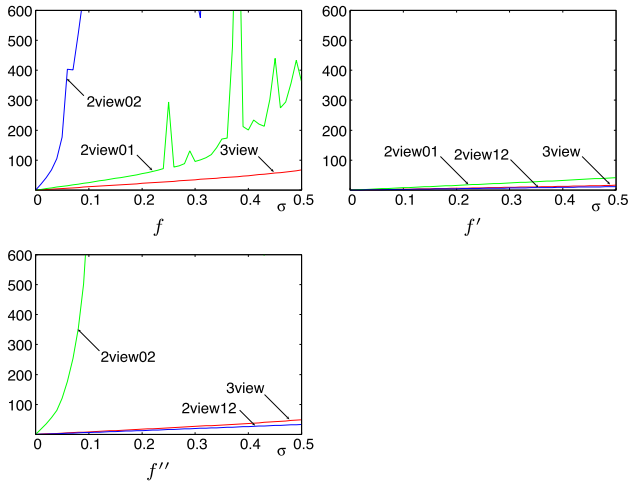

Fig. 8 The RMS error of focal length computation for σ, where "2view01," etc. denote the values computed from the 0th-1st image pair, etc., and "3view" means the value computed from the three views.

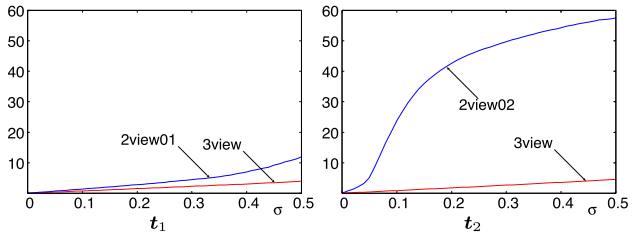

Fig. 9 The RMS error (in degree) of translation computation for σ. "2view01," etc. denote the values computed from the 0th-1st image pair, etc., and "3view" means the value computed from the three views.

we let $\Delta\theta = 90°$. For a computed rotation $\boldsymbol{R}$, we evaluated the angle $\Delta\Omega$ (in degree) of the relative rotation $\boldsymbol{R}\bar{\boldsymbol{R}}^\top$ from the true value $\bar{\boldsymbol{R}}$. If the computation failed due to imaginary focal lengths, we let $\Delta\Omega = 90°$. Then, we evaluated the RMSs

$$E_f = \sqrt{\frac{1}{K}\sum_{a=1}^{K}\Delta f_a^2}, \qquad E_{\boldsymbol{t}} = \sqrt{\frac{1}{K}\sum_{a=1}^{K}\Delta\theta_a^2},$$
$$E_{\boldsymbol{R}} = \sqrt{\frac{1}{K}\sum_{a=1}^{K}\Delta\Omega_a^2}, \tag{29}$$

over $K = 10{,}000$ independent trials, each time using different noise, where the subscript a indicates the value of the ath trial.

Figure 8 compares the accuracy of focal lengths computed from two views and from three views. We see that f and f'' computed from the 0th-2nd image pair have large errors. This is because the 0th and 2nd cameras are nearly in a fixating configuration. The large fluctuations of the plots indicate the occurrence of imaginary focal lengths. However, we can obtain accurate values for all the focal lengths if we use three images. In this noise range, no imaginary focal lengths occurred for three-view computation. **Figure 9** and **Fig. 10** compare the accuracy of translation and rotation. The error is large for the values computed from the 0th-2nd image pair due to the low accuracy of the focal length computation from them. However, we can obtain accurate values by using three views despite the fixating camera configuration of the 0th and 2nd cameras.

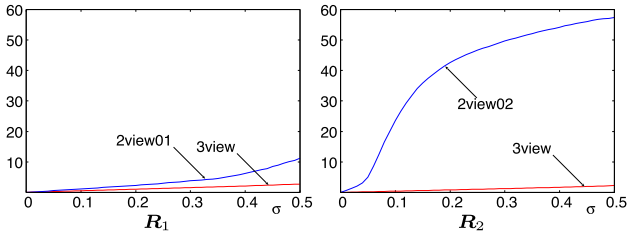

Fig. 10 The RMS error (in degree) of rotation computation for σ. "2view01," etc. denote the values computed from the 0th-1st image pair, etc., and "3view" means the value computed from the three views.

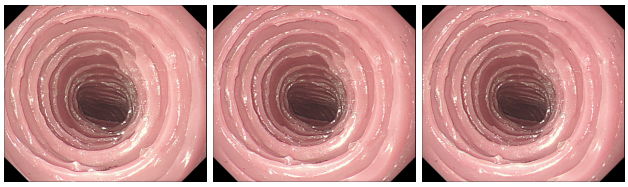

Fig. 11 Three consecutive frames of endoscopic intestine tract images.

11. Endoscopic Image Experiments

Figure 11 shows three consecutive frames of intestine tract images taken by an endoscope receding along the tract. These images were provided by medical researchers without any information about the endoscopic camera used. It is well known that if a camera is moved forward or backward, two-view reconstruction frequently fails because any two camera positions are nearly in a fixating configuration, frequently resulting in imaginary focal lengths. Hence, this is a good testbed for examining the degeneracy tolerance capability of our method. At the same time, our method, if successful, would bring about a new medical application of reconstructing 3-D structures from endoscopic images.

We extracted feature points and matched them between each pair of frames, using the method of Hirai et al. [6]. **Figure 12** (a) shows the reconstruction from the three frames in Fig. 11. For comparison, Fig. 12 (b), (c), (d) show the two-view reconstructions from the 0th-1st frame pair, the 0th-2nd frame pair, and the 1st-2nd frame pair, respectively; only those points viewed in the corresponding image pairs are reconstructed.

Since the ground truth is not known, we cannot tell which of (a), (b), (c), and (d) is the most accurate. As we can see, however, the three-view reconstruction (a) provides a detailed shape in a longer range along the tract with a larger number of points than the two-view reconstructions (b), (c), and (d). Ideally, the superimposition of (b), (c), and (d) should coincide with (a) if we correctly adjust the scale of the two-view reconstructions in (b), (c), and (d) (recall that the scale is indeterminate in each reconstruction). For real data, however, the two-view reconstructions do not necessarily agree with the three-view reconstruction. In this sense, our three-view reconstruction can be viewed as automatically adjusting the scales of two-view reconstructions and optimally merging them into a single shape.

Figure 13 shows another set of three consecutive frames of intestine tract images, and **Fig. 14** shows the reconstruction from them. Figure 14 (a) shows the resulting three-view reconstruction. In this case, two-view reconstruction was possible only from the 1st-2nd frame pair (Fig. 14 (b)); the computation failed both

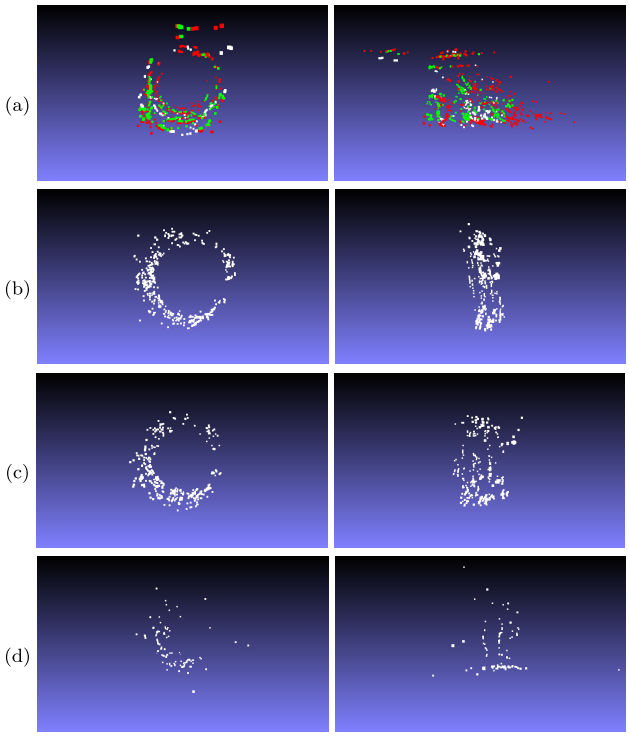

Fig. 12 Front views (left) and side views (right) of the 3-D reconstruction from the data of Fig. 11. (a) Using the three frames. Different colors indicate different image pairs they originate from. (b) Using the 0th-1st frame pair. (c) Using the 0th-2nd frame pair. (d) Using the 1st-2nd frame pair.

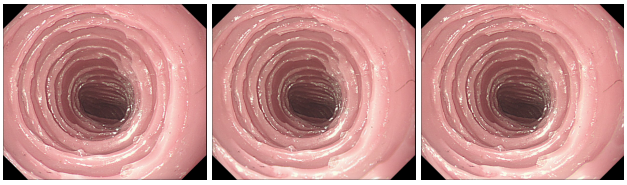

Fig. 13 Another set of three consecutive frames of endoscopic intestine tract images.

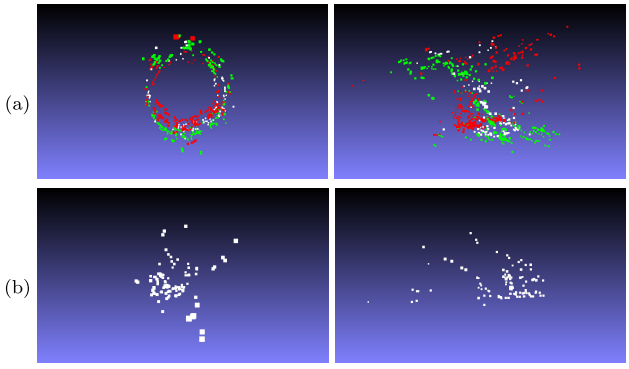

Fig. 14 Front views (left) and side views (right) of the 3-D reconstruction from the data set in Fig. 13. (a) Using the three frames. (b) Using the 1st-2nd frame pair. Reconstruction from the 0th-1st frame pair and reconstruction from the 0th-2nd frame pair both fail.

for the 0th-1st frame pair and for the 0th-2nd frame pair due to imaginary focal lengths. Yet, using three images, we can accurately compute the 3-D positions of all pairwise matched points and obtain a detailed structure in a longer range along the tract.

Figure 15 shows the result of applying bundle adjustment to the shape shown in Fig. 12 (a): red points by three-view reconstruction are corrected to white points (the details of our program

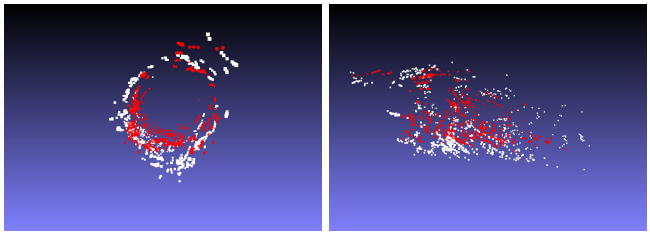

Fig. 15 Front view (left) and side view (right) of bundle adjustment. Red points correspond to the result in Fig. 12 (a), which are corrected to white points.

is described in Ref. [11]). The reprojection error reduced from 204.89 to 0.47058, which means that the distance between observed and predicted image positions reduced about 0.05 times on average. This reduction is partly due to optimizing the focal lengths and principal points of the three views independently, although the camera is the same, i.e., the same focal length and principal point. Since the ground truth is unknown, it is unable to tell if the accuracy has really improved by bundle adjustment. Bundle adjustment is very effective when object is viewed from many different viewpoints. For endoscopic images, however, one feature point is observed mostly by two or three consecutive frames. In such a case, we cannot expect much information is added by bundle adjustment.

12. Concluding Remarks

We have presented a new method for initializing 3-D reconstruction from three views, generating a candidate solution to be refined later. Our main focus is to prevent the imaginary focal length degeneracy, which two-view reconstruction frequently suffers. Our method does not require correspondences among the three images; all we need is three fundamental matrices separately computed from pair-wise correspondences. We exploited the redundant information provided by the three fundamental matrices to optimize the camera parameters and the 3-D structure. We conducted numerical simulation and observed that imaginary focal lengths never occurred in the experimented noise range while two-view computation frequently failed, resulting in higher average accuracy of our method than two-view reconstruction. We also tested the degeneracy tolerance capability of our method by using endoscopic intestine tract images, noting that the camera configuration is almost always near degeneracy. We observed that our method produced a more detailed structure in a wider range than combining pairwise two-view reconstructions, for which computation frequently fails. We also observed how our three-view reconstruction is refined by bundle adjustment. Our method is expected to broaden medical applications of endoscopic images.

Acknowledgments This work was supported in part by JSPS Grant-in-Aid for Young Scientists (B 23700202), for Scientific Research (C 24500202), and for Challenging Exploratory Research (24650086) and the 2013 Material and Device Joint Project (2013355).

References

[1] Bougnoux, S.: From projective to Euclidean space under any practical situation, a criticism of self-calibration, *Proc. 6th Int. Conf. Comput.*

Vis., Bombay, India, pp.790–796 (1998).
[2] Goldberger, J.: Reconstructing camera projection matrices from multiple pairwise overlapping views, *Comput. Vis. Image Understanding*, Vol.97, pp.283–296 (2005).
[3] Hartley, R.: Estimation of relative camera positions for uncalibrated cameras, *Proc. 2nd European Conf. Comput. Vis.*, Santa Margherita Ligure, Italy, pp.579–587 (1992).
[4] Hartley, R. and Silpa-Anan, C.: Reconstruction from two views using approximate calibration, *Proc. 5nd Asian Conf. Comput. Vis.*, Melbourne, Australia, pp.338–343 (2002).
[5] Hartley, R. and Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, Cambridge, U.K. (2003).
[6] Hirai, K., Kanazawa, Y., Sagawa, R. and Yagi, Y.: Endoscopic image matching for reconstructing the 3-D structure of the intestines, *Med. Imag. Tech.*, Vol.29, No.1, pp.36–46 (2011).
[7] Kanatani, K.: *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, U.K. (1993).
[8] Kanatani, K. and Matsunaga, C.: Closed-form expression for focal lengths from the fundamental matrix, *Proc. 4th Asian Conf. Comput. Vis.*, Taipei, Taiwan, Vol.1, pp.128–133 (2000).
[9] Kanatani, K., Nakatsuji, A. and Sugaya, Y: Stabilizing the focal length computation for 3-D reconstruction from two uncalibrated views, *Int. J. Comput. Vis.*, Vol.66, No.2, pp.109–122 (2006).
[10] Kanatani, K. and Sugaya, Y.: Compact fundamental matrix computation, *IPSJ Trans. Comput. Vis. Appl.*, Vol.2, pp.59–70 (2010).
[11] Kanatani, K. and Sugaya, Y.: Implementation and evaluation of bundle adjustment for 3-D reconstruction, *Proc. 17th Symp. Sensing via Imaging Information*, Yokohama, Japan, pp.IS4/02/1–IS4/02/8 (2011).
[12] Kanatani, K., Sugaya, Y. and Kanazawa, Y.: Latest algorithms for 3-D reconstruction from two views, Chen, C.H. (Ed.), *Handbook of Pattern Recognition and Computer Vision*, 4th ed., World Scientific Publishing, pp.201–234 (2009).
[13] Kanatani, K., Sugaya, Y. and Niitsuma, H.: Triangulation from two views revisited: Hartley-Sturm vs. optimal correction, *Proc. 19th British Machine Vis. Conf.*, Leeds, U.K., pp.173–182 (2008).
[14] Kanazawa, Y., Sugaya, Y. and Kanatani, K.: Initializing 3-D reconstruction from three views using three fundamental matrices, Huang, F. and Sugimoto, A. (Eds.), *PSIVT 2013 Workshops*, LNCS 8334, Springer, Berlin/Heidelberg, Germany, pp.169–180 (2014).
[15] Lourakis, M.I.A. and Argyros, A.A.: SBA: A software package for generic sparse bundle adjustment, *ACM Trans. Math. Software*, Vol.36, No.1 (2), pp.1–30 (2009).
[16] Pollefeys, M., Koch, R. and Gool, L.V.: Self-calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters, *Int. J. Comput. Vis.*, Vol.32, No.1, pp.7–25 (1999).
[17] Snavely, N., Seitz, S. and Szeliski, R.: Photo tourism: Exploring photo collections in 3d, *ACM Trans. Graphics*, Vol.25, No.8, pp.835–846 (1995).
[18] Snavely, N., Seitz, S. and Szeliski, R.: Modeling the world from Internet photo collections, *Int. J. Comput. Vis.*, Vol.80, No.2, pp.189–210 (2008).

Appendix

A.1 Rotation Computation

Consider the computation of the rotation $\boldsymbol{R}$ that minimizes the square Frobenius norm

$$\|c\boldsymbol{E} - \boldsymbol{t}\times\boldsymbol{R}\|^2 \tag{A.1}$$

for given matrix $\boldsymbol{E}$ and vector $\boldsymbol{t}$ and for *some* scale constant c. Equation (A.1) is rewritten as

$$c^2\|\boldsymbol{E}\|^2 - 2c\mathrm{tr}[\boldsymbol{E}^\top(\boldsymbol{t}\times\boldsymbol{R})] + \|\boldsymbol{t}\times\boldsymbol{R}\|^2. \tag{A.2}$$

The second term can be written as

$$-2c\mathrm{tr}[\boldsymbol{E}^\top(\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R}] = -2c\mathrm{tr}[((\boldsymbol{t}\times\boldsymbol{I})^\top\boldsymbol{E})^\top\boldsymbol{R}] = -2c\mathrm{tr}[\boldsymbol{K}^\top\boldsymbol{R}] \tag{A.3}$$

where we let

$$\boldsymbol{K} = (\boldsymbol{t}\times\boldsymbol{I})^\top\boldsymbol{E} = -(\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{E} = -\boldsymbol{t}\times\boldsymbol{E}. \tag{A.4}$$

If we recall that $\|(\cdots)\boldsymbol{R}\|^2 = \|\cdots\|^2$ for the Frobenius norm, the third term of Eq. (A.2) is

$$\|(\boldsymbol{t}\times\boldsymbol{I})\boldsymbol{R}\|^2 = \|\boldsymbol{t}\times\boldsymbol{I}\|^2 = 2\|\boldsymbol{t}\|^2. \tag{A.5}$$

Thus, Eq. (A.2) becomes

$$c^2 - 2c\mathrm{tr}[\boldsymbol{K}^\top\boldsymbol{R}] + 2\|\boldsymbol{t}\|^2 \tag{A.6}$$

Here, we assume that the sign of $\boldsymbol{E}$ is correctly chosen so that $c > 0$ (the issue of sign selection is discussed in Section 6). Then, minimizing Eq. (A.2) is equivalent to maximizing $\mathrm{tr}[\boldsymbol{K}^\top\boldsymbol{R}]$. Thus, the scale constant c is irrelevant as long as the sign of $\boldsymbol{E}$ is correctly chosen.

A.2 Derivation of Optimal Correction

We correct vectors $\boldsymbol{x}$, $\boldsymbol{x}'$, and $\boldsymbol{x}''$ to $\boldsymbol{x}-\Delta\boldsymbol{x}$, $\boldsymbol{x}'-\Delta\boldsymbol{x}'$, and $\boldsymbol{x}''-\Delta\boldsymbol{x}''$, respectively, such that the following epipolar equations satisfy

$$\begin{aligned}&(\boldsymbol{x}-\Delta\boldsymbol{x}, \boldsymbol{E}_{01}(\boldsymbol{x}'-\Delta\boldsymbol{x}')) = 0, \quad (\boldsymbol{x}-\Delta\boldsymbol{x}, \boldsymbol{E}_{02}(\boldsymbol{x}''-\Delta\boldsymbol{x}'')) = 0,\\ &(\boldsymbol{x}'-\Delta\boldsymbol{x}', \boldsymbol{E}_{12}(\boldsymbol{x}''-\Delta\boldsymbol{x}'')) = 0.\end{aligned} \tag{A.7}$$

Expanding these with respect to the correction terms $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}'$, and $\Delta\boldsymbol{x}''$ and ignoring higher order terms, we obtain

$$\begin{aligned}&(\Delta\boldsymbol{x}, \boldsymbol{E}_{01}\boldsymbol{x}') + (\Delta\boldsymbol{x}', \boldsymbol{E}_{01}^\top\boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{E}_{01}\boldsymbol{x}'),\\ &(\Delta\boldsymbol{x}, \boldsymbol{E}_{02}\boldsymbol{x}'') + (\Delta\boldsymbol{x}'', \boldsymbol{E}_{02}^\top\boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{E}_{02}\boldsymbol{x}''),\\ &(\Delta\boldsymbol{x}', \boldsymbol{E}_{12}\boldsymbol{x}'') + (\Delta\boldsymbol{x}'', \boldsymbol{E}_{12}^\top\boldsymbol{x}') = (\boldsymbol{x}', \boldsymbol{E}_{12}\boldsymbol{x}'').\end{aligned} \tag{A.8}$$

Among those $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}'$, and $\Delta\boldsymbol{x}''$ that satisfy these, we choose the values that minimizes $\|\Delta\boldsymbol{x}\|^2 + \|\Delta\boldsymbol{x}'\|^2 + \|\Delta\boldsymbol{x}''\|^2$. Since the correction occurs in the image plane, the third component of $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}'$, and $\Delta\boldsymbol{x}''$ are zero. In other words,

$$(\boldsymbol{k}, \Delta\boldsymbol{x}) = 0, \qquad (\boldsymbol{k}, \Delta\boldsymbol{x}') = 0, \qquad (\boldsymbol{k}, \Delta\boldsymbol{x}'') = 0, \tag{A.9}$$

where $\boldsymbol{k} = (0, 0, 1)^\top$. Introducing Lagrange multipliers for Eqs. (A.8) and (A.9), we consider

$$\begin{aligned}&\frac{1}{2}\Big(\|\Delta\boldsymbol{x}\|^2 + \|\Delta\boldsymbol{x}'\|^2 + \|\Delta\boldsymbol{x}''\|^2\Big)\\ &-\lambda\Big((\Delta\boldsymbol{x}, \boldsymbol{E}_{01}\boldsymbol{x}') + (\Delta\boldsymbol{x}', \boldsymbol{E}_{01}^\top\boldsymbol{x})\Big)\\ &-\lambda'\Big((\Delta\boldsymbol{x}, \boldsymbol{E}_{02}\boldsymbol{x}'') + (\Delta\boldsymbol{x}'', \boldsymbol{E}_{02}^\top\boldsymbol{x})\Big)\\ &-\lambda''\Big((\Delta\boldsymbol{x}', \boldsymbol{E}_{12}\boldsymbol{x}'') + (\Delta\boldsymbol{x}'', \boldsymbol{E}_{12}^\top\boldsymbol{x}')\Big)\\ &-\mu(\boldsymbol{k}, \Delta\boldsymbol{x}) - \mu'(\boldsymbol{k}, \Delta\boldsymbol{x}') - \mu''(\boldsymbol{k}, \Delta\boldsymbol{x}'').\end{aligned} \tag{A.10}$$

Differentiating this with respect to $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}'$, and $\Delta\boldsymbol{x}''$ and letting the result be 0, we have

$$\begin{aligned}\Delta\boldsymbol{x} &= \lambda\boldsymbol{E}_{01}\boldsymbol{x}' + \lambda'\boldsymbol{E}_{02}\boldsymbol{x}'' + \mu\boldsymbol{k},\\ \Delta\boldsymbol{x}' &= \lambda\boldsymbol{E}_{01}^\top\boldsymbol{x} + \lambda''\boldsymbol{E}_{12}\boldsymbol{x}'' + \mu'\boldsymbol{k},\\ \Delta\boldsymbol{x}'' &= \lambda'\boldsymbol{E}_{02}^\top\boldsymbol{x} + \lambda''\boldsymbol{E}_{12}^\top\boldsymbol{x}' + \mu''\boldsymbol{k}.\end{aligned} \tag{A.11}$$

Multiplying these by $\boldsymbol{P}_k = \mathrm{diag}(1, 1, 0)$ from left on both sides and noting that $\boldsymbol{P}_k\Delta\boldsymbol{x} = \Delta\boldsymbol{x}$, $\boldsymbol{P}_k\Delta\boldsymbol{x}' = \Delta\boldsymbol{x}'$, $\boldsymbol{P}_k\Delta\boldsymbol{x}'' = \Delta\boldsymbol{x}''$, and $\boldsymbol{P}_k\boldsymbol{k} = \boldsymbol{0}$, we obtain

$$\begin{aligned}\Delta\boldsymbol{x} &= \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}\boldsymbol{x}' + \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}\boldsymbol{x}'',\\ \Delta\boldsymbol{x}' &= \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\boldsymbol{x} + \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}\boldsymbol{x}'',\end{aligned}$$

$$\Delta\boldsymbol{x}'' = \lambda' \boldsymbol{P}_k \boldsymbol{E}_{02}^\top \boldsymbol{x} + \lambda'' \boldsymbol{P}_k \boldsymbol{E}_{12}^\top \boldsymbol{x}'. \tag{A.12}$$

Substituting these into Eq. (A.8), we obtain the following linear equation in λ, λ', and λ'':

$$\begin{pmatrix} (\boldsymbol{P}_k\boldsymbol{E}_{01}\boldsymbol{x}', \boldsymbol{E}_{01}\boldsymbol{x}') + (\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\boldsymbol{x}, \boldsymbol{E}_{01}^\top\boldsymbol{x}) & (\boldsymbol{P}_k\boldsymbol{E}_{01}\boldsymbol{x}', \boldsymbol{E}_{02}\boldsymbol{x}'') & (\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\boldsymbol{x}, \boldsymbol{E}_{12}\boldsymbol{x}'') \\ (\boldsymbol{P}_k\boldsymbol{E}_{02}\boldsymbol{x}'', \boldsymbol{E}_{01}\boldsymbol{x}') & (\boldsymbol{P}_k\boldsymbol{E}_{02}\boldsymbol{x}'', \boldsymbol{E}_{02}\boldsymbol{x}'') + (\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\boldsymbol{x}, \boldsymbol{E}_{02}^\top\boldsymbol{x}) & (\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\boldsymbol{x}, \boldsymbol{E}_{12}^\top\boldsymbol{x}') \\ (\boldsymbol{P}_k\boldsymbol{E}_{12}\boldsymbol{x}'', \boldsymbol{E}_{01}^\top\boldsymbol{x}) & (\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\boldsymbol{x}', \boldsymbol{E}_{02}^\top\boldsymbol{x}) & (\boldsymbol{P}_k\boldsymbol{E}_{12}\boldsymbol{x}'', \boldsymbol{E}_{12}\boldsymbol{x}'') + (\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\boldsymbol{x}', \boldsymbol{E}_{12}^\top\boldsymbol{x}') \end{pmatrix} \begin{pmatrix} \lambda \\ \lambda' \\ \lambda'' \end{pmatrix} = \begin{pmatrix} (\boldsymbol{x}, \boldsymbol{E}_{01}\boldsymbol{x}') \\ (\boldsymbol{x}, \boldsymbol{E}_{02}\boldsymbol{x}'') \\ (\boldsymbol{x}', \boldsymbol{E}_{12}\boldsymbol{x}'') \end{pmatrix}. \tag{A.13}$$

We solve this for λ, λ', and λ'' and correct $\boldsymbol{x}$, $\boldsymbol{x}'$, and $\boldsymbol{x}''$ according to Eq. (A.12) as follows:

$$\begin{aligned} \hat{\boldsymbol{x}} &= \boldsymbol{x} - \lambda \boldsymbol{P}_k \boldsymbol{E}_{01} \boldsymbol{x}' - \lambda' \boldsymbol{P}_k \boldsymbol{E}_{02} \boldsymbol{x}'', \\ \hat{\boldsymbol{x}}' &= \boldsymbol{x}' - \lambda \boldsymbol{P}_k \boldsymbol{E}_{01}^\top \boldsymbol{x} - \lambda'' \boldsymbol{P}_k \boldsymbol{E}_{12} \boldsymbol{x}'', \\ \hat{\boldsymbol{x}}'' &= \boldsymbol{x}'' - \lambda' \boldsymbol{P}_k \boldsymbol{E}_{02}^\top \boldsymbol{x} - \lambda'' \boldsymbol{P}_k \boldsymbol{E}_{12}^\top \boldsymbol{x}'. \end{aligned} \tag{A.14}$$

However, this is only first approximation. So, we add higher order correction terms $\Delta\hat{\boldsymbol{x}}$, $\Delta\hat{\boldsymbol{x}}'$, and $\Delta\hat{\boldsymbol{x}}''$ in such a way that

$$\begin{aligned} &(\hat{\boldsymbol{x}} - \Delta\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}(\hat{\boldsymbol{x}}' - \Delta\hat{\boldsymbol{x}}')) = 0, \\ &(\hat{\boldsymbol{x}} - \Delta\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}(\hat{\boldsymbol{x}}'' - \Delta\hat{\boldsymbol{x}}'')) = 0, \\ &(\hat{\boldsymbol{x}}' - \Delta\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}(\hat{\boldsymbol{x}}'' - \Delta\hat{\boldsymbol{x}}'')) = 0. \end{aligned} \tag{A.15}$$

Expanding these with respect to $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}'$, and $\Delta\boldsymbol{x}''$ and ignoring higher order terms, we obtain

$$\begin{aligned} &(\Delta\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') + (\Delta\hat{\boldsymbol{x}}', \boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}}) = (\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}\hat{\boldsymbol{x}}'), \\ &(\Delta\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') + (\Delta\hat{\boldsymbol{x}}'', \boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}}) = (\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}\hat{\boldsymbol{x}}''), \\ &(\Delta\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') + (\Delta\hat{\boldsymbol{x}}'', \boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}') = (\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}''). \end{aligned} \tag{A.16}$$

Among those $\Delta\hat{\boldsymbol{x}}$, $\Delta\hat{\boldsymbol{x}}'$, and $\Delta\hat{\boldsymbol{x}}''$ that satisfy these, we choose the values that minimizes

$$\begin{aligned} &\|\boldsymbol{x} - (\hat{\boldsymbol{x}} - \Delta\hat{\boldsymbol{x}})\|^2 + \|\boldsymbol{x}' - (\hat{\boldsymbol{x}}' - \Delta\hat{\boldsymbol{x}}')\|^2 + \|\boldsymbol{x}' - (\hat{\boldsymbol{x}}'' - \Delta\hat{\boldsymbol{x}}'')\|^2 \\ &= \|\tilde{\boldsymbol{x}} + \Delta\hat{\boldsymbol{x}}\|^2 + \|\tilde{\boldsymbol{x}}' + \Delta\hat{\boldsymbol{x}}'\|^2 + \|\tilde{\boldsymbol{x}}'' + \Delta\hat{\boldsymbol{x}}''\|^2, \end{aligned} \tag{A.17}$$

where we put

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} - \hat{\boldsymbol{x}}, \qquad \tilde{\boldsymbol{x}}' = \boldsymbol{x}' - \hat{\boldsymbol{x}}', \qquad \tilde{\boldsymbol{x}}'' = \boldsymbol{x}'' - \hat{\boldsymbol{x}}''. \tag{A.18}$$

Since the correction occurs in the image plane, we have $(\boldsymbol{k}, \Delta\hat{\boldsymbol{x}}) = 0$, $(\boldsymbol{k}, \Delta\hat{\boldsymbol{x}}') = 0$, and $(\boldsymbol{k}, \Delta\hat{\boldsymbol{x}}'') = 0$. Introducing Lagrange multipliers, we differentiate

$$\begin{aligned} &\frac{1}{2}\Big(\|\tilde{\boldsymbol{x}} + \Delta\hat{\boldsymbol{x}}\|^2 + \|\tilde{\boldsymbol{x}}' + \Delta\hat{\boldsymbol{x}}'\|^2 + \|\tilde{\boldsymbol{x}}'' + \Delta\hat{\boldsymbol{x}}''\|^2\Big) \\ &-\lambda\Big((\Delta\hat{\boldsymbol{x}}, \boldsymbol{E}_{01}\hat{\boldsymbol{x}}') + (\Delta\hat{\boldsymbol{x}}', \boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}})\Big) \\ &-\lambda'\Big((\Delta\hat{\boldsymbol{x}}, \boldsymbol{E}_{02}\hat{\boldsymbol{x}}'') + (\Delta\hat{\boldsymbol{x}}'', \boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}})\Big) \\ &-\lambda''\Big((\Delta\hat{\boldsymbol{x}}', \boldsymbol{E}_{12}\hat{\boldsymbol{x}}'') + (\Delta\hat{\boldsymbol{x}}'', \boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}')\Big) \\ &-\mu(\boldsymbol{k}, \Delta\hat{\boldsymbol{x}}) - \mu'(\boldsymbol{k}, \Delta\hat{\boldsymbol{x}}') - \mu''(\boldsymbol{k}, \Delta\hat{\boldsymbol{x}}''), \end{aligned} \tag{A.19}$$

with respect to $\Delta\boldsymbol{x}$, $\Delta\boldsymbol{x}'$, and $\Delta\boldsymbol{x}''$ and let the result be 0. We obtain

$$\begin{aligned} \Delta\boldsymbol{x} &= \lambda\boldsymbol{E}_{01}\hat{\boldsymbol{x}}' + \lambda'\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'' + \mu\boldsymbol{k} - \tilde{\boldsymbol{x}}, \\ \Delta\boldsymbol{x}' &= \lambda\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}} + \lambda''\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'' + \mu'\boldsymbol{k} - \tilde{\boldsymbol{x}}', \\ \Delta\boldsymbol{x}'' &= \lambda'\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}} + \lambda''\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}' + \mu''\boldsymbol{k} - \tilde{\boldsymbol{x}}''. \end{aligned} \tag{A.20}$$

Multiplying these by $\boldsymbol{P}_k$ from left on both sides and noting that $\boldsymbol{P}_k\Delta\boldsymbol{x} = \Delta\boldsymbol{x}$, $\boldsymbol{P}_k\Delta\boldsymbol{x}' = \Delta\boldsymbol{x}'$, $\boldsymbol{P}_k\Delta\boldsymbol{x}'' = \Delta\boldsymbol{x}''$, $\boldsymbol{P}_k\boldsymbol{k} = \boldsymbol{0}$, $\boldsymbol{P}_k\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}$, $\boldsymbol{P}_k\tilde{\boldsymbol{x}}' = \tilde{\boldsymbol{x}}'$, and $\boldsymbol{P}_k\tilde{\boldsymbol{x}}'' = \tilde{\boldsymbol{x}}''$, we obtain

$$\begin{aligned} \Delta\boldsymbol{x} &= \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}\hat{\boldsymbol{x}}' + \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'' - \tilde{\boldsymbol{x}}, \\ \Delta\boldsymbol{x}' &= \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}} + \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'' - \tilde{\boldsymbol{x}}', \\ \Delta\boldsymbol{x}'' &= \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}} + \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}' - \tilde{\boldsymbol{x}}''. \end{aligned} \tag{A.21}$$

Substituting these into Eq. (A.16), we obtain the linear equation in Eq. (22). Solving it for λ, λ', and λ'', we correct $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}'$, and $\hat{\boldsymbol{x}}''$ according to Eq. (A.21) as follows:

$$\begin{aligned} \hat{\hat{\boldsymbol{x}}} &= \hat{\boldsymbol{x}} - \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}\hat{\boldsymbol{x}}' - \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'' + \tilde{\boldsymbol{x}} \\ &= \boldsymbol{x} - \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}\hat{\boldsymbol{x}}' - \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}\hat{\boldsymbol{x}}'', \\ \hat{\hat{\boldsymbol{x}}}' &= \hat{\boldsymbol{x}}' - \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}} - \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'' + \tilde{\boldsymbol{x}}' \\ &= \boldsymbol{x}' - \lambda\boldsymbol{P}_k\boldsymbol{E}_{01}^\top\hat{\boldsymbol{x}} - \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}\hat{\boldsymbol{x}}'', \\ \hat{\hat{\boldsymbol{x}}}'' &= \hat{\boldsymbol{x}}'' - \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}} - \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}' + \tilde{\boldsymbol{x}}'' \\ &= \boldsymbol{x}'' - \lambda'\boldsymbol{P}_k\boldsymbol{E}_{02}^\top\hat{\boldsymbol{x}} - \lambda''\boldsymbol{P}_k\boldsymbol{E}_{12}^\top\hat{\boldsymbol{x}}'. \end{aligned} \tag{A.22}$$

Regarding these as new $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{x}}'$, and $\hat{\boldsymbol{x}}''$, we add higher order corrections again and repeat this until the reprojection error E no longer decreases. The procedure is arranged in the form given in Section 8.

Yasushi Kanazawa received his B.E. and M.S. degree in information engineering from Toyohashi University of Technology in 1985 and 1987, respectively and his Ph.D. in information and computer science from Osaka University in 1997. After engaging in research and development of image processing systems at Fuji Electric Co., Tokyo, Japan, and serving as Lecturer of Information and Computer Engineering at Gunma College of Technology, Gunma, Japan, he is currently Associate Professor of computer science and engineering at Toyohashi University of Technology, Aichi, Japan. His research interests include image processing and computer vision.

Yasuyuki Sugaya received his B.E., M.S., and Ph.D. in computer science from the University of Tsukuba, Ibaraki, Japan, in 1996, 1998, and 2001, respectively. After serving as Assistant Professor of computer science at Okayama University, Okayama, Japan, he is currently Associate Professor of computer science and engineering at Toyohashi University of Technology, Toyohashi, Aichi, Japan. His research interests include image processing and computer vision. He received the IEICE Best Paper Award in 2005.

Kenichi Kanatani received his B.E., M.S., and Ph.D. in applied mathematics from the University of Tokyo in 1972, 1974 and 1979, respectively. After serving as Professor of computer science at Gunma University, Gunma, Japan, and Okayama University, Okayama, Japan, he retired in 2013 and is now Professor Emeritus of Okayama University. He is the author of many books on computer vision and received many awards including the Best Paper Awards from IPSJ (1987) and IEICE (2005). He is a Fellow of IEICE and IEEE.

(Communicated by *Anton van den Hengel*)