5-2

Statistical Optimization for 3-D Reconstruction from a Single View

Kenichi Kanatani* Yasuyuki Sugaya*

∗ Department of Computer Science, Okayama University

Abstract

We analyze the noise sensitivity of the focal length computation for single-view 3-D reconstruction based on vanishing points and orthogonality. We point out that due to the nonlinearity of the computation the standard statistical optimization is not very effective. We present a practical compromise between preventing the computational failure and maintaining the accuracy and demonstrate that our method can produce a consistent 3-D shape in the presence of however large noise.

1. Introduction

Given two or more views of a scene, we can reconstruct its 3-D structure based on triangulation [2, 3]. However, the 3-D shape can be reconstructed from only a single view if we have sufficient knowledge about the scene [1, 3]. For example, if there are parallel lines in the scene, their projections define their vanishing point, which constrains the orientation of these lines in the scene. If we can find three vanishing points of mutually orthogonal parallel lines in the scene, we can compute the camera focal length and the principal point, from which we can compute the positions and orientations of the lines in the scene.

This type of single-view 3-D reconstruction has been studied in various forms and is widely used today not only in industrial environments such as robotic manufacturing and navigation but also in many other fields including entertainment, education, and scholastic research through virtual reality generation and 3-D reconstruction from paintings and historical photographs.

The major disadvantage of such single-view reconstruction is that because it is based on the perspective projection geometry, according to which objects further away look smaller, we cannot reconstruct 3-D if there are no perspective effects. Even if there are, the computation often fails when the perspective effects are small, which is often the case for a distant scene. A typical symptom is that the inside of the square root becomes negative when we estimate the camera focal length by the standard method, resulting in an imaginary focal length.

This paper studies the effects of the noise on the focal length computation. We point out that due to the nonlinearity of the computation the standard statistical optimization is not very effective. We present a practical compromise between preventing the computational failure and maintaining the accuracy and demonstrate that our method can produce a consistent 3-D shape in the presence of however large noise.

2. Camera Model

We define an XYZ coordinate system with the origin O at the center of the camera lens (the *viewpoint*) and the Z-axis along the lens optical axis and regard the camera imaging geometry as *perspective projection*: A point in the scene is projected onto the intersection of the plane $Z = f$ (the *image plane*) with the line (the *line of sight*) starting from the view point O and passing through that point (Fig. 1). The constant f is called the *focal length*.

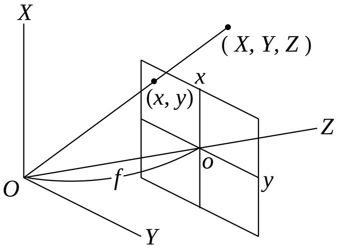


Figure 1: Perspective projection.

The input image is identified with the plane $Z = f$, on which we define an xy coordinate system with the image origin at the point on the Z-axes (the *principal point*) and the x- and y-axes parallel to the camera X- and Y-axes, respectively. The image coordinate system is assume to have zero skew with aspect ratio 1. For the time being, the principal point is assumed to be known, typically at the center of the image frame (we later consider its estimation).

We represent an image point (x, y) by the following 3-D vectors:

$$\boldsymbol{x} = \begin{pmatrix} x/f_0 \\ y/f_0 \\ 1 \end{pmatrix}, \quad \boldsymbol{m} = \frac{1}{\sqrt{x^2+y^2+f_0^2}} \begin{pmatrix} x \\ y \\ f_0 \end{pmatrix}. \qquad (1)$$

Here, f_0 is a default focal length[1] measured in pixels. The two vectors $\boldsymbol{x}$ and $\boldsymbol{m}$ are transformed to each other as follows:

$$\boldsymbol{x} = Z[\boldsymbol{m}], \qquad \boldsymbol{m} = N[\boldsymbol{x}]. \qquad (2)$$

Throughout this paper, $Z[\,\cdot\,]$ normalization to make the third component 1, and $N[\,\cdot\,]$ means normalization into a unit vector. We call $\boldsymbol{x}$ and $\boldsymbol{m}$ their *Z-vector* and *N-vector*, respectively [3].

A line in the image is written as $ax+by+c = 0$. Since the coefficients a, b, and c can be specified only up to multiplication by a nonzero constant, we can normalize them to $a^2 + b^2 + (c/f_0)^2 = 1$. We call the unit vector

$$\boldsymbol{n} = N[\begin{pmatrix} a \\ b \\ c/f_0 \end{pmatrix}] \qquad (3)$$

the *N-vector* of the line [3]. Using the vector notation of eqs. (1), we can write the equation of the line as

*Okayama-shi, Okayama 700-8530 Japan
{kanatani,sugaya}@suri.it.okayama-u.ac.jp

[1] In theory, its value is arbitrary. In our experiment, we used the value $f_0 = 600$ (pixels).

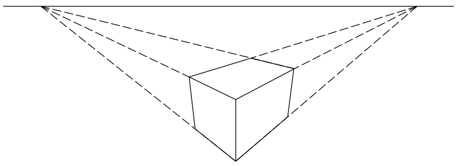


Figure 2: Vanishing points.

$(\boldsymbol{n}, \boldsymbol{x}) = 0$. Throughout this paper, we denote the inner product of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$.

The N-vector $\boldsymbol{n}$ of the line passing through two points with Z-vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ and the Z-vector $\boldsymbol{x}$ of the intersection of two lines with N-vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are given as follows [3]:

$$\boldsymbol{n} = N[\boldsymbol{x}_1 \times \boldsymbol{x}_2], \qquad \boldsymbol{x} = Z[\boldsymbol{n}_1 \times \boldsymbol{n}_2]. \tag{4}$$

3. Focal Length Estimation

The first step of 3-D reconstruction is to compute the *vanishing point* of parallel lines in the scene: it is defined as the intersection of their projections on the image (Fig. 2). The orientation of the lines in the scene coincides with the direction toward the vanishing point on the image plane $Z = f$ seen from the viewpoint O [2, 3].

In the presence of noise, the projections of parallel lines do not necessarily intersect at a single point in the image. An optimal procedure for estimating the true intersection, called *renormalization*, was presented by Kanazawa and Kanatani [5]. This procedure produces not only an optimal estimate of the vanishing point but also the *covariance matrix*[2] $V[\boldsymbol{m}]$ of the N-vector of the estimated vanishing point as a byproduct.

Suppose the vanishing points of three mutually orthogonal parallel lines are located in the image. Let $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ be their N-vectors, and $V[\boldsymbol{m}_1]$, $V[\boldsymbol{m}_2]$, and $V[\boldsymbol{m}_3]$ be their covariance matrices. The unit vectors $\hat{\boldsymbol{m}}_1$, $\hat{\boldsymbol{m}}_2$, and $\hat{\boldsymbol{m}}_3$ that start from the viewpoint O and point toward these vanishing points are given by

$$\hat{\boldsymbol{m}}_i = N[\boldsymbol{I}_f \boldsymbol{m}_i], \qquad i = 1, 2, 3, \tag{5}$$

where

$$\boldsymbol{I}_f \equiv \mathrm{diag}(1, 1, \frac{f}{f_0}). \tag{6}$$

The symbol diag($\cdots$) denotes the diagonal matrix with $\cdots$ as the diagonal elements in that order. Since the three vanishing point orientations should be mutually orthogonal, we have the following constraints:

$$e_1 \equiv (\hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3) = 0, \qquad e_2 \equiv (\hat{\boldsymbol{m}}_3, \hat{\boldsymbol{m}}_1) = 0,$$
$$e_3 \equiv (\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2) = 0. \tag{7}$$

In the presence of noise, these do not necessarily hold exactly. An optimal estimate for the focal length f is obtained by minimizing the following function [4]:

$$J = \sum_{i,j=1}^{3} W_{ij} e_i e_j. \tag{8}$$

Here, we define the matrix $\boldsymbol{W} = (W_{ij})$ as the inverse of the matrix $\boldsymbol{V} = (V_{ij})$ having the covariance V_{ij} of e_i and e_j as the ij element (V_{ii} is the variance of e_i). Using eq. (5), we obtain

$$\begin{aligned}
V_{11} &= (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_3) + (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_2), \\
V_{22} &= (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_1) + (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_3), \\
V_{33} &= (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_2) + (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_1), \\
V_{23} &= V_{32} = (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_3), \\
V_{31} &= V_{13} = (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_1), \\
V_{12} &= V_{21} = (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_2).
\end{aligned} \tag{9}$$

The matrix $\boldsymbol{W} = (W_{ij})$ weights the three constraints of eqs. (7) according to the reliability of the three vanishing points evaluated in terms of their covariance matrices $V[\boldsymbol{m}_i]$. If the uncertainty of each vanishing point were the same and independent of each other, the matrix $\boldsymbol{W}$ would be a constant times the unit matrix $\boldsymbol{I}$, so the minimization of eq. (8) would reduce to the *least-squares method* that minimizes $\sum_{i=1}^{3} e_i^2$.

Eq. (8) can be minimized by first regarding $\boldsymbol{W}$ as a constant matrix. Then, eq. (8) is a quadratic polynomial in

$$\alpha = \left(\frac{f}{f_0}\right)^2, \tag{10}$$

so the value α that minimizes eq. (8) can be analytically obtained. We update $\boldsymbol{W}$ by substituting this value, recompute α, and iterate this until it converges. The focal length f is given by

$$f = f_0 \sqrt{\alpha}. \tag{11}$$

4. Composite Method

The above method sounds satisfactory, because it is theoretically guaranteed to be optimal. However, the optimality is based on linear analysis: the covariance matrix $V[\boldsymbol{m}]$ is defined as the expectation $E[\Delta\boldsymbol{m}\Delta\boldsymbol{m}^\top]$ for the deviation $\Delta\boldsymbol{m}$ of $\boldsymbol{m}$ evaluated to a first approximation via Taylor expansion [4].

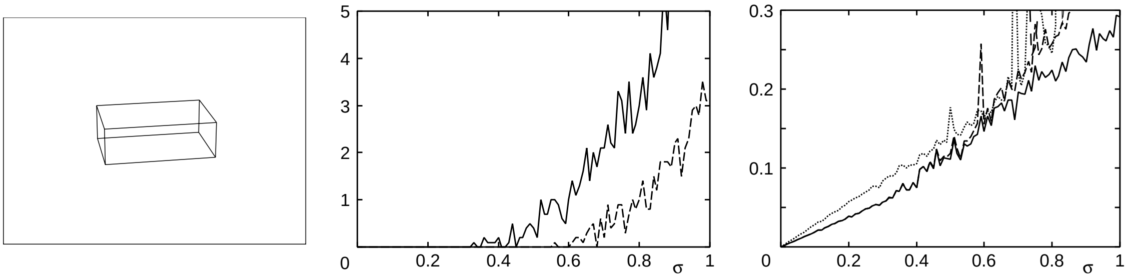
In reality, the vanishing point computation is highly nonlinear, particularly so when it is located very far a way from the frame center, resulting in far larger deviations than predicted by the covariance analysis. A typical symptom is that the value α computed by eq. (8) becomes negative, resulting in an imaginary focal length f.

The reason why a real solution does not exist while geometrically there should be one is that some of the necessary geometric constraints are violated. In fact, the three vanishing points cannot be anywhere but should be at the vertices of a triangle whose orthocenter is at the principal point [2, 3], implying that the directions toward any two of them from the principal point make an obtuse angle.

However, the vanishing point locations can be perturbed to a large extent even by slight noise due to the nonlinearity, so these conditions may be violated. For such an inadmissible configuration, a real focal length solution may no longer exist.

From these considerations, we take a strategy of complementing the insufficiency of linear analysis by checking to what extent the necessary geometric conditions are violated. We consider the following four cases for the angles made by the directions toward the computed vanishing points from the image origin:

[2] Since multiplication of the covariance matrix by a common factor does not affect the results in this paper, we compute it with an appropriate scale normalization [4].



(a) (b) (c)

Figure 3: (a) Simulated image of a box. (b) The percentage (%) of computational failure. Solid line: optimal computation. Dotted line: least squares. (c) Accuracy of focal length computation. Solid line: composite method. Dashed line: optimal computation. Dotted line: least squares.

Case 1: Three obtuse angles. We regard the three vanishing points as sufficiently reliable and do the optimal computation as described in Sec. 3.

Case 2: One acute angle. We remove from among the three constraints in eqs. (7) the one involving the two directions that make an acute angle and minimize eq. (8) subject to the remaining two constraints.

Case 3: Two acute angles. We retain from among the three constraints in eqs. (7) only the one involving the two directions that make an obtuse angle. In this case, we need not minimize eq. (8): the solution that makes eq. (8) zero can be analytically obtained.

Case 4: Three acute angles. We regard no vanishing points as reliable and formally returns $f = \infty$ (a sufficiently large value in practice).

Fig. 3(a) is a simulated image of a rectangular box. The image size is supposed to be 300×400 (pixels), and the focal length is set to $f = 1000$ (pixels). We added Gaussian noise of mean 0 and standard deviation σ (pixels) to the x and y coordinates of all the vertices positions independently and estimated the focal length.

Fig. 3(b) plots the percentage (%) of computational failures (resulting in an imaginary focal length or no convergence[3]) of the optimal computation of Sec. 3. (solid line) over 1000 trials using different noise for each σ on the horizontal axis. The dotted line shows, for comparison, the corresponding result for the least squares (eq. (8) is replaced by $\sum_{i=1}^{3} e_i^2$). The composite computation is not plotted here, because it does not fail.

We can see that the rate of computational failures increases as the noise increases. It is larger for the optimal computation than for the least squares, indicating that *pursuit for high accuracy is generally not compatible with computational robustness*.

We then evaluated the relative accuracy of the composite method by the following root-mean-square:

$$D = \sqrt{\frac{1}{1000}\sum_{a=1}^{1000}\Big(\frac{f^{(a)} - f}{f^{(a)}}\Big)^2}. \qquad (12)$$

Here, $f^{(a)}$ is the ath instance of the 1000 trials. We let $f^{(a)} = \infty$ (i.e., $(f^{(a)} - f)/f^{(a)} = 1$) if the computation failed. Fig. 3(c) plots the value D for the noise standard deviation σ on the horizontal axis. The solid line is for the composite method; the dashed line is for the optimal computation; the dotted line is for the least squares.

We can immediately see that the least-squares method, which ignores the statistical error behavior, yields poor results. The optimal computation indeed achieves high accuracy when the noise level is small, but the error irregularly fluctuates for a large noise level. The composite method retains high accuracy for small noise, yet preserves the accuracy expected of the optimal computation even for large noise.

5. Image Data Correction

So far, we have assumed that the principal point is known and taken to be the image origin. It can be computed as the orthocenter of the triangle defined by the vanishing points of three mutually orthogonal set of parallel lines in the scene [2, 3]. According to our simulation experiments, however, the principal point is very sensitive to noise; it is perturbed to an extraordinary degree by very small noise. We conclude that estimating the principal point is not realistic unless the vanishing points can be estimated with very high accuracy. Rather, it is more reasonable to assume an appropriate principal point instead of estimating it. The possible distortions resulting from the use of a wrong principal point can be compensated for by correcting the *image itself* so that it conforms to the estimated parameters. We now describe the procedure.

The three directions from the viewpoint to the vanishing points may not be exactly orthogonal even though the focal length is optimally computed. So, we correct them to be exactly orthogonal. First, we convert the N-vectors $\{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$ of the three vanishing points into $\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$ for the computed focal length f, using eqs. (5) and (6). A statistically optimal method for computing an orthonormal system $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ that best approximates a given set of vectors $\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$ is to minimize

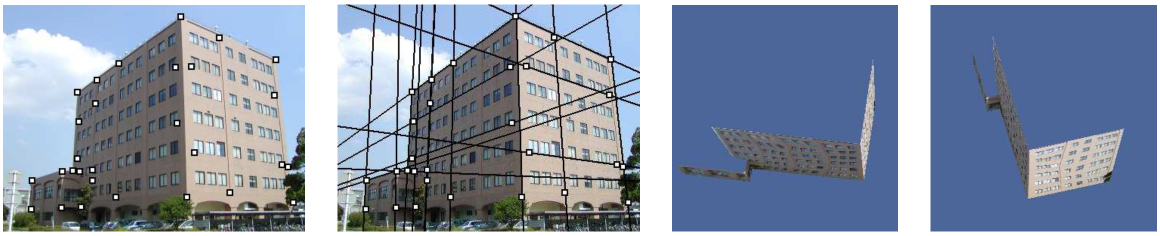
$$\frac{\|\boldsymbol{e}_1 - \hat{\boldsymbol{m}}_1\|^2}{\mathrm{tr}V[\boldsymbol{m}_1]} + \frac{\|\boldsymbol{e}_2 - \hat{\boldsymbol{m}}_2\|^2}{\mathrm{tr}V[\boldsymbol{m}_2]} + \frac{\|\boldsymbol{e}_3 - \hat{\boldsymbol{m}}_3\|^2}{\mathrm{tr}V[\boldsymbol{m}_3]}, \qquad (13)$$

where $V[\boldsymbol{m}_i]$ is the covariance matrix of the ith vanishing point computed by renormalization, and tr denotes the trace[4]. The solution is obtained by computing singular value decomposition (SVD) of the matrix that has $\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$ as its columns:

$$\begin{pmatrix} \frac{\hat{\boldsymbol{m}}_1}{\mathrm{tr}V[\boldsymbol{m}_3]} & \frac{\hat{\boldsymbol{m}}_2}{\mathrm{tr}V[\boldsymbol{m}_3]} & \frac{\hat{\boldsymbol{m}}_3}{\mathrm{tr}V[\boldsymbol{m}_3]} \end{pmatrix} = \boldsymbol{V}\mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)\boldsymbol{U}^\top. \qquad (14)$$

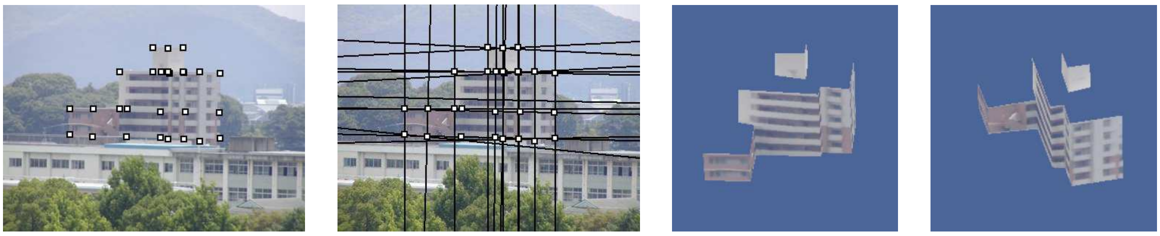
[3] We regarded the iterations as convergent when the increment in f is less than 1 pixel and as divergent when the number of iterations exceeds 10.

[4] By the definition of the covariance matrix $V[\boldsymbol{m}_i]$, the trace $\mathrm{tr}V[\boldsymbol{m}_i]$ is the mean square $E[\|\Delta\boldsymbol{m}_i\|^2]$.



(a) (b) (c) (d)

Figure 4: Input image (short-range view) and its 3-D reconstruction.



(a) (b) (c) (d)

Figure 5: Input image (distant view) and its 3-D reconstruction.

Here, $\boldsymbol{V}$ and $\boldsymbol{U}$ are orthogonal matrix, and σ_1, σ_2, and σ_3 the singular values. The solution $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is given as follows [3, 4]:

$$\begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{pmatrix} = \boldsymbol{V}\boldsymbol{U}^\top. \tag{15}$$

If we modify the vanishing point locations, the projections of parallel lines no longer pass through them in the image. So, we correct all the lines so that they pass through their respective vanishing points. Let $\boldsymbol{n}$ be the N-vector of the line in question, and $V[\boldsymbol{n}]$ its covariance matrix. Let $\bar{\boldsymbol{m}}_i$ be the N-vector of the corrected vanishing point. The optimal correction $\Delta\boldsymbol{n}$ of $\boldsymbol{n}$ is is given as follows [4]:

$$\bar{\boldsymbol{n}} = N\Big[\boldsymbol{n} - \frac{(\boldsymbol{n}, \boldsymbol{m}_i)}{(\boldsymbol{m}_i, V[\boldsymbol{n}]\boldsymbol{m}_i)} V[\boldsymbol{n}]\boldsymbol{m}_i\Big]. \tag{16}$$

If the lines are corrected in this way, the Z-vectors of their intersections are replaced by the second of eqs. (4). Points on these lines but not at the intersections with other lines are orthogonally displaced onto the nearest position on the corrected lines. The N-vectors of the lines passing through displaced points are replaced by the first of eqs. (4), the Z-vectors of the intersections of the displaced lines are replaced by the second of eq. (4), and this process is propagated.

Using the constraints on the orientations of lines and planes provided by the corrected vanishing points and vanishing lines, we can determine the 3-D shape of the scene up to a scale factor [1, 3].

6. Real Image Examples

Fig. 4(a) is a short-range view of a building with a strong perspective effect (300 × 400 pixels). The selected feature points are marked in it. Using the three orthogonal directions drawn in Fig. 4(b), we estimated the focal length to be 416 pixels by least squares and 431 pixels by the optimal computation. In this case, the three angles defined by the vanishing points are all obtuse, so the composite method reduces to the optimal computation. Fig. 4(c),(d) are views of the reconstructed 3-D shape seen from two different angles.

Fig. 5(a) is a distant view of a building with little perspective effect (300 × 400 pixels); the projection is almost orthographic. Using the feature points marked there and the three orthogonal directions drawn Fig. 5(b), we estimated the focal length to be 812 pixels by least squares and 2825 pixels by the optimal computation. In this case, only one of the three angles defined by the vanishing points is obtuse, and the composite method yields 3103 pixels. Applying our correction procedures, we obtain a consistent 3-D shape, as displayed in the two right images in Fig. 5(c),(d): lines that should be parallel are exactly parallel, and lines that should be orthogonal are exactly orthogonal.

7. Concluding Remarks

We analyzed the noise sensitivity of the focal length computation for single-view 3-D reconstruction. We pointed out that due to the nonlinearity of the computation the standard statistical optimization is not very effective. We presented a practical compromise between preventing the computational failure and maintaining the accuracy and demonstrated that our method can produce a consistent 3-D shape in the presence of however large noise.

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. 15500113). The authors thank Richard Hartley of the Australian National University and Sabry El-Hakim of NRC, Canada, for helpful discussions. They also thank Naoki Ikeda of Okayama University for helping the synthetic and real image experiments.

References

[1] A. Criminisi, I. Reid and A. Zisserman, Single view metrology, *Proc. 7th Int. Conf. Comput. Vision*, Kerkyra, Greece, Vol. 1, pp. 434–441, Sept. 1999.

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

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

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

[5] Y. Kanazawa and K. Kanatani, Optimal line fitting and reliability evaluation, *IEICE Trans. Inf. & Syst.*, **E79-D**-9 (1996), 1317–1322.