STATISTICAL RELIABILITY OF 3-D COMPUTING FROM IMAGES

Kenichi Kanatani
Department of Computer Science, Gunma University
Kiryu, Gunma 376, Japan

ABSTRACT

The reliability of 3-D interpretations computed from images is analyzed in statistical terms. The reliability of line fitting to edges is evaluated quantitatively, and the reliability of vanishing point estimation is deduced quantitatively. The result is applied to focal length calibration. The reliability of fitting an orthogonal frame to three orientations is also discussed, and statistical criteria for testing geometric hypotheses are derived.

1. STATISTICAL MODEL OF NOISE

In robotics applications of computer vision, reliability of computation is of the utmost importance. In the past, reliability of computation has often been tested empirically by using synthetic and real data. For further progress, we need a *theoretical* basis for it.

Assume the camera imaging model [3, 4] shown in Fig. 1: the origin O is called the *viewpoint*; the constant f is called the *focal length*. A point P on the image plane is represented by the unit vector $\boldsymbol{m}$ starting from O and pointing toward P; A line l on the image plane is represented by the unit surface normal $\boldsymbol{n}$ to the plane passing through O and l. We call $\boldsymbol{m}$ and $\boldsymbol{n}$ *N-vectors* [4].

Let $\boldsymbol{m}$ be the N-vector of a point data on the image plane when there is no noise. In the presence of noise, a perturbed N-vector $\boldsymbol{m}' = \boldsymbol{m} + \Delta\boldsymbol{m}$ is observed. The error term $\Delta\boldsymbol{m}$ is regarded as a *random variable*. Consider the *covariance matrix*

$$V[\boldsymbol{m}] = E[\Delta\boldsymbol{m}\Delta\boldsymbol{m}^\top], \tag{1}$$

where $E[\cdot]$ denotes expectation. Suppose noise occurs at each pixel on the image plane and is equally likely in all orientations with the same root-mean-square ϵ (measured in pixels), which we call the *image accuracy*. If the size of the image is small compared with the focal length f, the covariance matrix $V[\boldsymbol{m}]$ of the N-vector of a data pixel has the following form [5]:

$$V[\boldsymbol{m}] = \frac{\tilde{\epsilon}^2}{2}(\boldsymbol{I} - \boldsymbol{m}\boldsymbol{m}^\top), \qquad \tilde{\epsilon} = \frac{\epsilon}{f}. \tag{2}$$

2. RELIABILITY OF EDGES

Let $\boldsymbol{n}$ be the N-vector of a line fitted to edge pixels in the absence of noise. In the presence of noise, each edge pixel is displaced [1]. Let $\boldsymbol{n}' = \boldsymbol{n} + \Delta\boldsymbol{n}$ be the N-vector of the line fitted to displaced edge pixels (Fig. 2). The reliability of the fit is described by the covariance matrix

$$V[\boldsymbol{n}] = E[\Delta\boldsymbol{n}\Delta\boldsymbol{n}^\top]. \tag{3}$$

If a line is fitted to an edge segment of length w in orientation $\boldsymbol{u}$, the covariance matrix $V[\boldsymbol{n}]$ of the N-vector $\boldsymbol{n}$ of the fitted line has the following form [5]:

$$V[\boldsymbol{n}] = \frac{6\kappa}{w^3}\boldsymbol{u}\boldsymbol{u}^\top + \frac{\kappa}{2f^2 w}\boldsymbol{m}_G\boldsymbol{m}_G^\top, \quad \kappa = \frac{\epsilon^2}{\gamma}. \tag{4}$$

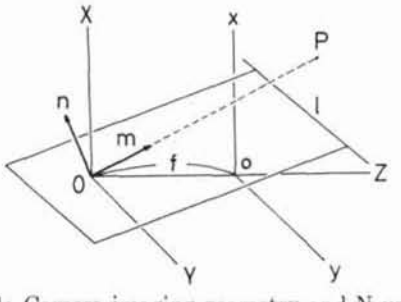


Fig. 1: Camera imaging geometry and N-vectors.

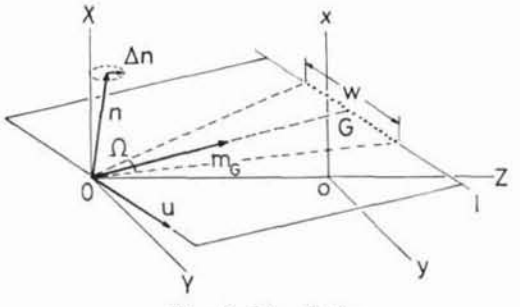


Fig. 2: Line fitting.

Here, $\boldsymbol{m}_G$ is the N-vector of the center point of the edge segment and γ is the *edge density* (the number of edge pixels per unit pixel length). We call the constant κ the *image resolution*.

The covariance matrix $V[\boldsymbol{n}]$ of the N-vector $\boldsymbol{n}$ of the line passing through two data pixels has the following form [5]:

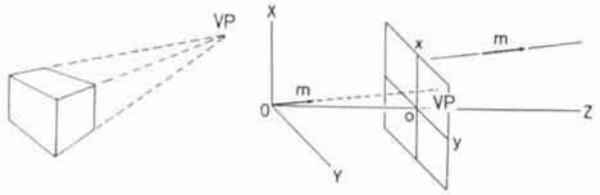
$$V[\boldsymbol{n}] \approx (\frac{\epsilon}{w})^2\boldsymbol{u}\boldsymbol{u}^T. \tag{5}$$

Here, $\boldsymbol{u}$ is the unit vector indicating the orientation of the line and w is the distance between the two pixels.

3. RELIABILITY OF VANISHING POINTS

If lines $\{l_\alpha\}$, $\alpha = 1, ..., N$, are projections of parallel lines in the scene, they are *concurrent* on the image plane, meeting at a common *vanishing point* (Fig. 3(a)); its N-vector $\boldsymbol{m}$ indicates the 3-D orientation of the corresponding lines (Fig. 3(b)) [3, 4, 5]. Let $\boldsymbol{n}_G$ be the N-vector of a hypothetical *center line* l_G of the N lines (Fig. 4). Let $\boldsymbol{m}_C$ be the unit vector orthogonal to both $\boldsymbol{n}_G$ and the N-vector $\boldsymbol{m}$ of the vanishing point; $\boldsymbol{m}_C$ can be identified with the N-vector of a point P_C on the center line l_G, which we call the *conjugate point*.

Suppose each line l_α is obtained by fitting a line to an edge segment of length w_α. Let $\boldsymbol{m}_{G\alpha}$ be the N-vector of its center point G_α. Let ϕ_α be the angle between $\boldsymbol{n}_G$ and $\boldsymbol{n}_\alpha$ and call it the *deviation angle* (from the hypothetical center line) of line l_α. Let θ_α be the angle between $\boldsymbol{m}_{G\alpha}$ and $\boldsymbol{m}$ and call it the *disparity* of the vanishing point from



(a) (b)

Fig. 3: (a) Vanishing point. (b) The N-vector $\boldsymbol{m}$ of the vanishing point indicates the 3-D orientation of the line.

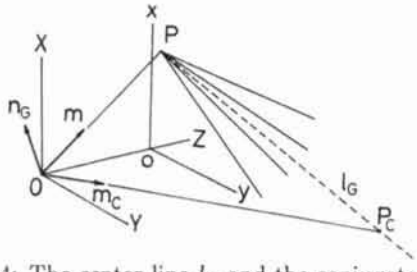


Fig. 4: The center line l_G and the conjugate point point P_G.

the center point of the αth edge segment. The covariance matrix $V[\boldsymbol{m}]$ of the N-vector of the estimated vanishing point is evaluated in the following form [5]:

$$V[\boldsymbol{m}] \approx \frac{6\kappa \boldsymbol{m}_C \boldsymbol{m}_C^T}{\sum_{\alpha=1}^{N} {w_\alpha}^3 \sin^2\phi_\alpha / \sin^2\theta_\alpha} \tag{6}$$

4. FOCAL LENGTH CALIBRATION

The camera model must be adjusted so that it agrees with the actual camera. The most important parameter is the focal length f. Its determination is based on the following fact [3, 5, 6]. Let $\boldsymbol{m}$ and $\boldsymbol{m}'$ be the N-vectors of the vanishing points of mutually orthogonal lines in the scene defined with respect to a tentative focal length f. The true focal length $\hat{f}$ is given by

$$\hat{f} = f\sqrt{-\frac{m_1 m_1' + m_2 m_2'}{m_3 m_3'}}. \tag{7}$$

Suppose the two vanishing points are detected as intersections of N and N' concurrent lines fitted to edge segments of lengths w_α and ${w_\alpha}'$, respectively. Let ϕ_α and ${\phi_\alpha}'$ be the deviation angles of the individual edge segments. If the center lines of the two sets of lines meet at the image origin o and if θ and θ' are the disparities of the two vanishing points from o, the variance $V[f]$ of the focal length f is given as follows [5]:

$$V[f] \approx \frac{3\kappa f^2}{2}\left(\frac{1/\cos^2\theta}{\sum_{\alpha=1}^{N} {w_\alpha}^3 \sin^2\phi_\alpha} + \frac{1/\cos^2\theta'}{\sum_{\alpha=1}^{N'} {w_\alpha}'^3 \sin^2{\phi_\alpha}'}\right) \tag{8}$$

Consider a square grid pattern placed in the scene (Fig. 5). Let r be the distance from O to the center of the grid pattern. The pattern consists of two sets of N (= an odd number) lines. Let l be the size of the pattern, and d the size of the individual grid squares. If l/f is small, it can be proved [5] that the minimum of eq. (8) is attained

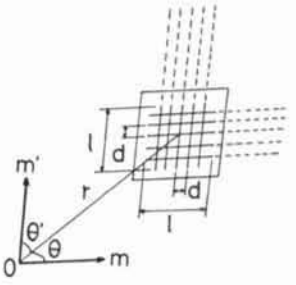
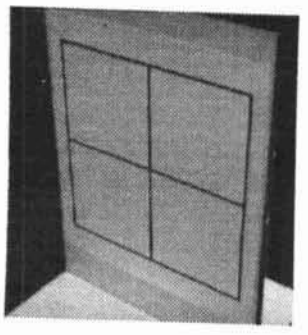


Fig. 5 The 3-D configuration of the grid pattern.



α	f_α	$V[f_\alpha]$	W_α
1	1027.0	110.1	0.021
2	1136.0	28.4	0.083
3	1149.9	18.6	0.126
4	1146.4	13.5	0.174
5	1092.4	21.4	0.110
6	1167.6	55.3	0.042
7	1147.3	19.3	0.122
8	1142.2	17.9	0.131
9	1092.5	20.0	0.117
10	1056.9	31.4	0.075

(a) (b)

Fig. 6: (a) A real image of a planar grid pattern. (b) Focal length f_α, variance $V[f_\alpha]$, and optimal weights W_α for each trial.

when $\theta = \theta' = 58.61...°$, for which the projections of the two sets of lines meet on the image plane at angle $\gamma = 111.85...°$ [5].

Let f_α, $\alpha = 1, ..., N$, be N estimates obtained from different images. It can be shown [5] that the optimally weighted average is given by

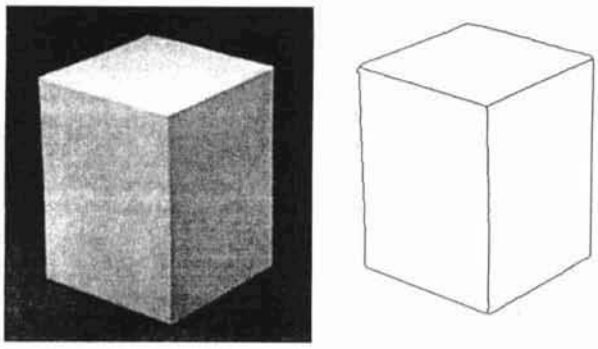
$$\bar{f} = \sum_{\alpha=1}^{N} W_\alpha f_\alpha, \quad W_\alpha = \frac{1}{V[f_\alpha]} \Big/ \sum_{\beta=1}^{N} \frac{1}{V[f_\alpha]}, \tag{9}$$

and its theoretical variance is

$$V[\bar{f}] = 1 \Big/ \sum_{\alpha=1}^{N} \frac{1}{V[f_\alpha]}. \tag{10}$$

The *confidence interval* can be computed by employing the Gaussian approximation: the true value f is inferred with $(100-a)\%$ confidence to be in the interval $[\bar{f} - \lambda_a\sqrt{V[\bar{f}]},\ \bar{f} + \lambda_a\sqrt{V[\bar{f}]}]$, where λ_a is the $a\%$ point of the standard normal distribution. However, this analysis indirectly involves the image resolution κ. If κ is difficult to estimate, we can do without it by using the Student statistic [5]: the $(100-a)\%$ confidence interval is $[\bar{f} - t_{a,N-1}s/\sqrt{N-1},\ \bar{f} + t_{a,N-1}s/\sqrt{N-1}]$, where $t_{a,N}$ is the $a\%$ point of the Student distribution with N degrees of freedom and $s = \sqrt{\sum_{\alpha=1}^{N} W_\alpha (f_\alpha - \bar{f})^2}$. The resolution κ can be chosen arbitrarily [5].

The planar grid pattern of Fig. 6(a) was placed in various locations and orientations in the scene. Fig. 6(b) shows the estimated focal lengths f_α, their variances $V[f_\alpha]$, and the corresponding optimal weights W_α. Trials 1 and 10 correspond to the case in which the pattern



(a) (b)

Fig. 7: (a) A real image of a rectangular box. (b) Detected edges.

is nearly parallel to the camera image plane. The corresponding variances are very large, meaning that the reliability is very low for such configurations. The optimally weighted average is $\bar{f} = 1125.0$ (pixels). The 95% confidence interval is [1099.6, 1150.4]. Since $l/r \approx 1/3$ and $d/r \approx 1/6$, the theoretical lower bound of $V[f]$ is $V[f] \geq 9.1...$, meaning that trial 4 (Fig. 6(a)) is very close to the optimal configuration; the grid lines intersect at angle 112° in Fig. 6(a).

5. ORTHOGONALITY FITTING

In many robotics applications, we encounter the problem of fitting an orthogonal frame to three axes computed from real data. For given unit vectors $\{\boldsymbol{m}_i\}$, $i = 1, 2, 3$, consider the computation of a right-handed orthonormal system $\{\boldsymbol{r}_i\}$ such that

$$\sum_{i=1}^{3} W_i \|\boldsymbol{r}_i - \boldsymbol{m}_i\|^2 \rightarrow \min, \qquad \sum_{i=1}^{3} W_i = 1. \tag{11}$$

The solution is analytically obtained by the method of *singular value decomposition*, the method of *polar decomposition*, or the method of *quaternion representation* [5].

Since $\{\boldsymbol{r}_i\}$ is a right-handed orthonormal system, the matrix $\boldsymbol{R}$ having them as columns is a rotation matrix, which we call the *best-fitting rotation*. If $\boldsymbol{R}$ is computed from image data, it may be perturbed into $\boldsymbol{R}'$. Since its columns still form an orthonormal system, the transformation from $\boldsymbol{R}$ to $\boldsymbol{R}'$ is a rotation by some angle $\Delta\Omega$ around some axis $\boldsymbol{l}$. Let $\Delta\boldsymbol{l} = \Delta\Omega\boldsymbol{l}$, and define *covariance matrix* of rotation $\boldsymbol{R}$ by

$$V[\boldsymbol{R}] = E[\Delta\boldsymbol{l}\Delta\boldsymbol{l}^{\top}]. \tag{12}$$

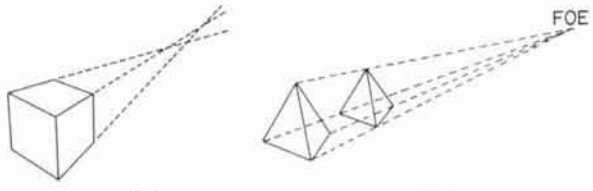
The covariance matrix $V[\boldsymbol{R}]$ of the best-fitting rotation $\boldsymbol{R}$ to $\{\boldsymbol{m}_i\}$, $i = 1, 2, 3$, is given as follows [5] ($(\cdot, \cdot)$ denotes the inner product of vectors):

$$V[\boldsymbol{R}] = \sum_{i,j=1}^{3} \frac{\sum_{k=1}^{3} W_k{}^2 (\boldsymbol{r}_i \times \boldsymbol{r}_k, V[\boldsymbol{m}_k](\boldsymbol{r}_j \times \boldsymbol{r}_k))}{(1 - W_i)(1 - W_j)} \boldsymbol{r}_i \boldsymbol{r}_j^{\top}. \tag{13}$$

Fig. 7(a) is a real image of a rectangular box. Fig. 7(b) shows detected edges. The N-vectors $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ of the three vanishing points are estimated by optimal least-squares [5]. Their covariance matrices $V[\boldsymbol{m}_1]$, $V[\boldsymbol{m}_2]$, and $V[\boldsymbol{m}_3]$ are given by eq. (6). The discrepancies of $\boldsymbol{m}_1$, $\boldsymbol{m}_2$, and $\boldsymbol{m}_3$ from the fitted orientations are 1.35°, 1.25°, and 0.97°, respectively. Eq. (13) gives the covariance matrix $V[\boldsymbol{R}]$, which tells us that the root-mean-square error $\Delta\Omega$ of the angle of error rotation (from the true frame, which we do not know) is 0.49°.

(a) (b)

Fig. 8: (a) Model matching for object recognition. (b) How should we group these edge segments together?



(a) (b)

Fig. 9: (a) How can we judge the concurrency of edge segments? (b) Focus of expansion.

6. GEOMETRIC TESTING

In recognizing objects in an image, edges are detected and candidate 3-D models stored in a database one are matched by one by changing the spatial position and orientation. Then, the one that best matches is chosen as the true object. It is natural to measure the degree of matching by the discrepancies of the detected edge segments from the supposed line segments of the model (Fig. 8(a)). In the past, various discrepancy measures were heuristically introduced [2, 7], but from eq. (4) we obtain the following criterion: Let $\boldsymbol{m}_{G\alpha}$, $\boldsymbol{u}_\alpha$, and w_α be the N-vector of the center point of an edge segment, its orientation, and its length, respectively, and let $\bar{\boldsymbol{n}}$ be the N-vector of a line. If

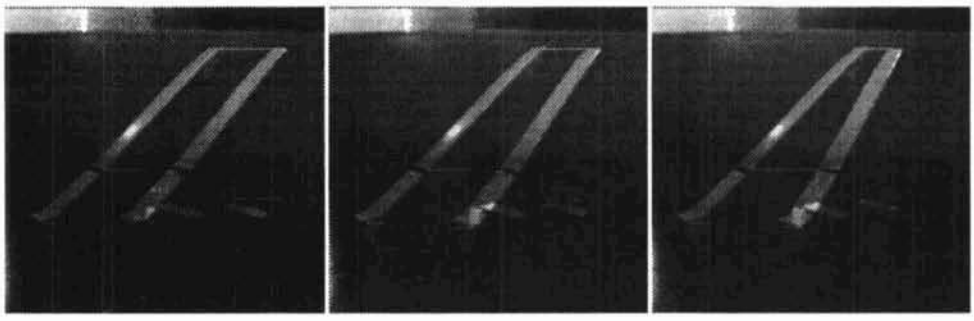
$$\frac{{w_\alpha}^3}{6\kappa}(\bar{\boldsymbol{n}}, \boldsymbol{u}_\alpha)^2 + \frac{2f^2 w_\alpha}{\kappa}(\bar{\boldsymbol{n}}, \boldsymbol{m}_{G\alpha})^2 > \chi^2_{a,2} \tag{14}$$

is satisfied, the edge segment cannot be regarded as lying on the line with confidence $(100 - a)\%$.

We want to know if multiple fragmented edge segments can be combined together and replaced by a single line $\bar{l}$ (Fig. 8(b)) [8]. Let $\boldsymbol{n}_\alpha$, $\boldsymbol{m}_{G\alpha}$, $\boldsymbol{u}_\alpha$, and w_α be the N-vector of the αth edge segment, the N-vector of its center point, its orientation, and its length, respectively, $\alpha = 1, ..., N$. It can be shown [5] that if the N-vector $\bar{\boldsymbol{n}}$ of the line $\bar{l}$ fitted to all the edge segments satisfies

$$\sum_{\alpha=1}^{N} \left(\frac{{w_\alpha}^3}{6}(\bar{\boldsymbol{n}}, \boldsymbol{u}_\alpha)^2 + 2f^2 w_\alpha (\bar{\boldsymbol{n}}, \boldsymbol{m}_{G\alpha})^2 \right) > \kappa \chi^2_{a,2N}, \tag{15}$$

the N edge segments are not regarded as collinear with confidence $(100 - a)\%$, where $\chi^2_{a,2}$ is the $a\%$ point of the χ^2-distribution with two degrees of freedom.



(a) (b) (c)

Fig. 10: Superimposed images of a translating stapler.

Vanishing points provide important clues to 3-D interpretation [9], but projections of parallel lines may not be concurrent due to noise (Fig. 9(a)). Let $\boldsymbol{n}_\alpha$, $\boldsymbol{m}_{G\alpha}$, and w_α be the N-vector of the αth edge segment, the N-vector of its center point, and its length, respectively, $\alpha = 1, ..., N$. It can be shown [5] that a point of N-vector $\boldsymbol{m}$ cannot be regarded as their vanishing point with confidence $(100 - a)\%$ if

$$\sum_{\alpha=1}^{N} {w_\alpha}^3 \left(1 - \frac{|\boldsymbol{m}, \boldsymbol{n}_\alpha, \boldsymbol{m}_{G\alpha}|^2}{1 - (\boldsymbol{m}, \boldsymbol{m}_{G\alpha})^2}\right) > 6\kappa\chi^2_{a,2N}. \qquad (16)$$

If points are rigidly translating in the scene (or the camera is translating relative to them), their trajectories are parallel in the scene, defining a common *focus of expansion* (Fig. 9(b)): its N-vector indicates the 3-D orientation of the corresponding 3-D translations [4]. Let $\boldsymbol{n}_\alpha$, $\boldsymbol{m}_{G\alpha}$, and w_α be the N-vector of the trajectory passing through the αth pair, the N-vector of its center point, and the distance between the two points, respectively, $\alpha = 1, ..., N$. It can be shown [5] that a point of N-vector $\boldsymbol{m}$ cannot be regarded as their focus of expansion with confidence $(100 - a)\%$ if

$$\sum_{\alpha=1}^{N} {w_\alpha}^2 \left(1 - \frac{|\boldsymbol{m}, \boldsymbol{n}_\alpha, \boldsymbol{m}_{G\alpha}|^2}{1 - (\boldsymbol{m}, \boldsymbol{m}_{G\alpha})^2}\right) > \epsilon^2\chi^2_{a,2N}. \qquad (17)$$

Fig. 10 shows superimpositions of two real images in which a stapler undergoes (a) a pure translation, (b) a translation and a small rotation, and (c) a translation and a large rotation. Seven feature points are chosen, and their trajectories are defined by connecting the corresponding positions. We hypothesize that all the trajectories are concurrent. The validity of this hypothesis depends on the image accuracy ϵ (in pixels) with which the feature points are detected. If $\epsilon \leq 2.5$, the hypothesis is accepted for (a) but rejected for (b) and (c) with 95% confidence. If $2.5 < \epsilon \leq 7$, the hypothesis is accepted for (a) and (b) but rejected for (c) with 95% confidence. If $\epsilon > 7$, the hypothesis is accepted for (a), (b), and (c) with 95%confidence.

7. CONCLUDING REMARKS

In this paper, the reliability of 3-D interpretations computed from images has been analyzed by applying the statistical theory of Kanatani [5]. First, the reliability of edge fitting was evaluated in terms of image noise characteristics. Then, the reliability of vanishing point estimation was deduced from the reliability of edge fitting. The result was applied to focal length calibration, and an optimal scheme was derived. We also discussed the reliability of fitting an orthogonal frame to three orientations obtained by sensing.

Finally, statistical criteria were derived for model matching, testing edge groupings, vanishing points, focuses of expansion, and vanishing lines. For these problems, it has been customary to introduce ad hoc parameters to be thresholded, but all the criteria given here do not involve any ad hoc parameters; they are build on a rigorous statistical basis.

References

[1] B. Brillault-O'Mahony, New method for vanishing point detection, *CVGIP: Image Understanding*, **54** (1991), 289–300.

[2] R. A. Brooks, *Model-Based Computer Vision*, UMI Research Press, Ann Arbor, MI, U.S.A., 1984

[3] K. Kanatani, *Group-Theoretical Methods in Image Understanding*, Springer, Berlin, F.R.G., 1990.

[4] K. Kanatani, Computational Projective geometry, *CVGIP: Image Understanding*, **54** (1991), 333–448.

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

[6] K. Kanatani and Y. Onodera, Anatomy of camera calibration using vanishing points, *IEICE Trans.*, **74**-10 (1991), 3369–3378.

[7] D. G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic, Boston, MA, U.S.A., 1985.

[8] R. Weiss and M. Boldt, Geometric grouping applied to straight lines, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Miami Beach, FL, U.S.A., June 1986, pp. 656–661.

[9] R. Weiss, H. Nakatani and E. M. Riseman, An error analysis for surface orientation from vanishing points, *IEEE Trans. Pattern Anal. Machine Intell.*, **12** (1990), 1179–1185.