PAPER   *Special Section on Image Recognition and Understanding*

Statistical Optimization for 3-D Reconstruction from a Single View

Kenichi KANATANI[†a)]   *and* Yasuyuki SUGAYA[†], *Members*

SUMMARY   We analyze the noise sensitivity of the focal length computation, the principal point estimation, and the orthogonality enforcement for single-view 3-D reconstruction based on vanishing points and orthogonality. We point out that due to the nonlinearity of the problem the standard statistical optimization is not very effective. We present a practical compromise for avoiding the computational failure and preserving high accuracy, allowing a consistent 3-D shape in the presence of however large noise.
key words: *3-D reconstruction, statistical optimization, vanishing point, vanishing line, principal point*

1. Introduction

Given two or more views of a scene, we can reconstruct its 3-D structure based on triangulation [5], [8]. However, 3-D reconstruction is possible using only a single view if we have sufficient knowledge about the scene [1], [2], [8]. 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 detect three vanishing points of mutually orthogonal sets of parallel lines in the scene, we can compute the camera focal length and the principal point, from which we can compute the orientations of the lines in the scene.

This type of single-view 3-D reconstruction is widely used not only in industrial environments such as robotic manufacturing and navigation but also for entertainment, education, and scholastic research through 3-D reconstruction from paintings and historical photographs [2]–[4].

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. A typical symptom is that the inside of the square root becomes negative when we estimate the camera focal length, resulting in an imaginary value.

The purpose of this paper is twofold: We 1) analyze in detail the effects of noise on the focal length computation, the principal point estimation, and the orthogonality enforcement and 2) present a practical compromise for avoiding computational failure and preserving high accuracy.

First, we apply the standard statistical optimization procedure [9]: we evaluate covariances of intermediate quantities by cascading Jacobian matrices and minimize the final covariance. Then, we point out that the results are not very satisfactory because of the strong nonlinearity of the problem. To cope with them, we incorporate a practical remedy and show that our method can avoid computational failure for however large noise, yet maintains nearly optimal accuracy. We evaluate its performance by simulation and show that our method is relatively insensitive to the principal point position. We give a geometric interpretation to our observations.

Finally, we describe a procedure that can reconstruct a consistent 3-D shape in the presence of however large noise, incorporating orthogonality enforcement and image correction techniques.

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 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*.

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 assumed 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).

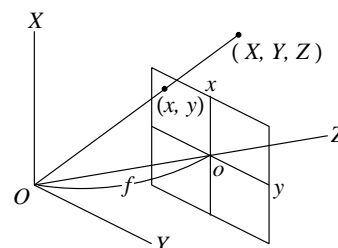We represent an image point (x, y) by the following

Fig. 1   Perspective projection.

3-D vectors:

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

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

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

Throughout this paper, $Z[\,\cdot\,]$ designates normalization to make the third component 1, and $N[\,\cdot\,]$ normalization into a unit vector. We call $\boldsymbol{x}$ and $\boldsymbol{m}$ their *Z-vector* and *N-vector*, respectively [8].

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

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

the *N-vector* of the line [8]. 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 $(\boldsymbol{a}, \boldsymbol{b})$.

Let $\boldsymbol{n}$ be the N-vector $\boldsymbol{n}$ of the line passing through two points with Z-vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and let $\boldsymbol{x}$ be the Z-vector $\boldsymbol{x}$ of the intersection of two lines with N-vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$. They are given as follows [8]:

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

3. Vanishing Point 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 [5], [8].

If we observe non-parallel coplanar lines in the scene, their vanishing points are collinear in the image, defining the *vanishing line* (Fig. 2). 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 [5], [8].
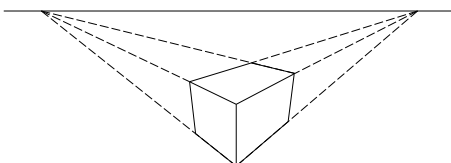

Fig. 2　Vanishing points and vanishing line.

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 [10]. It is summarized as follows [9]:

The reliability of a line in the image is evaluated by the following *normalized covariance matrix*[††] [9]:

$$V_0[\boldsymbol{n}] = \frac{\boldsymbol{P}_{\mathbf{n}}(\boldsymbol{x} \times \boldsymbol{P}_{\mathbf{k}} \times \boldsymbol{x} + \boldsymbol{y} \times \boldsymbol{P}_{\mathbf{k}} \times \boldsymbol{y})\boldsymbol{P}_{\mathbf{n}}}{\|\boldsymbol{x} \times \boldsymbol{y}\|}. \tag{5}$$

Here, $\boldsymbol{x}$ and $\boldsymbol{y}$ are the Z-vectors of the endpoints that define the line, and $\boldsymbol{n}$ is the N-vector of that line. We define

$$\boldsymbol{P}_{\mathbf{n}} = \boldsymbol{I} - \boldsymbol{n}\boldsymbol{n}^{\top}, \quad \boldsymbol{P}_{\mathbf{k}} = \boldsymbol{I} - \boldsymbol{k}\boldsymbol{k}^{\top}, \tag{6}$$

where $\boldsymbol{I}$ is the unit matrix and $\boldsymbol{k} = (0\ 0\ 1)^{\top}$. The superscript $\top$ denotes transpose. The matrices $\boldsymbol{P}_{\mathbf{n}}$ and $\boldsymbol{P}_{\mathbf{k}}$, respectively, define projections along $\boldsymbol{n}$ and $\boldsymbol{k}$ onto the planes perpendicular to them.

In Eq. (5), the product $\boldsymbol{u} \times \boldsymbol{A}$ of vector $\boldsymbol{u}$ and matrix $\boldsymbol{A}$ is a matrix whose columns are the vector products of $\boldsymbol{u}$ and the columns of $\boldsymbol{A}$, while the product $\boldsymbol{A} \times \boldsymbol{v}$ of matrix $\boldsymbol{A}$ and vector $\boldsymbol{v}$ is a matrix whose rows are the vector products of $\boldsymbol{v}$ and the rows of $\boldsymbol{A}$. The product $\boldsymbol{u} \times \boldsymbol{A} \times \boldsymbol{v}$ is unambiguously defined by associativity [9].

Let $\boldsymbol{n}_1, ..., \boldsymbol{n}_N$ be the N-vectors of lines that have a common intersection, and $V_0[\boldsymbol{n}_1], ..., V_0[\boldsymbol{n}_N]$ their normalized covariance matrices. Vanishing points are usually located very far away from the center of the image, so we represent it by its N-vector $\boldsymbol{m}$. It can be optimally computed together with its normalized covariance matrix $V_0[\boldsymbol{m}]$ as follows [9]:

1. Let $c = 0$ and $W_\alpha = 1$, $\alpha = 1, ..., N$.
2. Compute the following matrices:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{n}_\alpha \boldsymbol{n}_\alpha^{\top},$$

$$\boldsymbol{N} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{n}_\alpha]. \tag{7}$$

3. Compute the three eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and the corresponding orthonormal system of eigenvectors $\{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3\}$ of the matrix

$$\hat{\boldsymbol{M}} = \boldsymbol{M} - c\boldsymbol{N}. \tag{8}$$

4. If $|\lambda_3| \approx 0$, return $\boldsymbol{m} = \boldsymbol{m}_3$ and

$$V_0[\boldsymbol{m}] = \frac{1}{N}\Big(\frac{\boldsymbol{m}_1\boldsymbol{m}_1^{\top}}{\lambda_1} + \frac{\boldsymbol{m}_2\boldsymbol{m}_2^{\top}}{\lambda_2}\Big). \tag{9}$$

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

[††]The covariance matrix of the N-vector $\boldsymbol{n}$ of the line connecting two points, whose x and y coordinates are perturbed by independent Gaussian noise of mean 0 and standard deviation σ (pixels) is $\sigma^2 V_0[\boldsymbol{n}]$ [9]. Since multiplication by a positive constant does not affect the subsequent computation, we normalize σ^2 to 1 and call Eq. (5) the "normalized" covariance matrix.

5. Else, update c and W_α by

$$c \leftarrow c + \frac{\lambda_3}{(\boldsymbol{m}_3, \boldsymbol{N}\boldsymbol{m}_3)},$$

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{m}_3, V_0[\boldsymbol{n}_\alpha]\boldsymbol{m}_3)}, \tag{10}$$

and go back to Step 2.

4. Focal Length Estimation

Suppose the vanishing points of three sets of parallel lines with mutually orthogonal directions 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 $V_0[\boldsymbol{m}_3]$ their normalized covariance matrices (computed by the renormalization procedure; see Eq. (9)). The unit vectors $\hat{\boldsymbol{m}}_1$, $\hat{\boldsymbol{m}}_2$, and $\hat{\boldsymbol{m}}_3$ starting from the viewpoint O and pointing toward these vanishing points are given by

$$\hat{\boldsymbol{m}}_i = N[\boldsymbol{I}_f \boldsymbol{m}_i], \quad i = 1, 2, 3, \tag{11}$$

where

$$\boldsymbol{I}_f \equiv \text{diag}\left(1, 1, \frac{f}{f_0}\right). \tag{12}$$

The symbol $\text{diag}(\cdots)$ 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 constraints,

$$e_1 \equiv (\hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3) = (\boldsymbol{m}_2, \boldsymbol{I}_f^2 \boldsymbol{m}_3) = 0,$$
$$e_2 \equiv (\hat{\boldsymbol{m}}_3, \hat{\boldsymbol{m}}_1) = (\boldsymbol{m}_3, \boldsymbol{I}_f^2 \boldsymbol{m}_1) = 0,$$
$$e_3 \equiv (\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2) = (\boldsymbol{m}_1, \boldsymbol{I}_f^2 \boldsymbol{m}_2) = 0. \tag{13}$$

from which the focal length f can be solved. This fact has been well known [5], [8].

These do not necessarily hold strictly in the presence of noise, so many researchers use a naive least-squares method [5], [8]. According to the standard statistical optimization theory, however, an optimal estimate f is obtained by minimizing the following function [9]:

$$J = \sum_{i,j=1}^{3} W_{ij} e_i e_j. \tag{14}$$

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

$$\boldsymbol{W} = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix}^{-1}, \tag{15}$$

where V_{ij} is the covariance of e_i and e_j (V_{ii} is the variance of e_i). Using Eq. (11), we obtain

$$V_{11} = (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_3) + (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_2),$$
$$V_{22} = (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_1) + (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_3),$$
$$V_{33} = (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_2) + (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_1),$$

$$V_{23} = V_{32} = (\boldsymbol{m}_2, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_1] \boldsymbol{I}_f^2 \boldsymbol{m}_3),$$
$$V_{31} = V_{13} = (\boldsymbol{m}_3, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_2] \boldsymbol{I}_f^2 \boldsymbol{m}_1),$$
$$V_{12} = V_{21} = (\boldsymbol{m}_1, \boldsymbol{I}_f^2 V_0[\boldsymbol{m}_3] \boldsymbol{I}_f^2 \boldsymbol{m}_2). \tag{16}$$

The matrix $\boldsymbol{W} = (W_{ij})$ weights the three constraints of Eqs. (13) 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 constant times the unit matrix $\boldsymbol{I}$, so the minimization of Eq. (14) would reduce to the *least-squares method* that minimizes $\sum_{i=1}^{3} e_i^2$.

Equation (14) 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. (14) is a quadratic polynomial in

$$\alpha = \left(\frac{f}{f_0}\right)^2, \tag{17}$$

so the value α that minimizes Eq. (14) 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

$$f = f_0 \sqrt{\alpha}. \tag{18}$$

5. Composite Method

5.1 Motivations

The optimality of the above procedure is based on linear analysis. Specifically, 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 for small noise [9]. This means that if 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, the vanishing point computation is highly nonlinear, particularly so when it is located very far a way from the frame center. Hence, even if noise in the feature points is zero-mean, the errors in the vanishing point may have large bias, and the equiprobability contours can be parabolic or hyperbolic, diverging 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. (14) is optimal no longer holds. A typical symptom is that the value α computed by Eq. (14) becomes negative, resulting in an imaginary focal length f.

Since points and lines are represented by unit vectors, no computational failure occurs if the vanishing point is at infinity. However, the uncertainty of the final output is independent of the mathematical representations of intermediate quantities.
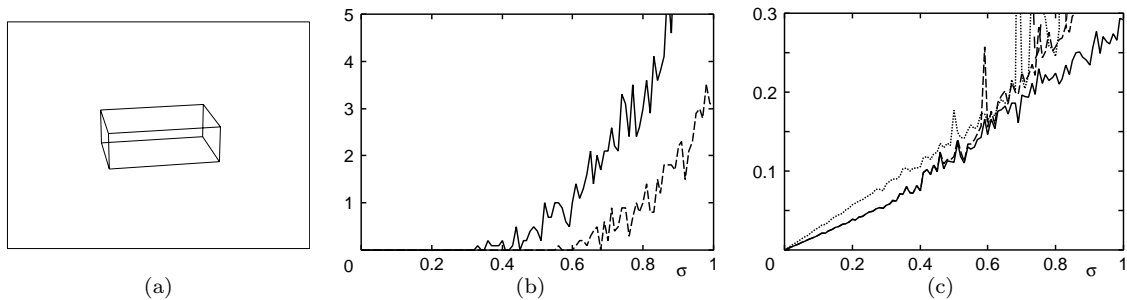
Fig. 3 (a) Simulated image of a box. (b) The percentage (%) of computational failure. Solid line: optimal computation. Dotted line: least squares. (c) Accuracy of focal length computation. Solid line: composite method. Dashed line: optimal computation. Dotted line: least squares.

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. To be specific, the three vanishing points cannot be anywhere but should be at the vertices of a triangle whose orthocenter is at the principal point [1], [5], [8], 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, and these conditions may be violated. For such an inadmissible configuration, a real focal length solution may no longer exist.

5.2 Procedure

From the above 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 principal point:

Case 1: Three obtuse angles. We regard the three vanishing points as sufficiently reliable and do the optimal computation as described in Sect. 4.

Case 2: One acute angle. We remove from among the three constraints in Eqs. (13) the one involving the two directions that make an acute angle and minimize Eq. (14) subject to the remaining two constraints.

Case 3: Two acute angles. We retain from among the three constraints in Eqs. (13) only the one involving the two directions that make an obtuse angle, removing the others as unreliable. In this case, we need not minimize Eq. (14): the solution that makes Eq. (14) (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).

Our strategy can be viewed as reducing the influence of the principal point estimation, since if some of the three angles are close to 90° or 180°, a small displacement of the principal point significantly affects the focal length (often resulting in an imaginary value). This sensitivity to the principal point position agrees with the analysis of Hartley and Kauric [6].

An alternative approach may be to displace the principal point so that the three angles become all obtuse. To determine the displacement optimally, we need an empirical distribution of the possible principal point positions (the Bayesian approach), as Hartley and Silpa-Anan [7] did for two-view analysis, but the result strongly depends on the choice of the prior, and there is no good way to define it. So, we adopt the above non-Bayesian approach.

5.3 Experiments

Figure 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 vertex positions independently and estimated the focal length.

Figure 3(b) plots the percentage (%) of computational failures (resulting in an imaginary focal length or no convergence[†]) of the optimal computation of Sect. 4 (solid line) over 1000 trials using different noise for each σ. The dotted line shows the corresponding result by least squares (Eq. (14) is replaced by $\sum_{i=1}^{3} e_i^2$). The composite computation is not plotted because it did not fail. We can see that the rate of computational failures increases as noise increases. It is larger for the optimal computation than for the least squares. This indicates that pursuit for high accuracy is generally not compatible with computational robustness.

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

$$D = \sqrt{\frac{1}{1000} \sum_{a=1}^{1000} \left(\frac{f^{(a)} - f}{f^{(a)}} \right)^2}, \qquad (19)$$

where $f^{(a)}$ is the ath instance of the 1000 trials. We let $f^{(a)} = \infty$ if the computation failed, so that D is

---

[†]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.

defined even if the computation fails: $(f^{(a)} - f)/f^{(a)}$ $= 1 - f/f^{(a)} = 1$ for $f^{(a)} = \infty$. Figure 3(c) plots the value D for the noise standard deviation σ. 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 noise is small, but the error irregularly fluctuates as noise increases. The composite method retains high accuracy for small noise and preserves the accuracy expected of the optimal computation even for large noise.

6. Principal Point Estimation

6.1 Procedure

So far, we have assumed that the principal point is known and taken to be the image origin. Now, 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 sets of parallel lines with mutually orthogonal directions [5], [8].

Let $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ be the N-vectors of the three vanishing points. The intersection of the line of sight of $\boldsymbol{m}_i$ with the plane $Z = 1$ is at $\boldsymbol{m}_i/m_{iz}$, where m_{iz} is the Z component of $\boldsymbol{m}_i$. The condition that vector $\boldsymbol{h}$ represents the orthocenter of the triangle defined by $\boldsymbol{m}_1/m_{1z}$, $\boldsymbol{m}_2/m_{2z}$, and $\boldsymbol{m}_3/m_{3z}$ is

$$\left(\frac{\boldsymbol{m}_1}{m_{1z}} - \boldsymbol{h}, \frac{\boldsymbol{m}_2}{m_{2z}} - \frac{\boldsymbol{m}_3}{m_{3z}} \right) = 0,$$

$$\left(\frac{\boldsymbol{m}_2}{m_{2z}} - \boldsymbol{h}, \frac{\boldsymbol{m}_3}{m_{3z}} - \frac{\boldsymbol{m}_1}{m_{1z}} \right) = 0,$$

$$\left(\frac{\boldsymbol{m}_3}{m_{3z}} - \boldsymbol{h}, \frac{\boldsymbol{m}_1}{m_{1z}} - \frac{\boldsymbol{m}_2}{m_{2z}} \right) = 0. \qquad (20)$$

Although any two of these are sufficient[†], we use all of them to retain symmetry. Let

$$\boldsymbol{u}_1 = (m_{2z}m_{3z})\boldsymbol{m}_1, \qquad \boldsymbol{u}_2 = (m_{3z}m_{1z})\boldsymbol{m}_2,$$

$$\boldsymbol{u}_3 = (m_{1z}m_{2z})\boldsymbol{m}_3, \qquad \boldsymbol{g} = (m_{1z}m_{2z}m_{3z})\boldsymbol{h}, \quad (21)$$

Multiplying Eqs. (20) by $m_{1z}m_{2z}m_{3z}$ and noting that $h_z = 1$, we can rearrange Eqs. (20) in the form

$$\boldsymbol{A}^\top \boldsymbol{g} = \boldsymbol{b}, \qquad (22)$$

where

$$\boldsymbol{A} = (\ \boldsymbol{u}_2 - \boldsymbol{u}_3 \quad \boldsymbol{u}_3 - \boldsymbol{u}_1 \quad \boldsymbol{u}_1 - \boldsymbol{u}_2 \quad \boldsymbol{k}\ ), \quad (23)$$

$$\boldsymbol{b} = \begin{pmatrix} (\boldsymbol{u}_1, \boldsymbol{u}_2 - \boldsymbol{u}_3) \\ (\boldsymbol{u}_2, \boldsymbol{u}_3 - \boldsymbol{u}_1) \\ (\boldsymbol{u}_3, \boldsymbol{u}_1 - \boldsymbol{u}_2) \\ m_{1z}m_{2z}m_{3z} \end{pmatrix}. \qquad (24)$$

(We put $\boldsymbol{k} = (0\ 0\ 1)^\top$ as before.) Eq. (22) can be solved by least squares as follows:

$$\boldsymbol{g} = (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\boldsymbol{b}, \qquad \boldsymbol{h} = Z[\boldsymbol{g}]. \qquad (25)$$

The principal point (x_c, y_c) is at $(f_0 h_x, f_0 h_y)$.

6.2 Experiments

The above procedure produces the orthocenter for whatever vanishing point configuration, but it can be the principal point only if it is inside the triangle defined by the vanishing points. However, even small noise can shift the principal point out of the triangle.

Using the simulated image of Fig. 3(a), we estimated the principal point after adding Gaussian noise to the vertices. We evaluated its accuracy by the root-mean-square distance from its true position over 1000 trials

$$E = \sqrt{\frac{1}{1000} \sum_{a=1}^{1000} \left((x_c^{(a)})^2 + (y_c^{(a)})^2 \right)}, \qquad (26)$$

where $(x_c^{(a)}, y_c^{(a)})$ is the ath estimates. The solid line (left scale) in Fig. 4(a) plots E in pixels vs. σ. The dashed line (right scale) shows the percentage (%) of the computed principal point being outside the triangle of the vanishing points.

We also computed the focal length by the composite method described in Sect. 5 using the estimated principal point. The solid line in Fig. 4(b) shows the accuracy of the computed focal length measured in D in Eq. (19). The dashed line is for using the true principal point.

Fig. 4(c) shows 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 away from by 0 pixels (solid line), 50 pixels (dashed line), and 100 pixels (dotted line) from the image origin.

6.3 Observations

Fig. 4(a) shows that 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 can stably obtain better results using its approximate position, as we can see from Fig. 4(b), which shows that the accuracy significantly deteriorates if principal point estimation is incorporated. Fig. 4(c) shows that the focal length is not significantly affected even if the principal point is assumed to be several dozens of pixels away from its true position.

This seems to contradict Hartley and Kauric [6], who maintained after theoretical analysis that the focal length can be significantly affected by the principal point position, particularly so when it is close to the horizon. It turns out, however, that such critical configurations exactly coincide with the situations in which estimation of the three vanishing points and the focal

---

[†]As is well known, the orthocenter is at the intersection of two lines, each passing through one vertex and perpendicularly intersecting the opposite side; the remaining line automatically passes through the orthocenter.
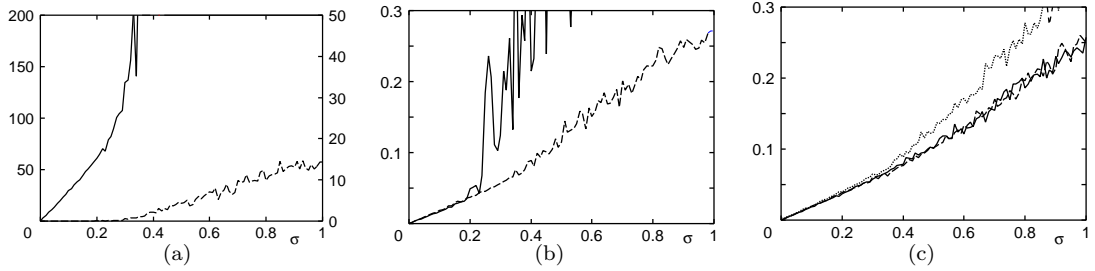
Fig. 4   (a) Accuracy of principal point estimation. Solid line (left scale): root-mean-square error (pixels). Dashed line (right scale): percentage (%) of being outside the triangle of the vanishing points. (b) Accuracy of focal length estimation. Solid line: estimating the principal point. Dashed line: 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).

length from them is very unreliable, since in such configurations, some of the three obtuse angles mentioned in the preceding section are close to 90° or 180°.

Our conclusion is that as long as the focal length is stably computed from vanishing points, its accuracy is not much affected by the principal point position, while if the focal length cannot be accurately determined from vanishing points, *it is already inaccurate whether the principal point is correct or not.* Since our composite method selectively avoids the use of angles close to 90° or 180°, it effectively reduces the influence of the principal point position on the computed focal length.

If we want to analyze exiting photographs and paintings without any information about the principal point, our experiments suggest that it suffices 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. That should result in 3-D shapes that better agree with human impression. We now describe such a procedure.

7.   Orthogonality Correction

The three directions from the viewpoint to the vanishing points may not be strictly orthogonal even though the focal length is optimally computed. So, we correct them to be strictly orthogonal with respect to the computed focal length. This process is necessary for lines that should be orthogonal in the scene to be strictly 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 $\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$ for the computed focal length f, using Eqs. (11) and (12). 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 minimizing

$$\|\boldsymbol{e}_1 - \hat{\boldsymbol{m}}_1\|^2 + \|\boldsymbol{e}_2 - \hat{\boldsymbol{m}}_2\|^2 + \|\boldsymbol{e}_3 - \hat{\boldsymbol{m}}_3\|^2. \quad (27)$$

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 [8], [9]. First, we compute the singular value decomposition (SVD) of the matrix that has

$\{\hat{\boldsymbol{m}}_1, \hat{\boldsymbol{m}}_2, \hat{\boldsymbol{m}}_3\}$ as its columns:

$$\begin{pmatrix} \hat{\boldsymbol{m}}_1 & \hat{\boldsymbol{m}}_2 & \hat{\boldsymbol{m}}_3 \end{pmatrix} = \boldsymbol{V}\mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)\boldsymbol{U}^\top. \quad (28)$$

Here, $\boldsymbol{V}$ and $\boldsymbol{U}$ are orthogonal matrices, 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

$$\begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{pmatrix} = \boldsymbol{V}\boldsymbol{U}^\top. \quad (29)$$

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. (27). As is well known in statistics, the optimal weights are given by

$$W_i = \frac{1}{\mathrm{tr}V_0[\boldsymbol{m}_i]}, \quad (30)$$

where $V_0[\boldsymbol{m}_i]$ is the normalized covariance matrix of the ith vanishing point computed by Eq. (9), and tr denotes the trace†. The optimal solution is obtained by minimizing

$$W_1\|\boldsymbol{e}_1 - \hat{\boldsymbol{m}}_1\|^2 + W_2\|\boldsymbol{e}_2 - \hat{\boldsymbol{m}}_2\|^2 + W_3\|\boldsymbol{e}_3 - \hat{\boldsymbol{m}}_3\|^2 (31)$$

instead of Eq. (27). The solution is still given by Eq. (29) if the left-hand side of Eq. (28) is replaced by $\begin{pmatrix} W_1\hat{\boldsymbol{m}}_1 & W_2\hat{\boldsymbol{m}}_2 & W_3\hat{\boldsymbol{m}}_3 \end{pmatrix}$ [9].

Using the simulated image of Fig. 3(a), we conducted the orthogonality correction after computing the focal length by the composite method of Sect. 5 1000 times using different noise each time. 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

$$F = \sqrt{\frac{1}{1000} \sum_{a=1}^{1000} \sum_{i=1}^{3} \|\boldsymbol{e}_i^{(a)} - \bar{\boldsymbol{m}}_i\|^2}, \quad (32)$$

where $\{\boldsymbol{e}_i^{(a)}\}$ is the ath value. Since the orientation of an N-vector is indeterminate, we adjusted the sign so

---

†From 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.

that $(\boldsymbol{e}_i, \bar{\boldsymbol{m}}_i) \geq 0$ before evaluating Eq. (32).

Figure 5 plots F vs. the noise standard deviation σ (pixels). The solid line is for the optimal solution (minimization of Eq. (31)); the dashed line is for the minimization of Eq. (27). For comparison, the dotted line shows the value of F for $\boldsymbol{e}_i = \hat{\boldsymbol{m}}_i$ (with no corrections). We can see that the optimal solution indeed achieves the lowest error.

8. 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 lines so that they pass through their respective vanishing points. In contrast to vanishing points, the deviations of lines are small when noise in the feature points that define them is small, so the optimal correction can be done using the covariance matrix of Eq. (5) [9].

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 [9]:

$$\bar{\boldsymbol{n}} = N\Big[\boldsymbol{n} - \frac{(\boldsymbol{n}, \boldsymbol{m}_i)V_0[\boldsymbol{n}]\boldsymbol{m}_i}{(\boldsymbol{m}_i, V_0[\boldsymbol{n}]\boldsymbol{m}_i)}\Big]. \tag{33}$$

If 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 [9]:

$$\bar{\boldsymbol{x}} = Z[\boldsymbol{x} - (\bar{\boldsymbol{n}}, \boldsymbol{x})\bar{\boldsymbol{n}}] \tag{34}$$

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.

9. 3-D Shape Reconstruction

A plane in the scene is represented by $AX + BY + CZ$
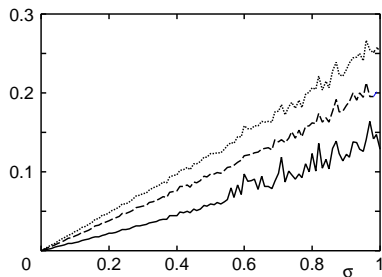
Fig. 5 Accuracy of orthogonality correction. Solid line: optimal computation. Dashed line: least squares. Dotted line: no corrections.

$= h$. To remove the scale indeterminacy, we normalize the coefficients into $A^2 + B^2 + C^2 = 1$. Thus, $\boldsymbol{\nu} = (A\ B\ C)^\top$ is the unit surface normal, and h is the (signed) distance of the plane from the origin O (positive in the $\boldsymbol{\nu}$ direction). If we write $\boldsymbol{r} = (X\ Y\ Z)^\top$, the equation of the plane is written as

$$(\boldsymbol{\nu}, \boldsymbol{r}) = h. \tag{35}$$

If a point $\boldsymbol{r}$ on the plane (35) is projected onto a point in the image with Z-vector $\boldsymbol{x}$, the point $\boldsymbol{r}$ is at the intersection of the plane (35) with the line of sight starting from the origin O and extending in the direction $\boldsymbol{x}$. Hence, the position $\boldsymbol{r}$ is given by the following *backprojection* of $\boldsymbol{x}$ (Fig. 6):

$$\boldsymbol{r} = \frac{h\boldsymbol{x}}{(\boldsymbol{\nu}, \boldsymbol{x})}. \tag{36}$$

If two non-parallel lines on a plane in the scene are projected onto the image and define vanishing points with N-vectors $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$, the unit surface normal $\boldsymbol{\nu}$ to that plane is give by

$$\boldsymbol{\nu} = N[\boldsymbol{m}_1 \times \boldsymbol{m}_2]. \tag{37}$$

The distance of this plane from the origin O is computed in the following three ways.

Case 1: Distance known

If we know that the backprojections of particular two points with Z-vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ should have distance d_{12} in the scene, the absolute distance $|h|$ to the plane is determined as follows:

$$|h| = d_{12} \Big/ \Big\| \frac{\boldsymbol{x}_1}{(\boldsymbol{\nu}, \boldsymbol{x}_1)} - \frac{\boldsymbol{x}_2}{(\boldsymbol{\nu}, \boldsymbol{x}_2)} \Big\|. \tag{38}$$

The equation of the plane is $(\pm\boldsymbol{\nu}, \boldsymbol{r}) = |h|$, and the choice of the sign depends on whether the image origin is inside the projection of that plane or outside it. In the former case, $\boldsymbol{\nu}$ is signed so that its Z-component is positive; in the latter, negative (the former is the case in the usual shooting condition[†]).
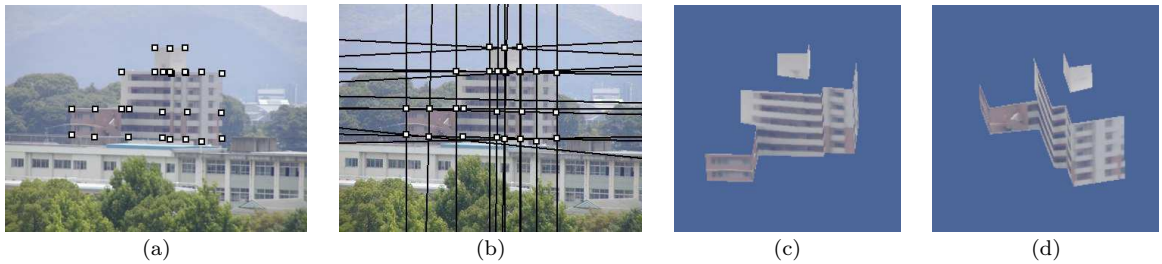
Case 2: Intersecting with a plane
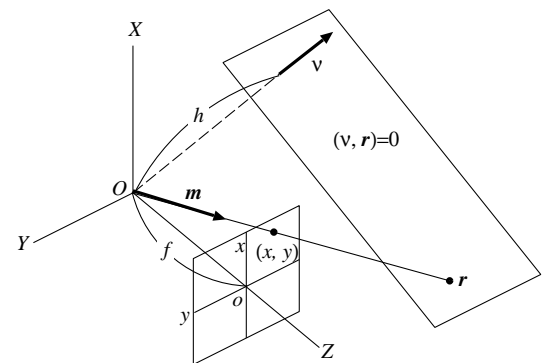
Fig. 6 Backprojection of image points.

[†]The sign is indeterminate if the image origin happens to be exactly on the vanishing line of the plane. Such a case is exceptional, so we can ignore it.
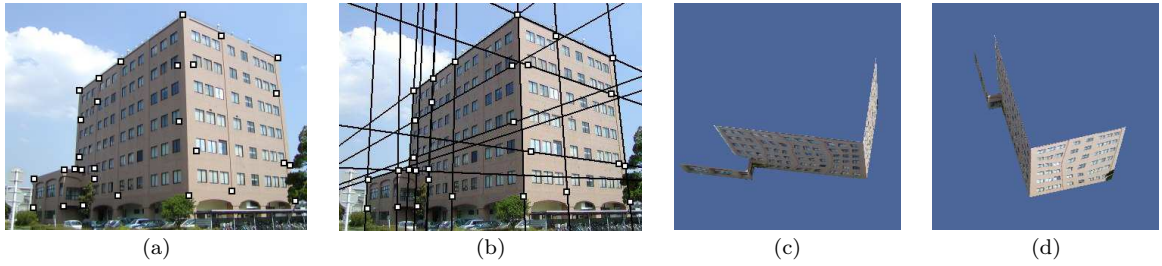
Fig. 7 Input image (short-range view) and its 3-D reconstruction.

Fig. 8 Input image (distant view) and its 3-D reconstruction.

If the plane $(\boldsymbol{\nu}, \boldsymbol{r}) = h$ intersects with another plane $(\boldsymbol{\nu}_0, \boldsymbol{r}) = h_0$ whose position and orientation are already computed, the backprojection of a point on their intersection should be the same whichever plane is used. Hence, if we let $\boldsymbol{x}$ be the Z-vector of an arbitrary point on the intersection, we obtain

$$h = \frac{(\boldsymbol{\nu}, \boldsymbol{x})}{(\boldsymbol{\nu}_0, \boldsymbol{x})} h_0. \tag{39}$$

Case 3: Intersecting with two planes

If the plane $(\boldsymbol{\nu}, \boldsymbol{r}) = h$ intersects with two planes whose positions and orientations are already computed, we can determine both the orientation $\boldsymbol{\nu}$ and the position h. First, we can compute the 3-D positions of points on the intersections with the two known planes by the backprojection (36) using the known planes. If we take points $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ on one intersection and points $\boldsymbol{r}_3$ and $\boldsymbol{r}_4$ on the other, we have $(\boldsymbol{x}_1 - \boldsymbol{x}_2, \boldsymbol{\nu}) = 0$ and $(\boldsymbol{x}_3 - \boldsymbol{x}_4, \boldsymbol{\nu}) = 0$. Hence, we have

$$\boldsymbol{\nu} = N[(\boldsymbol{r}_1 - \boldsymbol{r}_2) \times (\boldsymbol{r}_3 - \boldsymbol{r}_4)]. \tag{40}$$

The distance h can be computed by Eq. (39) using either of the two planes.

10. Real Image Examples

Thus, we can determine the positions and orientations of all visible planes by specifying the distance between two points on one plane and propagating that information to the remaining planes by the above procedure. If the distance of the initial two points is not known, we can set it to an arbitrary value; the 3-D shape is determined up to a scale factor.

Figure 7(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. 7(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. Figure 7(c),(d) are views of the reconstructed 3-D shape seen from two different angles.

Figure 8(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. 8(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. Our image correction procedure reconstructs a consistent 3-D shape, as displayed in the two right images in Fig. 8(c),(d). Lines that should be parallel are indeed strictly parallel, and lines that should be orthogonal are strictly orthogonal.

We conducted a simple camera calibration after shooting the images in Figs. 7 and 8, using a reference pattern. The focal length was computed to be 457 and 4060 pixels, respectively. These values do not so well agree with the above estimates, but our proposed method gives the closest values to them.

11. 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 problem the standard statistical optimization is not very effective. We presented a practical compromise between avoiding the computational

failure and preserving high accuracy. We evaluated its performance by simulation and showed that our method is relatively insensitive to the principal point position. We gave a geometric interpretation to our observations. Finally, we described a procedure for reconstructing a consistent 3-D shape in the presence of however large noise, incorporating orthogonality enforcement and image correction techniques.

Acknowledgments

References

[1] B. Caprile and V. Torre, "Using vanishing points for camera calibration," Int. J. Comput. Vis., vol.4, no. 2, pp.127–140, March, 1990.

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

[3] P.E. Debevec, C.J. Taylor, and J.Malik, "Modelling and rendering architecture for photographs: A hybrid geometry- and image-based approach," SIGGRAPH'96, pp.11–20, New Orleans, LA, U.S.A., Aug. 1996.

[4] S. El-Hakim, "A flexible approach to 3D reconstruction from single views," SIGGRAPH 2001, Sketches and Applications, San Antonio, TX, U.S.A., July 2001.

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

[6] R. I. Hartley and R. Kauric, "Sensitivity of calibration to principal point position," Proc. 7th Euro. Conf. Comput. Vision, vol.2, pp.433–446, Copenhagen, Denmark, May 2002.

[7] R. Hartley and C. Silpa-Anan, "Reconstruction from two views using approximate calibration," Proc. 5th Asian Conf. Comput. Vision, pp.23–25, Melbourne, Australia, Jan. 2002.

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

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

[10] Y. Kanazawa and K. Kanatani, "Optimal line fitting and reliability evaluation," IEICE Trans. Inf. & Syst., vol.E79-D, no.9, pp.1317–1322, Sept. 1996.

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, he is currently Professor of computer science at Okayama University, Okayama, Japan. He is the author of many books on computer vision, and he has received many awards including best paper awards from IPSJ (1987) and IEICE (2005). He is an IEEE Fellow.

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. He is currently Assistant Professor of computer science at Okayama University, Okayama, Japan. His research interests include image processing and computer vision. He received the IEICE best paper award in 2005.