Proc. 10th Int. Conf. Virtual Sys. MultiMedia (VSMM2004), Nov. 2004, Ogaki, Japan, pp. 554–563.

Statistical Optimization for 3-D Reconstruction from a Single View

Kenichi Kanatani and Yasuyuki Sugaya

Department of Information Technology, Okayama University, Okayama, Japan

Abstract. We analyze the noise sensitivity of the focal length computation, the principle point estimation, and the orthogonality enforcement for single-view 3-D reconstruction based on vanishing points and orthogonality. We first point out that due to the nonlinearity of the computation the standard optimal computation is not actually optimal. We then 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.

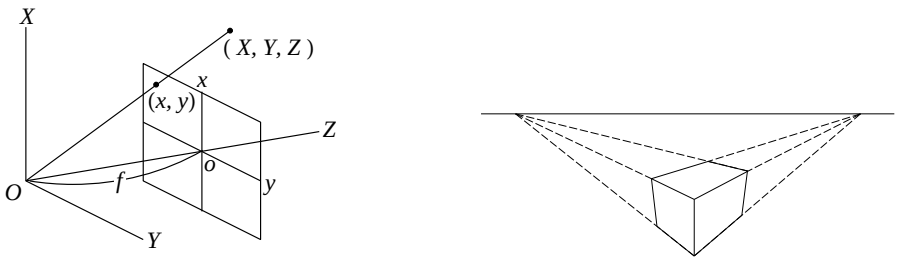
This type of single-view 3-D reconstruction has been studied in various forms and is widely used today not only in industrial environments such as robotic manufacturing and navigation but also in many other fields including entertainment, education, and scholastic research through virtual reality generation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

Most research on single-view reconstruction in the past seems to have focused on computational procedures and novel applications; little consideration seems to have been made on exact analysis of the accuracy and stability of computation. The purpose of this paper is to study the effects of the noise on the focal length computation, the principal point estimation, and the orthogonality correction.

We assume that the "noise", by which we mean *the inaccuracy of feature point locations*, is small. So, one may be tempted to use the Gaussian noise model with mean 0 and a small variance, since in this situation an effective optimization technique is well known [4]: we evaluate the covariance of the final output by computing the intermediate Jacobian matrices via Taylor expansion and cascading them using the chain rule and devise a method that minimizes that covariance.

In the following, we first point out that such an optimization technique does not necessarily produce optimal results, because the nonlinearity of the computation grows



(a) (b)

Figure 1: (a) Perspective projection. (b) Vanishing points and the vanishing line.

as the scene is further away to such an extent that even the slightest noise causes some quantities to diverge. Considering this nonlinearity, we then present a new technique that does not fail for however large noise but still maintains the accuracy. We demonstrate that our method can produce a consistent 3-D shape in the presence of however large noise.

2. Camera Model

We first describe our camera model and define our notations and terminologies [3]. Defining an XYZ coordinate system with the origin O at the center of the camera lens (called the *viewpoint*) and the Z-axis along the lens optical axis, we regard the camera imaging geometry as *perspective projection*: A point in the scene is projected onto the intersection of the plane $Z = f$ (called the *image plane*) with the line (called the *line of sight*) starting from the view point O and passing through that point (Fig. 1(a)). The constant f is called the *focal length*.

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 (called 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:

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

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:

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

Throughout this paper, $N[\,\cdot\,]$ means normalization into a unit vector, and $Z[\,\cdot\,]$ normalization to make the third component 1. 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 normalize them to $a^2 + b^2 + (c/f_0)^2 = 1$. We call the unit vector

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

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

the *N-vector* of the line [3]. Using the vector notation of Eqs. (1), we can write the equation of the line as $(\boldsymbol{n}, \boldsymbol{x}) = 0$. 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. 1(b)). 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].

If we observe non-parallel coplanar lines in the scene, their vanishing points are collinear in the image, defining the *vanishing line* (Fig. 1(b)). The orientation of the supporting plane coincides with that of the plane passing through the viewpoint O and intersecting the image plane $Z = f$ along the vanishing line [2, 3].

In the presence of noise, however, 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 *normalized covariance matrix*[2] $V_0[\boldsymbol{m}]$ 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 $V_0[\boldsymbol{m}_1]$, $V_0[\boldsymbol{m}_2]$, and $V_0[\boldsymbol{m}_3]$ be their normalized covariance matrices (computed by the renormalization procedure). 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 $\mathrm{diag}(\cdots)$ 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:

$$e_1 \equiv (\hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3) = (\boldsymbol{m}_2, \boldsymbol{I}_f^2 \boldsymbol{m}_3) = 0, \qquad e_2 \equiv (\hat{\boldsymbol{m}}_3, \hat{\boldsymbol{m}}_1) = (\boldsymbol{m}_3, \boldsymbol{I}_f^2 \boldsymbol{m}_1) = 0,$$

$$e_3 \equiv (\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2) = (\boldsymbol{m}_1, \boldsymbol{I}_f^2 \boldsymbol{m}_2) = 0. \tag{7}$$

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

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

[2]The covariance matrix of the N-vector $\boldsymbol{m}$ is $\sigma^2 V_0[\boldsymbol{n}]$ if the standard deviation of the inaccuracy of feature point location is σ (pixels). Since multiplication by a positive constant does not affect the subsequent computation, we normalize σ^2 to 1, hence the name "normalized" covariance matrix.

Here, we define the matrix $\boldsymbol{W} = (W_{ij})$ by

$$\boldsymbol{W} = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix}^{-1}, \tag{9}$$

where V_{ij} is the covariance of e_i and e_j (V_{ii} is the variance ofe_i). Using Eq. (5), we obtain

$$\begin{aligned} V_{11} &= (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_3) + (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_2), \\ V_{22} &= (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_1) + (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_3), \\ V_{33} &= (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_2) + (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_1), \\ V_{23} = V_{32} &= (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_3), \qquad V_{31} = V_{13} = (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_1), \\ V_{12} &= V_{21} = (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_2). \end{aligned} \tag{10}$$

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 normalized covariance matrices $V_0[\boldsymbol{m}_i]$. If the uncertainty of each vanishing point were the same and independent of each other, the matrix $\boldsymbol{W}$ would be a multiple of the unit matrix $\boldsymbol{I}$ by a constant, so the minimization of Eq. (8) would reduce to the *least-squares method* that minimizes $\sum_{i=1}^{3} e_i^2$.

Equation (8) can be minimized as follows. Although the matrix $\boldsymbol{W}$ depends on f through the matrix $\boldsymbol{I}_f$, the dependence is very small if the default value f_0 is chosen so that $f/f_0 \sim 1$. So, we tentatively regard $\boldsymbol{W}$ approximately as a constant matrix by substituting $f = f_0$ into $\boldsymbol{W}$. Then, Eq. (8) is a quadratic polynomial in

$$\alpha = \Big(\frac{f}{f_0}\Big)^2, \tag{11}$$

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

$$f = f_0 \sqrt{\alpha}. \tag{12}$$

4. Composite Method for Focal Length Estimation

The above method sounds satisfactory, because it is theoretically guaranteed to be optimal. However, the optimality is based on linear analysis. In fact, the normalized covariance matrix $V_0[\boldsymbol{n}]$ is defined as the expectation $E[\Delta\boldsymbol{n}\Delta\boldsymbol{n}^\top]$ for the deviation $\Delta\boldsymbol{n}$ of $\boldsymbol{n}$ evaluated to a first approximation via Taylor expansion for small noise [4]. This type of first approximation assumes that if the noise is zero-mean Gaussian, the errors in the computed vanishing point are also zero-mean Gaussian, having an elliptic equiprobability contours around it.

In reality, however, the vanishing point computation is highly nonlinear, particularly so when it is located very far a way, so even if the noise in the feature points is zero-mean, the errors in the vanishing point may have large bias, and the equiprobability contours may be parabolic and divergent toward infinity.

In such a case, the covariance analysis based on first approximation is unable to capture the error behavior. Hence, the guarantee that the minimizer of Eq. (8) is

optimal no longer holds. 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 to complement the insufficiency of linear analysis by checking to what extent the necessary geometric conditions are violated. We also consider the possibility that the vanishing point defined by nearly parallel lines may appear in the opposite direction. With this in mind, we consider the following three cases for the angles made by the directions toward the computed vanishing points from the image origin:

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, removing the others as unreliable (regarding one vanishing point direction as reversed). In this case, we need not minimize Eq. (8): the solution that makes Eq. (8) (quadratic in α) 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).

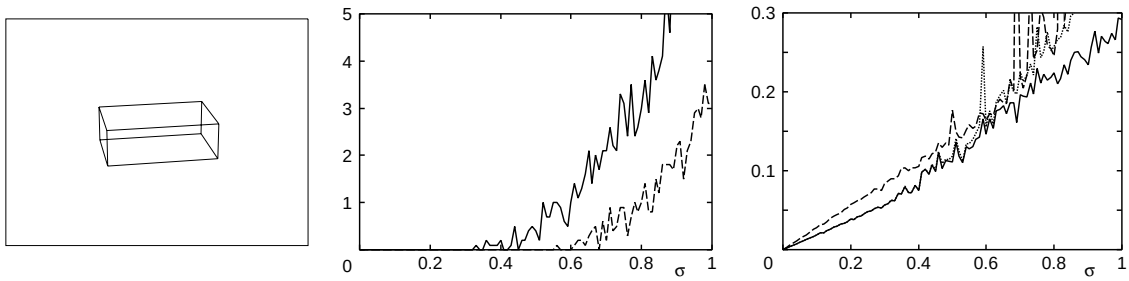
Fig. 2(a) is a simulated image of a rectangular box. The image size is supposed to be 300×400 (pixels), and the focal length is set to $f = 1000$ (pixels). We added Gaussian noise of mean 0 and standard deviation σ (pixels) to the x and y coordinates of all the vertices positions independently and estimated the focal length. Fig. 2(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 method (Eq. (8) is replaced by $\sum_{i=1}^{3} e_i^2$). 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 method. The composite computation is not plotted here, because it does not fail.

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

$$D = \sqrt{\frac{1}{1000} \sum_{a=1}^{1000} \Big(\frac{f^{(a)} - f}{f^{(a)}} \Big)^2}. \tag{13}$$

Here, $f^{(a)}$ is the ath instance of the 1000 trials. We let $f^{(a)} = \infty$ (i.e., $(f^{(a)} - f)/f^{(a)} = 1 - f/f^{(a)} = 1$) if the computation failed. Fig. 2(c) plots the value D for the noise

[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



(a) (b) (c)

Figure 2: (a) Simulated image of a box. (b) The percentage (%) of computational failure. Solid line: optimal computation. Dotted line: least-squares method. (c) Accuracy of focal length computation. Solid line: composite method. Dashed line: optimal computation. Dotted line: least-squares method.

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 method.

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

5. Principal Point Estimation

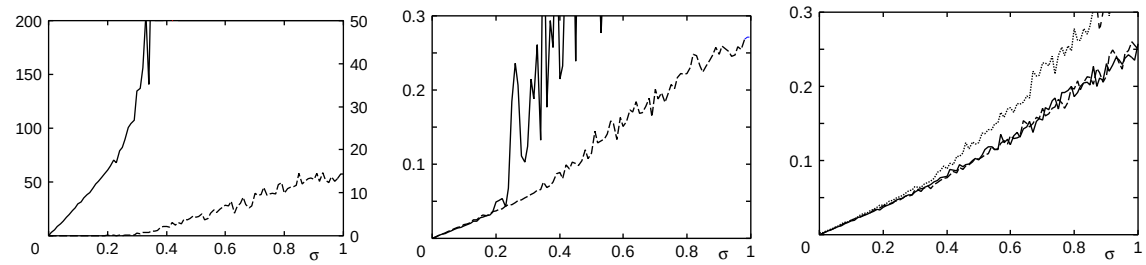
So far, we have assumed that the principal point is known and taken to be the image origin. Here, we consider the case where it is unknown. As mentioned earlier, it should be at the orthocenter of the triangle defined by the vanishing points of three mutually orthogonal set of parallel lines in the scene [2, 3]. However, the computed orthocenter can be the principal point only if it is inside the triangle with the vanishing points as its vertices, and even small noise in the image can shift the principal point out of the triangle.

Using the simulated image of Fig. 2(a), we estimated the principal point after adding Gaussian noise to the vertices. Since the true principal point is at the image origin, we evaluate the accuracy of the computed principal point by its root-mean-square distance from the image origin over 1000 trials

$$E = \sqrt{\frac{1}{1000}\sum_{a=1}^{1000}\Big((x_c^{(a)})^2 + (y_c^{(a)})^2\Big)}, \tag{14}$$

where $(x_c^{(a)}, y_c^{(a)})$ is the ath estimates The solid line (left scale) in Fig. 3(a) plots E in pixels vs. σ on the horizontal axis. The dashed line (right scale) shows the percentage (%) of the computed principal point being outside the triangle of the vanishing points.

As we can see, the principal point is very sensitive to noise; it is perturbed to an extraordinary degree by very small noise. This implies that estimating the principal point is not realistic unless the vanishing points can be estimated with very high accuracy. If the principal point is known to be somewhere near the center of the image, we may stably obtain better results using its approximate position. We investigated this by experiments.



(a) (b) (c)

Figure 3: (a) Accuracy of principal point estimation. Solid line (left scale): root-mean-square error (pixels). Dashed line (right scale): percentage (%) of being outside the triangle of the vanishing points. (b) Accuracy of focal length estimation. Solid line: estimating the principal point. Dashed line: using the true principal point. (c) Accuracy of focal length computation using a default principal point. The true principal point is displaced from it by 0 pixels (solid line), 50 pixels (dashed line), and 100 pixels (dotted line).

Using the simulated image of Fig. 2(a) with noise, we computed the focal length by the composite method described in Sec. 4 after shifting the image origin to the estimated principal point. The solid line in Fig. 3(b) shows the accuracy of the computed focal length measured in D in Eq. (13). The dashed line is for using the image origin (the true principal point). We can see that the accuracy significantly deteriorates if principal point estimation is incorporated.

For comparison, Fig. 3(c) is the accuracy of the focal length computed by assuming that the principal point is at the image origin, while the true principal point is horizontally shifted from by 0 pixels (solid line), 50 pixels (dashed line), and 100 pixels (dotted line) from the image origin. We can see that the focal length is not noticeably affected even if the principal point is several dozens of pixels away from its true position.

If we newly take images, we may calibrate the principal point of that camera using a reference pattern. If we are to analyze exiting photographs and paintings, however, the cameras used by the photographers or the perspective rules used by the painters are usually unknown. In such a case, our experiments suggest that it is more reasonable to assume an appropriate principal point rather than 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 describe this correction scheme later.

6. Orthogonality Correction

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. This process is necessary for lines that should be orthogonal to be exactly orthogonal after 3-D reconstruction.

First, we convert the N-vectors $\{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$ of the three vanishing points into $\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$, using the computed focal length f and Eqs. (5) and (6). A well known 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 the least-squares method minimizing

$$\|\boldsymbol{e}_1 - \hat{\boldsymbol{m}}_1\|^2 + \|\boldsymbol{e}_2 - \hat{\boldsymbol{m}}_2\|^2 + \|\boldsymbol{e}_3 - \hat{\boldsymbol{m}}_3\|^2. \tag{15}$$

The solution that minimizes this subject to the constraint that $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be orthonormal is analytically obtained as follows [3, 4]. First, we compute the singular

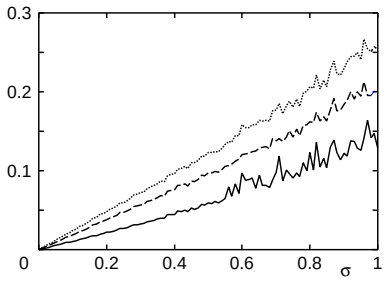


Figure 4: Accuracy of orthogonality correction. Solid line: optimal computation. Dashed line: least-squares method. Dotted line: no corrections.

value decomposition (SVD) of the matrix that has $\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$ as its columns:

$$\begin{pmatrix} \hat{\boldsymbol{m}}_1 & \hat{\boldsymbol{m}}_2 & \hat{\boldsymbol{m}}_3 \end{pmatrix} = \boldsymbol{V}\mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)\boldsymbol{U}^\top. \tag{16}$$

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 by

$$\begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{pmatrix} = \boldsymbol{V}\boldsymbol{U}^\top. \tag{17}$$

However, this solution treats the three vanishing points equally without considering the difference in their reliability. In order to account for this, we need to introduce appropriate weights to Eq. (15). As is well known in statistics, the optimal weights are given by

$$W_i = \frac{1}{\mathrm{tr}V_0[\boldsymbol{m}_i]}, \tag{18}$$

where $V_0[\boldsymbol{m}_i]$ is the normalized covariance matrix of the ith vanishing point computed by renormalization, and tr denotes the trace[4]. Then, the optimal solution is obtained by minimizing

$$W_1\|\boldsymbol{e}_1 - \hat{\boldsymbol{m}}_1\|^2 + W_2\|\boldsymbol{e}_2 - \hat{\boldsymbol{m}}_2\|^2 + W_3\|\boldsymbol{e}_3 - \hat{\boldsymbol{m}}_3\|^2 \tag{19}$$

instead of Eq. (15). The solution is obtained by replacing the left-hand side of Eq. (16) by $\begin{pmatrix} W_1\hat{\boldsymbol{m}}_1 & W_2\hat{\boldsymbol{m}}_2 & W_3\hat{\boldsymbol{m}}_3 \end{pmatrix}$ [4].

Using the simulated image of Fig. 2(a), we conducted the orthogonality correction after computing the focal length by the composite method of Sec. 4 1000 times with different noise each time. Then, we evaluated the average discrepancy of the corrected directions $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ from the true directions $\{\bar{\boldsymbol{m}}_1, \bar{\boldsymbol{m}}_2, \bar{\boldsymbol{m}}_3\}$ by

$$F = \sqrt{\frac{1}{1000}\sum_{a=1}^{1000}\sum_{i=1}^{3}\|\boldsymbol{e}_i^{(a)} - \bar{\boldsymbol{m}}_i\|^2}, \tag{20}$$

where $\{\boldsymbol{e}_i^{(a)}\}$ is the ath value. Since the orientation of an N-vector is indeterminate, we adjusted the sign so that $(\boldsymbol{e}_i, \bar{\boldsymbol{m}}_i) \geq 0$ before evaluating Eq. (20).

Fig. 4 plots F vs. the noise standard deviation σ (pixels) on the horizontal axis. The solid line is for the optimal solution (minimization of Eq. (19)); the dashed line is for the minimization of Eq. (15) without introducing the weights. For comparison, the dotted line shows the value of F for $\boldsymbol{e}_i = \hat{\boldsymbol{m}}_i$ (with no corrections). We can see from this that the optimal solution indeed achieves the lowest error.

[4]By the definition of the normalized covariance matrix $V_0[\boldsymbol{m}_i]$, the trace $\mathrm{tr}V_0[\boldsymbol{m}_i]$ is the mean square $E[\|\Delta\boldsymbol{m}_i\|^2]$ scaled so that the noise standard deviation σ is 1.

7. Image Data Correction

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. In contrast to the vanishing points, the deviations of the lines are very small when the noise in the feature points is very small, so the optimal correction can be done based on linear analysis [4].

Let $\boldsymbol{n}$ be the N-vector of the line in question, and $V_0[\boldsymbol{n}]$ its normalized 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 obtained by minimizing the squared Mahalanobis distance $(\Delta\boldsymbol{n}, V_0[\boldsymbol{n}]^-\Delta\boldsymbol{n})$ subject to the constraint $(\boldsymbol{n} - \Delta\boldsymbol{n}, \bar{\boldsymbol{m}}_i) = 0$, where $V_0[\boldsymbol{n}]^-$ is the Moore-Penrose generalized inverse of $V_0[\boldsymbol{n}]$. The final correction is given as follows [4]:

$$\bar{\boldsymbol{n}} = N\Big[\boldsymbol{n} - \frac{(\boldsymbol{n}, \boldsymbol{m}_i)}{(\boldsymbol{m}_i, V_0[\boldsymbol{n}]\boldsymbol{m}_i)} V_0[\boldsymbol{n}]\boldsymbol{m}_i\Big]. \tag{21}$$

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. If $\boldsymbol{x}$ and $\bar{\boldsymbol{n}}$ are the Z-vectors of the initial point and the N-vector of the displaced line, respectively, the vector $\boldsymbol{x}$ is corrected into $\bar{\boldsymbol{x}}$ as follows [4]:

$$\bar{\boldsymbol{x}} = Z[\boldsymbol{x} - (\bar{\boldsymbol{n}}, \boldsymbol{x})\bar{\boldsymbol{n}}] \tag{22}$$

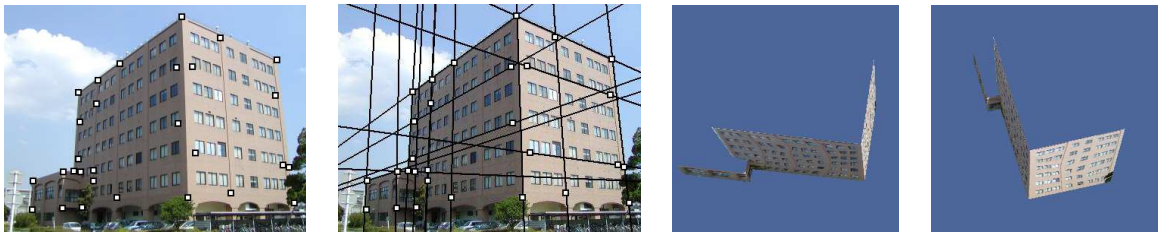
The N-vectors of the lines passing through displaced points are replaced by the first of Eqs. (4), the Z-vectors of the intersections of the displaced lines are replaced by the second of Eqs. (4), and this process is propagated.

8. Real Image Examples

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].

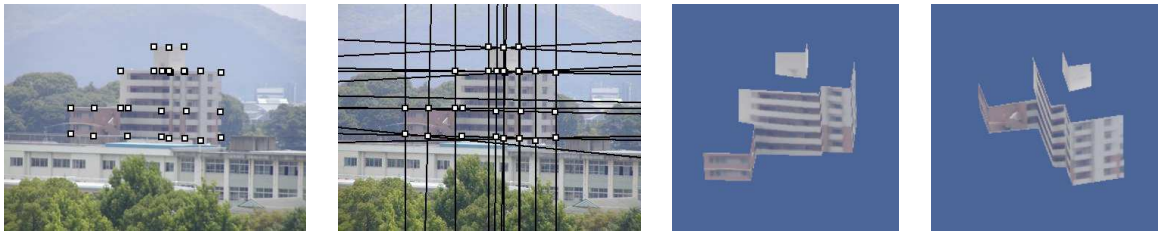
Fig. 5(a) is a short-range view of a building with a strong perspective effect (300×400 pixels). The selected feature points are marked in it. Using the three orthogonal directions drawn in Fig. 5(b), we estimated the focal length to be 416 pixels by least squares and 431 pixels by the optimal computation. In this case, the three angles defined by the vanishing points are all obtuse, so the composite method reduces to the optimal computation. Fig. 5(c),(d) are views of the reconstructed 3-D shape seen from two different angles.

Fig. 6(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. 6(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. 6(c),(d): lines that should be parallel are exactly parallel, and lines that should be orthogonal are exactly orthogonal.



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

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



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

Figure 6: Input image (distant view) and its 3-D reconstruction.

9. Concluding Remarks

In this paper, we analyzed the noise sensitivity of the focal length computation, the principal point estimation, and the orthogonality enforcement for single-view 3-D reconstruction. We pointed out that due to the nonlinearity of the computation the standard optimal computation is not actually optimal. 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 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*, September 1999, Kerkyra, Greece, Vol. 1, pp. 434–441.

[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 (1996-9), 1317–1322.