Statistical Analysis of Focal-Length Calibration Using Vanishing Points

Kenichi Kanatani

Abstract—**Camera focal-length calibration is analyzed by applying a statistical error model of edge fitting in digital images. The reliability of the focal length is computed from the reliability of the vanishing points, and the reliability of the vanishing points is computed from the reliability of the edges to which lines are fitted. A planar grid pattern is placed in the scene for this purpose, and the 3-D configuration that maximizes the reliability is theoretically derived. Then, multiple data are fused by optimally weighted averaging. A real image example is given.**

I. Introduction

Visual sensing plays a central role in controlling robots that recognize environments, and 3-D information is easily obtained by detecting "vanishing points" in robotics environments where many parallel lines usually exist [7], [8], [10]. Vanishing points also play a key role in autonomously navigating land vehicles by vision [12], [13], [16], [18].

Computing 3-D interpretations of images requires the imaging geometry of the camera, and such parameters as the 3-D location of the center of the lens, the 3-D orientation of the optical axis, and the focal length must be accurately calibrated. Again, the detection of vanishing points plays an essential role [2], [5], [11], [21].

Thus, accurate detection of vanishing points is one of the most essential components of 3-D visual sensing, and various techniques have been proposed [1], [3], [7], [17], [22]. However, the reliability of such techniques is usually tested empirically by using synthetic and real data. To further the progress of robotics applications, we need a *theory* to evaluate the reliability of computation based on a realistic model of the statistical behavior of noise and error, correctly estimating the *confidence bound* and theoretically proving the *optimality* of the method. This is particularly necessary when we attempt to enhance system performance by fusing multiple data obtained by multiple sensors (*sensor fusion*), because reliable data must be given large weights while unreliable data must be given small weights.

Recently, a comprehensive mathematical theory that serves this purpose was proposed by Kanatani [9]. In this paper, we apply his theory to camera focal length calibration. The reliability of the focal length is computed from the reliability of the vanishing points, and the reliability of the vanishing points is computed from the reliability of the edges to which lines are fitted. We derive a calibration procedure using a planar grid pattern, and the 3-D configuration that maximizes the reliability of the computed focal length is theoretically derived.

Manuscript received November 13, 1991; revised March 9, 1992.

The author is with the Department of Computer Science, Gunma University, Kiryu, Gunma 376, Japan.

IEEE Log Number 9203533.

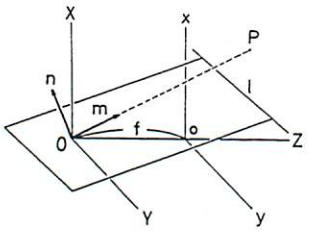


Fig. 1. Imaging geometry and N-vectors of a point and a line.

Then, multiple data are fused by optimally weighted averaging. A real image example is also given.

II. Perspective Projection and N-Vectors

Assume the following camera model. The camera is associated with an XYZ coordinate system with its origin O at the center of the lens and the Z axis along the optical axis (Fig. 1). The plane $Z = f$ is identified with the image plane, on which an xy coordinate system is defined so that the x and y axes are parallel to the X and Y axes, respectively. Let us call the origin O the *viewpoint* and the constant f the *focal length.*

A point (x, y) on the image plane is represented by the unit vector $\boldsymbol{m}$ indicating the orientation of the ray starting from the viewpoint O and passing through that point. A line $Ax + By + C = 0$ on the image plane is represented by the unit surface normal $\boldsymbol{n}$ to the plane passing through the viewpoint O and intersecting the image plane along that line (Fig. 1.) Their components are given by

$$\boldsymbol{m} = \pm N\left[\begin{pmatrix} x \\ y \\ f \end{pmatrix}\right], \qquad \boldsymbol{n} = \pm N\left[\begin{pmatrix} A \\ B \\ C/f \end{pmatrix}\right] \tag{1}$$

where $N[\cdot]$ denotes normalization into a unit vector. Let us call $\boldsymbol{m}$ and $\boldsymbol{n}$ the N-vectors of the point and the line [8].

A point (X, Y, Z) in the scene is perspectively projected onto a point (x, y) on the image plane given by

$$x = f\frac{X}{Z}, \qquad y = f\frac{Y}{Z}. \tag{2}$$

We define the N-vector of a point in the scene to be the N-vector of its projection on the image plane, and the N-vector of a line in the scene to be the N-vector of its projection on the image plane. In order to avoid the confusion of whether we are referring to a point in the scene or its projection on the image plane, we call a point in the scene a *space point* and a point on the image plane an *image point.* Similarly, we call a line in the scene a *space line* and a line on the image plane an *image line.*

If $\boldsymbol{m}$ and $\boldsymbol{n}$ are the N-vectors of an image point P and an image line l, respectively, image point P is *on* image line l, or image line l *passes through* image point P, if and only if

$$(\boldsymbol{m}, \boldsymbol{n}) = 0 \tag{3}$$

where $(\cdot, \cdot)$ denotes the inner product of vectors. Then we say that image point P and image line l are *incident* to each other and call (3) the *incidence equation.*

An image point that is on two distinct image lines is called their *intersection*, and an image line that passes through two distinct image points is called their *join.* If $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are the N-vectors of two distinct image lines, the N-vector $\boldsymbol{m}$ of their intersection is given by

$$\boldsymbol{m} = \pm N[\boldsymbol{n}_1 \times \boldsymbol{n}_2] \tag{4}$$

because $\boldsymbol{m}$ must satisfy the incidence equation (3) for both image lines: $(\boldsymbol{m}, \boldsymbol{n}_1) = 0$ and $(\boldsymbol{m}, \boldsymbol{n}_2) = 0$. Dually, if $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are the N-vectors of two distinct image points, the N-vector $\boldsymbol{n}$ of their join is given by

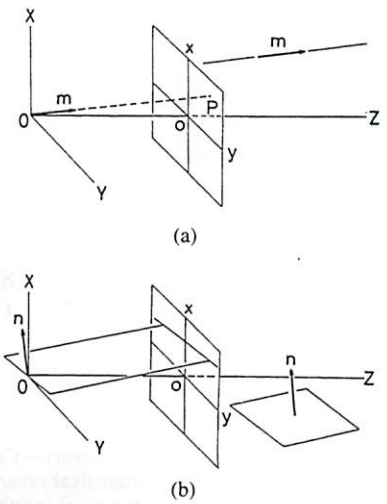


Fig. 2. (a) The vanishing point of a space line. (b) The vanishing line of a planar surface in the scene.

$$\boldsymbol{n} = \pm N[\boldsymbol{m}_1 \times \boldsymbol{m}_2] \tag{5}$$

because $\boldsymbol{n}$ must satisfy the incidence equation (3) for both image points: $(\boldsymbol{m}_1, \boldsymbol{n}) = 0$ and $(\boldsymbol{m}_2, \boldsymbol{n}) = 0$.

The use of N-vectors for representing image points and image lines is equivalent to using *homogeneous coordinates.* Although homogeneous coordinates can be multiplied by any nonzero number, computational problems arise if they are too large or too small. So, it is convenient to normalize them into a unit vector. Kanatani [8] reformulated projective geometry from this viewpoint. Rewriting the relationships of projective geometry as "computational procedures," he called his formalism *computational projective geometry.* In the following, we adopt his formalism, regarding unit vector $\boldsymbol{m}$ whose Z component is 0 as the N-vector of an *ideal point* (a *point at infinity*) and $\boldsymbol{n} = (0, 0, \pm 1)$ as the N vector of the *ideal line* (the *line at infinity*).

As is well known, projections of parallel space lines meet at a common "vanishing point" on the image plane. Formally, the *vanishing point* of a space line is the limit of the projection of a point that moves along the space line indefinitely in one direction (both directions define the same vanishing point). From Fig. 2(a), it is easy to confirm the following theorem:

Theorem 1: A space line extending along unit vector $\boldsymbol{m}$ has, when projected, a vanishing point of N-vector $\pm\boldsymbol{m}$.

Since the vanishing point is determined by the 3-D orientation of the space line alone, irrespective of its location in the scene, we see that:

Corollary 1: Projections of parallel space lines intersect at a common vanishing point.

As is also well known, projections of planar surfaces that are parallel in the scene define a common "vanishing line." Formally, the *vanishing line* of a planar surface in the scene is the set of all the vanishing points of space lines lying on it. From Fig. 2(b), it is easy to confirm the following theorem:

Theorem 2: A planar surface of unit surface normal $\boldsymbol{n}$ has, when projected, a vanishing line of N-vector $\pm\boldsymbol{n}$.

Since the vanishing line is determined by the 3-D orientation of the planar surface alone, irrespective of its location in the scene, we see that:

Corollary 2: Projections of planar surfaces that are parallel in the scene define a common vanishing line.

Thus, if a vanishing point is detected, its N-vector indicates the 3-D orientation of the corresponding space line, and if a vanishing line is detected, its N-vector indicates the surface normal to the corresponding planar surface. This fact plays an essential role in 3-D scene analysis of various robotics applications.

III. Statistical Model of Edge Fitting

In conventional image processing, edges are detected by the Hough transform or an edge operator, and a set of edge pixels is obtained by applying thresholding and thinning processes. Then, lines are fitted to the edge pixels, say, by least squares. Since vanishing points are estimated as intersections of these fitted lines, the reliability of the vanishing points depends on the reliability of the edges. Thus, we need a statistical model for the error behaviors of edge fitting.

Let $\boldsymbol{n}$ be the N-vector of an image line fitted to edge pixels in the absence of noise. In the presence of noise, each edge pixel is displaced. Let $\boldsymbol{n}' = \boldsymbol{n} + \Delta\boldsymbol{n}$ be the N-vector of the image line fitted to the displaced edge pixels. Since the noise behavior is random, the error $\Delta\boldsymbol{n}$ is regarded as a vector-valued random variable. We assume that $\Delta\boldsymbol{n}$ is sufficiently small as compared with $\boldsymbol{n}$. Then $\Delta\boldsymbol{n}$ is orthogonal to $\boldsymbol{n}$ to a first approximation, and $\boldsymbol{n}' = \boldsymbol{n} + \Delta\boldsymbol{n}$ is also a unit vector to a first approximation. We define the *covariance matrix* of $\boldsymbol{n}$ by

$$V[\boldsymbol{n}] = E[\Delta\boldsymbol{n}\Delta\boldsymbol{n}^\top] \tag{6}$$

where $E[\cdot]$ means expectation and $\top$ denotes transpose. From this definition, we have the following observations:

- The covariance matrix $V[\boldsymbol{n}]$ is symmetric and positive semidefinite.
- The covariance matrix $V[\boldsymbol{n}]$ is singular with $\boldsymbol{n}$ itself as the unit eigenvector for eigenvalue 0: $V[\boldsymbol{n}]\boldsymbol{n} = \boldsymbol{0}$.
- If ${\sigma_1}^2$, ${\sigma_2}^2$, and 0 ($\sigma_1 \geq \sigma_2 > 0$) are the three eigenvalues and if $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{m}\}$ is the corresponding orthonormal system of eigenvectors, the covariance matrix $V[\boldsymbol{n}]$ has the following "spectral decomposition" (see [10]):

$$V[\boldsymbol{n}] = {\sigma_1}^2\boldsymbol{u}_1\boldsymbol{u}_1^\top + {\sigma_2}^2\boldsymbol{u}_2\boldsymbol{u}_2^\top (+0^2\boldsymbol{m}\boldsymbol{m}^\top). \tag{7}$$

- The root mean square of the orthogonal projection of noise $\Delta\boldsymbol{n}$ onto orientation $\boldsymbol{l}$ (unit vector) takes its maximum for $\boldsymbol{l} = \boldsymbol{u}_1$ and its minimum for $\boldsymbol{l} = \boldsymbol{u}_2$. The maximum and minimum values are σ_1 and σ_2, respectively.
- The root mean square magnitude $\|\Delta\boldsymbol{n}\|$ is $\sqrt{\mathrm{tr}V[\boldsymbol{n}]} = \sqrt{{\sigma_1}^2 + {\sigma_2}^2}$.

In intuitive terms, error $\Delta\boldsymbol{n}$ is most likely to occur in orientation $\boldsymbol{u}_1$ (equal to the unit eigenvector of $V[\boldsymbol{n}]$ for the largest eigenvalue ${\sigma_1}^2$) and least likely to occur in orientation $\boldsymbol{u}_2$ (equal to the unit eigenvector of $V[\boldsymbol{n}]$ for the second largest eigenvalue ${\sigma_2}^2$). The magnitude $\|\Delta\boldsymbol{n}\|$ is σ_1 in orientation $\boldsymbol{u}_1$ and σ_2 in orientation $\boldsymbol{u}_2$ in the sense of root mean square.

The theoretical expression of this covariance matrix was derived from a realistic model of noise by Kanatani [9], who also conducted simulations to confirm the result. Since the derivation requires a lengthy analysis, we omit the details and list the final form (Fig. 3):

Theorem 3: The covariance matrix of the N-vector $\boldsymbol{n}$ of an edge segment of length w in orientation $\boldsymbol{u}$ is given by

$$V[\boldsymbol{n}] = \frac{6\kappa}{w^3}\boldsymbol{u}\boldsymbol{u}^\top + \frac{\kappa}{2f^2w}\boldsymbol{m}_G\boldsymbol{m}_G^\top \tag{8}$$

where $\boldsymbol{m}_G$ is the N-vector of the center point of the edge and κ is the image resolution.

The length w is measured in pixels. If $\boldsymbol{m}_a$ and $\boldsymbol{m}_b$ are the N-vectors of the endpoints of the edge segment, the vectors $\boldsymbol{u}$ and $\boldsymbol{m}_G$

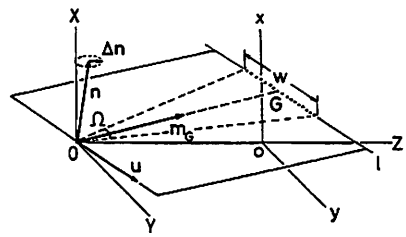


Fig. 3. Line fitting to an edge segment.

are formally defined by

$$\boldsymbol{u} = \pm N[\boldsymbol{m}_a - \boldsymbol{m}_b], \qquad \boldsymbol{m}_G = \pm N[\boldsymbol{m}_a + \boldsymbol{m}_b] \tag{9}$$

and the three vectors $\{\boldsymbol{u}, \boldsymbol{m}_G, \boldsymbol{n}\}$ form an orthonormal system of unit eigenvectors of $V[\boldsymbol{n}]$ for eigenvalues $6\kappa/w^3$, $\kappa/2f^2w$, and 0, respectively. The *image resolution* κ is defined by

$$\kappa = \frac{\epsilon^2}{\gamma} \tag{10}$$

where ϵ is the *image accuracy* defined as the root mean square of the displacement of each edge pixel, while γ is the *edge density* defined as the number of edge pixels per unit pixel length. If the image is ideal, the image accuracy is at most one pixel and edge pixels are aligned at one pixel intervals, so $\kappa \approx 1$.

If the length w of the edge segment is very small as compared with the focal length f, the right-hand side of (8) is dominated by the first term, and (8) is approximated by

$$V[\boldsymbol{n}] \approx \frac{6\kappa}{w^3}\boldsymbol{u}\boldsymbol{u}^\top. \tag{11}$$

It has been experimentally confirmed that this is indeed a very good approximation [9]. From this, we observe that

- The error in the N-vector $\boldsymbol{n}$ almost always occurs in the orientation $\boldsymbol{u}$ of the fitted line.
- The error is approximately proportional to $\kappa^{1/2}$ of the image resolution κ (hence, proportional to the image accuracy ϵ and to $\gamma^{-1/2}$ of the edge density γ).
- The error is approximately proportional to $w^{-3/2}$ of the length w of the edge segment.

Here, we are assuming that an image line is fitted to a dense sequence of edge pixels by least squares, but if an image line is defined as the join of two data points, its error behavior is somewhat different. It can be shown [9] that the covariance matrix $V[\boldsymbol{n}]$ of the N-vector $\boldsymbol{n}$ of the join has the form

$$V[\boldsymbol{n}] \approx (\frac{\epsilon}{w})^2\boldsymbol{u}\boldsymbol{u}^T \tag{12}$$

where $\boldsymbol{u}$ is defined by the first of (9) for the N-vectors $\boldsymbol{m}_a$ and $\boldsymbol{m}_b$ of the two points, and w is the distance (in pixels) between them.

IV. Optimal Estimation of Vanishing Points

If image lines $\{l_\alpha\}$, $\alpha = 1, \cdots, N$, are projections of parallel space lines, they are concurrent on the image plane. Their common intersection is the "vanishing point" whose N-vector $\boldsymbol{m}$ indicates the 3-D orientation of the corresponding space lines (Theorem 1). If $\{\boldsymbol{n}_\alpha\}$, $\alpha = 1, \cdots, N$, are the N-vectors of the image lines, the incidence equation $(\boldsymbol{m}, \boldsymbol{n}_\alpha) = 0$ (see (3)) must hold for $\alpha = 1, \cdots, N$. Hence, $\boldsymbol{m}$ is robustly computed by the least squares optimization

$$\sum_{\alpha=1}^{N} W_\alpha (\boldsymbol{m}, \boldsymbol{n}_\alpha)^2 \to \min. \tag{13}$$

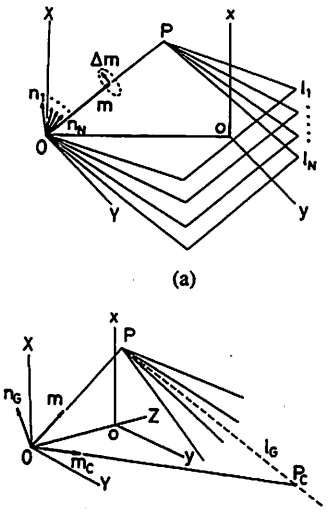


Fig. 4. (a) The common intersection of concurrent lines. (b) The center line l_G of concurrent lines. Point P_G is conjugate to the vanishing point P on line l_G.

(See Fig. 4(a).) The weights W_α should be determined so that reliable data are given large weights while unreliable data are given small weights. It can be shown [9] that the *optimal weights* in the sense of "maximum likelihood estimation" are

$$W_\alpha = \frac{1}{(\boldsymbol{m}, V[\boldsymbol{n}_\alpha]\boldsymbol{m})}. \tag{14}$$

It is easily seen that reliable data have small covariance matrices, and are thereby assigned large weights, while unreliable data have large covariance matrices and are thereby assigned small weights. Since

$$\sum_{\alpha=1}^{N} W_\alpha (\boldsymbol{m}, \boldsymbol{n}_\alpha)^2 = \left(\boldsymbol{m}, \left(\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{n}_\alpha \boldsymbol{n}_\alpha^\top\right)\boldsymbol{m}\right) \tag{15}$$

is a quadratic form in unit vector $\boldsymbol{m}$, it is minimized by the unit eigenvector of the *moment matrix*

$$\boldsymbol{N} = \sum_{\alpha=1}^{N} W_\alpha \boldsymbol{n}_\alpha \boldsymbol{n}_\alpha^\top \tag{16}$$

for the smallest eigenvalue.

Note that the optimal weights contain the N-vector $\boldsymbol{m}$ we want to compute. This difficulty can be avoided by computing approximately optimal weights by using an estimate of $\boldsymbol{m}$, say the solution for constant weights. It can be shown [9] that if the estimate of $\boldsymbol{m}$ has an error $\Delta\boldsymbol{m}$, the resulting solution $\boldsymbol{m}$ has an error of only $O(\|\Delta\boldsymbol{m}\|^2)$.

If there is no noise, the computed $\boldsymbol{m}$ is exact. If the N-vector $\boldsymbol{n}_\alpha$ of each line is perturbed, the resulting N-vector $\boldsymbol{m}$ is perturbed, say, by $\Delta\boldsymbol{m}$. It can be shown [9] that its covariance matrix $V[\boldsymbol{m}] = E[\Delta\boldsymbol{m}\Delta\boldsymbol{m}^\top]$ is given as follows:

Theorem 4: Let $\{\boldsymbol{m}, \boldsymbol{u}, \boldsymbol{v}\}$ be the orthonormal system of the eigenvectors of the optimal moment matrix $\boldsymbol{N}$ for eigenvalues 0, λ_u, and λ_v, respectively. The covariance matrix $V[\boldsymbol{m}]$ of the N-vector of the optimally estimated common intersection is

$$V[\boldsymbol{m}] = \frac{\boldsymbol{u}\boldsymbol{u}^\top}{\lambda_u} + \frac{\boldsymbol{v}\boldsymbol{v}^\top}{\lambda_v}. \tag{17}$$

Let $\boldsymbol{n}_G$ be the unit eigenvector of the moment matrix $\boldsymbol{N}$ of (16) for the largest eigenvalue. Vector $\boldsymbol{n}_G$ can be regarded as the N-vector of a hypothetical *center line* l_G of the N lines (Fig. 4(b)). Since the three eigenvectors form an orthonormal system, the unit eigenvector $\boldsymbol{m}_C$ for the second largest eigenvalue equals $\pm\boldsymbol{m} \times \boldsymbol{n}_G$. The vector $\boldsymbol{m}_C$ is orthogonal to both $\boldsymbol{n}_G$ and $\boldsymbol{m}$ and hence can be identified with the N vector of the point P_C "conjugate" [8] to the vanishing point P on the center line l_G.

From (16), the eigenvalue of $\boldsymbol{N}$ for $\boldsymbol{n}_G$ is given by

$$(\boldsymbol{n}_G, \boldsymbol{N}\boldsymbol{n}_G) = \sum_{\alpha=1}^{N} W_\alpha (\boldsymbol{n}_G, \boldsymbol{n}_\alpha)^2. \tag{18}$$

The eigenvalue for $\boldsymbol{m}_C$ is

$$(\boldsymbol{m}_C, \boldsymbol{N}\boldsymbol{m}_C) = \sum_{\alpha=1}^{N} W_\alpha (\boldsymbol{m} \times \boldsymbol{n}_G, \boldsymbol{n}_\alpha)^2 = \sum_{\alpha=1}^{N} W_\alpha |\boldsymbol{m}, \boldsymbol{n}_G, \boldsymbol{n}_\alpha|^2 \tag{19}$$

where $|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}|(= (\boldsymbol{a} \times \boldsymbol{b}, \boldsymbol{c}) = (\boldsymbol{b} \times \boldsymbol{c}, \boldsymbol{a}) = (\boldsymbol{c} \times \boldsymbol{a}, \boldsymbol{b}))$ denotes the scalar triple product of vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. From Theorem 4, the covariance matrix $V[\boldsymbol{m}]$ of the N-vector $\boldsymbol{m}$ of the vanishing point is given by

$$V[\boldsymbol{m}] = \frac{\boldsymbol{m}_C \boldsymbol{m}_C^\top}{\sum_{\alpha=1}^{N} W_\alpha |\boldsymbol{m}, \boldsymbol{n}_G, \boldsymbol{n}_\alpha|^2} + \frac{\boldsymbol{n}_G \boldsymbol{n}_G^\top}{\sum_{\alpha=1}^{N} W_\alpha (\boldsymbol{n}_G, \boldsymbol{n}_\alpha)^2}. \tag{20}$$

If the separations among the lines are small, the right-hand side is dominated by the first term. If ϕ_α is the angle between $\boldsymbol{n}_G$ and $\boldsymbol{n}_\alpha$, we have

$$|\boldsymbol{m}, \boldsymbol{n}_G, \boldsymbol{n}_\alpha|^2 = (\boldsymbol{m}, \boldsymbol{n}_G \times \boldsymbol{n}_\alpha)^2 = \sin^2 \phi_\alpha \tag{21}$$

because $\boldsymbol{m}$ is parallel to $\boldsymbol{n}_G \times \boldsymbol{n}_\alpha$. Let us call ϕ_α the *deviation angle* (from the hypothetical center line). If the covariance matrix $V[\boldsymbol{n}_\alpha]$ of each line is approximated in the form of (11), we have

$$V[\boldsymbol{n}_\alpha] \approx \frac{6\kappa}{{w_\alpha}^2} \boldsymbol{u}_\alpha \boldsymbol{u}_\alpha^\top = \frac{6\kappa}{{w_\alpha}^2} (\boldsymbol{m}_{G\alpha} \times \boldsymbol{n}_\alpha)(\boldsymbol{m}_{G\alpha} \times \boldsymbol{n}_\alpha)^\top \tag{22}$$

where $\boldsymbol{u}_\alpha$ is the orientation of the αth edge segment and $\boldsymbol{m}_{G\alpha}$ is the N-vector of its center point. The optimal weight W_α is then approximated by

$$W_\alpha = \frac{1}{(\boldsymbol{m}, V[\boldsymbol{n}_\alpha]\boldsymbol{m})} \approx \frac{{w_\alpha}^2}{6\kappa |\boldsymbol{m}, \boldsymbol{m}_{G\alpha}, \boldsymbol{n}_\alpha|^2}. \tag{23}$$

If θ_α is the angle between $\boldsymbol{m}_{G\alpha}$ and $\boldsymbol{m}$, we have

$$|\boldsymbol{m}, \boldsymbol{m}_{G\alpha}, \boldsymbol{n}_\alpha|^2 = (\boldsymbol{m} \times \boldsymbol{m}_{G\alpha}, \boldsymbol{n}_\alpha)^2 = \sin^2 \theta_\alpha \tag{24}$$

because $\boldsymbol{n}_\alpha$ is parallel to $\boldsymbol{m} \times \boldsymbol{m}_{G\alpha}$. The angle θ_α indicates the *disparity* of the vanishing point from the center point of the αth edge segment.

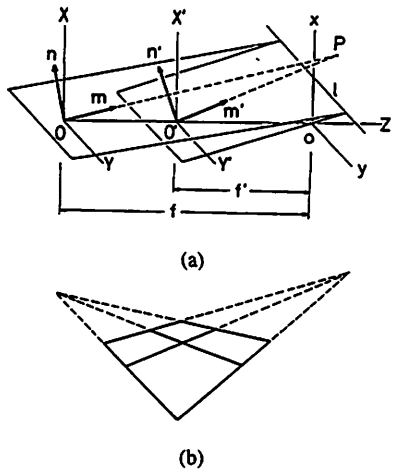
If these are substituted into (20), we obtain the approximation

$$V[\boldsymbol{m}] \approx \frac{6\kappa \boldsymbol{m}_C \boldsymbol{m}_C^\top}{\sum_{\alpha=1}^{N} {w_\alpha}^2 \sin^2 \phi_\alpha / \sin^2 \theta_\alpha}. \tag{25}$$

From this, we have the following observations:

- The error of the vanishing point is most likely to occur along the center line.
- The error of the vanishing point is approximately proportional to $\kappa^{1/2}$ of the image resolution κ.
- The error of the vanishing point is approximately proportional to ${w_\alpha}^{-3/2}$ of the lengths w_α of the individual edge segments.
- The error of the vanishing point is approximately proportional to $1/\sin\phi_\alpha$ of the deviation angles ϕ_α of the individual edge segments from the center line.
- The error of the vanishing point is approximately proportional to $\sin\theta_\alpha$ of the disparities θ_α of the vanishing point from the center points of the individual edge segments.

See [9] for numerical experiments that confirm these observations.



(a)

(b)

Fig. 5. (a) N-vectors of image points and image lines are altered if the focal length f is altered. (b) Vanishing points of two sets of parallel space lines mutually orthogonal in the scene.

V. Determination of the Focal Length

When we analyze images in terms of N-vectors, we are assuming a camera model (Fig. 1). This camera model can be hypothetical: it need not correspond to the actual camera *as long as no 3-D interpretation is involved.* For example, when we compute vanishing points as intersections of concurrent image lines, or vanishing lines as image lines passing through collinear image points, we can arbitrarily assume the camera model.

However, if we want to infer 3-D relationships such as 3-D orientations of space lines and surfaces and their orthogonality in the scene, the resulting 3-D interpretations are valid only when the camera model exactly agrees with the imaging geometry. Hence, the camera model must be adjusted so that it agrees with the actual camera. This process is called *camera calibration*, and is one of the most important problems in robotics applications [2], [5], [6], [11], [14], [15], [19]–[21].

The following proposition is easily obtained from the definition of the N-vector (see (1) and Fig. 5(a)):

Proposition 1: Let $\boldsymbol{m} = (m_1, m_2, m_3)^\top$ and $\boldsymbol{n} = (n_1, n_2, n_3)^\top$ be the N-vectors of an image point and an image line, respectively, defined with respect to focal length f, and let $\boldsymbol{m}' = (m_1', m_2', m_3')^\top$ and $\boldsymbol{n}' = (n_1', n_2', n_3')^\top$ be the N-vectors of the same image point and image line, respectively, defined with respect to a different focal length f'. Then

$$\boldsymbol{m}' = \pm N\left[\begin{pmatrix} m_1 \\ m_2 \\ (f'/f)m_3 \end{pmatrix}\right], \qquad \boldsymbol{n}' = \pm N\left[\begin{pmatrix} n_1 \\ n_2 \\ (f/f')n_3 \end{pmatrix}\right]. \tag{26}$$

Proof: The image coordinates (x, y) are given by

$$x = f\frac{m_1}{m_3}, \qquad y = f\frac{m_2}{m_3}. \tag{27}$$

If the equation of the image line is $Ax + By + C = 0$, we have

$$A = kn_1, \qquad B = kn_2, \qquad C = kfn_3 \tag{28}$$

where k is an arbitrary nonzero constant. From (1), we see that

$$\boldsymbol{m}' = \pm N\left[\begin{pmatrix} x \\ y \\ f' \end{pmatrix}\right] = \pm N\left[\begin{pmatrix} fm_1/m_3 \\ fm_2/m_3 \\ f' \end{pmatrix}\right] = \pm N\left[\begin{pmatrix} m_1 \\ m_2 \\ (f'/f)m_3 \end{pmatrix}\right] \tag{29}$$

$$\boldsymbol{n}' = \pm N\left[\begin{pmatrix} A \\ B \\ C/f' \end{pmatrix}\right] = \pm N\left[\begin{pmatrix} kn_1 \\ kn_2 \\ kfn_3/f' \end{pmatrix}\right] = \pm N\left[\begin{pmatrix} n_1 \\ n_2 \\ (f/f')n_3 \end{pmatrix}\right]. \tag{30}$$

■

From this, we obtain the following results:

Proposition 2: If $\boldsymbol{m} = (m_1, m_2, m_3)^\top$ and $\boldsymbol{m}' = (m_1', m_2', m_3')^\top$ are the N-vectors of the vanishing points of mutually orthogonal space lines defined with respect to a tentative focal length f, the true focal length $\hat{f}$ is given by

$$\hat{f} = f\sqrt{-\frac{m_1m_1' + m_2m_2'}{m_3m_3'}}. \tag{31}$$

Proof: Let $\hat{\boldsymbol{m}}$ and $\hat{\boldsymbol{m}}'$ be the N-vectors corresponding to $\boldsymbol{m}$ and $\boldsymbol{m}'$, respectively, expressed with respect to the true focal length $\hat{f}$. Since they indicate the 3-D orientations of the corresponding space lines (Theorem 1), they must be orthogonal to each other. From Proposition 1, $\hat{\boldsymbol{m}}$ and $\hat{\boldsymbol{m}}'$ are orthogonal if and only if

$$m_1m_1' + m_2m_2' + \frac{\hat{f}^2}{f^2}m_3m_3' = 0 \tag{32}$$

from which follows (31). ■

Proposition 3: If $\boldsymbol{n} = (n_1, n_2, n_3)^\top$ and $\boldsymbol{n}' = (n_1', n_2', n_3')^\top$ are the N-vectors of the vanishing lines of planar surfaces mutually orthogonal in the scene defined with respect to a tentative focal length f, the true focal length $\hat{f}$ is given by

$$\hat{f} = f\sqrt{-\frac{n_3n_3'}{n_1n_1' + n_2n_2'}}. \tag{33}$$

Proof: Let $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{n}}'$ be the N-vectors corresponding to $\boldsymbol{n}$ and $\boldsymbol{n}'$, respectively, expressed with respect to the true focal length $\hat{f}$. Since they indicate the unit surface normals to the corresponding planar surfaces (Theorem 2), they must be orthogonal to each other. From Proposition 1, $\hat{\boldsymbol{n}}$ and $\hat{\boldsymbol{n}}'$ are orthogonal if and only if

$$n_1n_1' + n_2n_2' + \frac{f^2}{\hat{f}^2}n_3n_3' = 0 \tag{34}$$

from which follows (33). ■

Proposition 2 provides the following simple procedure for calibrating the focal length (Fig. 5(b)):

1) Take an image of a rectangular grid pattern placed in the scene.
2) Fit lines to the grid line and compute the N-vectors of the vanishing points in the two directions with respect to a tentative focal length f.
3) Compute the true focal length $\hat{f}$ by (32).

This procedure is very simple, and essentially the same idea has been suggested by many researchers [2], [7], [11], [21]. However, (31) does not necessarily yield a reliable value if the vanishing points are located very far from the image origin o; as the vanishing points move away from o, both the numerator and the denominator approach 0. This means that the two vanishing points must be close to the image origin o, but as one approaches o, the other goes away from it, and the image lines that define the vanishing points become short or close to each other, thereby reducing the accuracy of the detected vanishing points. Thus, the reliability of calibration critically depends on the 3-D configuration of the space lines.

In the following, we analyze various effects that affect the reliability of this computation and compute an "optimal estimate" of the focal length and its "confidence interval."

VI. Reliability of Focal Length

If the tentative focal length f happens to be the exact value, the left-hand side of (31) should be precisely f. However, if the N-vectors $\boldsymbol{m}$ and $\boldsymbol{m}'$ are computed from a real image, they may contain errors, and the left-hand side of (31) may predict $f + \Delta f$. The variance $V[f] = E[(\Delta f)^2]$ is given as follows:

Theorem 5: If errors in the N-vectors $\boldsymbol{m} = (m_1, m_2, m_3)^\top$ and $\boldsymbol{m}' = (m_1', m_2', m_3')^\top$ are independent, the variance $V[f]$ is given by

$$V[f] = \frac{f^2}{4} \frac{(\boldsymbol{m}', V[\boldsymbol{m}]\boldsymbol{m}') + (\boldsymbol{m}, V[\boldsymbol{m}']\boldsymbol{m})}{(m_3 m_3')^2} \tag{35}$$

where $V[\boldsymbol{m}]$ and $V[\boldsymbol{m}']$ are the covariance matrices of N-vectors $\boldsymbol{m}$ and $\boldsymbol{m}'$, respectively.

Proof: Equation (31) was derived from (32). If m_i and m_i' are perturbed by Δm_i and $\Delta m_i'$, $i = 1, 2, 3$, respectively, we have to a first approximation

$$\Delta m_1 m_1' + m_1 \Delta m_1' + \Delta m_2 m_2' + m_2 \Delta m_2' + \frac{2\hat{f}\Delta\hat{f}}{f^2} m_3 m_3' + \frac{\hat{f}}{f} \Delta m_3 m_3' + \frac{\hat{f}}{f} m_3 \Delta m_3' = 0. \tag{36}$$

If we let $\hat{f} = f$, we obtain (37), given at the bottom of this page. Since errors in $\boldsymbol{m}$ and $\boldsymbol{m}'$ are independent, the variance is given by

$$E[(\Delta\hat{f})^2] = \frac{f^2}{4} \frac{\boldsymbol{m}'^\top E[\Delta\boldsymbol{m}\Delta\boldsymbol{m}^\top]\boldsymbol{m}' + \boldsymbol{m}^\top E[\Delta\boldsymbol{m}'\Delta\boldsymbol{m}'^\top]\boldsymbol{m}}{(m_3 m_3')^2} = \frac{f^2}{4} \frac{(\boldsymbol{m}', V[\boldsymbol{m}]\boldsymbol{m}') + (\boldsymbol{m}, V[\boldsymbol{m}']\boldsymbol{m})}{(m_3 m_3')^2}. \tag{38}$$

■

If the two vanishing points are detected as intersections of N and N' concurrent image lines fitted to edge segments, substitution of (25) into (35) yields the approximation given by (39) at the bottom of this page, where $\boldsymbol{m}_C$ and $\boldsymbol{m}_C'$ are the N-vectors of the points "conjugate" to the vanishing points on the "center lines" of the two sets of image lines, w_α and w_α' are the lengths of the edge segments, ϕ_α and ϕ_α' are their "deviation angles," and θ_α and θ_α' are the "disparities" of the vanishing points from the center points of the individual edge segments.

If θ and θ' are the disparities of the two vanishing points from the image origin o, we have

$$(m_3 m_3')^2 = \cos^2\theta \cos^2\theta'. \tag{40}$$

Suppose the center lines of the two sets of image lines meet at the image origin o. Then we have the following:

Proposition 4:

$$(\boldsymbol{m}_C, \boldsymbol{m}') = \frac{\cos\theta'}{\sin\theta}, \qquad (\boldsymbol{m}_C', \boldsymbol{m}) = \frac{\cos\theta}{\sin\theta'}. \tag{41}$$

Proof: In spherical coordinates, vectors $\boldsymbol{m}$ and $\boldsymbol{m}'$ have components

$$\boldsymbol{m} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}, \qquad \boldsymbol{m}' = \begin{pmatrix} \sin\theta'\cos\phi' \\ \sin\theta'\sin\phi' \\ \cos\theta' \end{pmatrix}. \tag{42}$$

Since $\boldsymbol{m}$ and $\boldsymbol{m}'$ are mutually orthogonal, we have

$$(\boldsymbol{m}, \boldsymbol{m}') = \sin\theta\sin\theta'(\cos\phi\cos\phi' + \sin\phi\sin\phi') + \cos\theta\cos\theta' = \sin\theta\sin\theta'\cos(\phi - \phi') + \cos\theta\cos\theta' = 0 \tag{43}$$

or

$$\cos(\phi - \phi') = -\frac{\cos\theta\cos\theta'}{\sin\theta\sin\theta'}. \tag{44}$$

The N-vectors $\boldsymbol{m}_C$ and $\boldsymbol{m}_C'$ are expressed in spherical coordinates in the form

$$\boldsymbol{m}_C = \begin{pmatrix} -\cos\theta\cos\phi \\ -\cos\theta\sin\phi \\ \sin\theta \end{pmatrix}, \qquad \boldsymbol{m}_C' = \begin{pmatrix} -\cos\theta'\cos\phi' \\ -\cos\theta'\sin\phi' \\ \sin\theta' \end{pmatrix}. \tag{45}$$

Hence,

$$(\boldsymbol{m}, \boldsymbol{m}_C') = -\sin\theta\cos\theta'(\cos\phi\cos\phi' + \sin\phi\sin\phi') + \cos\theta\sin\theta' = -\sin\theta\cos\theta'\cos(\phi - \phi') + \cos\theta\sin\theta' = \frac{\cos\theta\cos^2\theta'}{\sin\theta'} + \cos\theta\sin\theta' = \frac{\cos\theta}{\sin\theta'}. \tag{46}$$

The first of (41) is obtained similarly. ■

If the center points of the edge segments are located close to the image origin o, the disparities θ_α and θ_α' are all approximated by θ and θ', respectively. Hence, (39) is further approximated by

$$V[f] \approx \frac{3\kappa f^2}{2} \left(\frac{1/\cos^2\theta}{\sum_{\alpha=1}^{N} {w_\alpha}^3 \sin^2\phi_\alpha} + \frac{1/\cos^2\theta'}{\sum_{\alpha=1}^{N'} {w_\alpha}'^3 \sin^2\phi_\alpha'} \right). \tag{47}$$

From this, we observe that

- The error of f increases as the error of the detected vanishing points, which is approximately proportional to $\kappa^{1/2}$ of the image resolution κ, to ${w_\alpha}^{-3/2}$ and ${w_\alpha'}^{-3/2}$ of the individual lengths w_α and w_α' of the edge segments, and to $1/\sin\phi_\alpha$ and $1/\sin\phi_\alpha$ of their deviation angles ϕ_α and ϕ_α' from their center lines.
- The error of f is approximately proportional to $1/\cos\theta$ and $1/\cos\theta'$ of the disparities θ and θ' of the vanishing points from the image origin.

$$\Delta\hat{f} = -\frac{f}{2} \frac{\Delta m_1 m_1' + m_1 \Delta m_1' + \Delta m_2 m_2' + m_2 \Delta m_2' + \Delta m_3 m_3' + m_3 \Delta m_3'}{m_3 m_3'} = -\frac{f}{2} \frac{(\boldsymbol{m}', \Delta\boldsymbol{m}) + (\boldsymbol{m}, \Delta\boldsymbol{m}')}{m_3 m_3'}. \tag{37}$$

$$V[f] \approx \frac{3\kappa f^2}{2(m_3 m_3')^2} \left(\frac{(\boldsymbol{m}_C, \boldsymbol{m}')^2}{\sum_{\alpha=1}^{N} {w_\alpha}^3 \sin^2\phi_\alpha / \sin^2\theta_\alpha} + \frac{(\boldsymbol{m}_C', \boldsymbol{m})^2}{\sum_{\alpha=1}^{N'} {w_\alpha}'^3 \sin^2\phi_\alpha' / \sin^2\theta_\alpha'} \right). \tag{39}$$

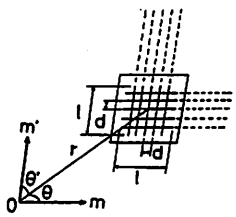


Fig. 6. The 3-D configuration of the grid pattern.

VII. Optimal Configuration of Space Lines

Consider a square grid pattern placed in the scene (Fig. 6). Let r be the distance from the viewpoint O to the center of the grid pattern. The pattern consists of two sets of N (equal to an odd number) lines. Let l be the size of the pattern, and d the size of the individual grid squares. If l/r is small, we have the approximation

$$w_\alpha \approx f\frac{l}{r}\sin\theta. \tag{48}$$

The length w'_α is approximated similarly. Let ϕ_0 and ϕ'_0 be the deviation angles between neighboring grid lines in the two directions, respectively. If Nd/r is small, we have the approximation

$$\sum_{\alpha=1}^{N}\sin^2\phi_\alpha \approx 2\sum_{k=1}^{(N-1)/2}k^2\sin^2\phi_0 = \frac{N(N^2-1)}{12}\sin^2\phi_0. \tag{49}$$

The summation $\sum_{\alpha=1}^{N}\sin^2\phi'_\alpha$ is approximated similarly. Let $\boldsymbol{k} = (0,0,1)^\top$. We assume that $\boldsymbol{m}$ and $\boldsymbol{m}'$ are oriented so that $|\boldsymbol{m},\boldsymbol{m}',\boldsymbol{k}| > 0$. The N-vector of the central grid line extending in the direction of $\boldsymbol{m}$ is

$$\boldsymbol{n}_0 = N[\boldsymbol{m}\times\boldsymbol{k}] = \frac{\boldsymbol{m}\times\boldsymbol{k}}{\sin\theta}. \tag{50}$$

If $d/r \ll 1$, the N-vector of the neighboring grid line is

$$\boldsymbol{n}_1 = N[\boldsymbol{m}\times(r\boldsymbol{k}+d\boldsymbol{m}')] = c(\boldsymbol{m}\times\boldsymbol{k}+\frac{d}{r}\boldsymbol{m}\times\boldsymbol{m}') \tag{51}$$

where

$$c = \frac{1}{\|\boldsymbol{m}\times\boldsymbol{k}+(d/r)\boldsymbol{m}\times\boldsymbol{m}'\|} = \frac{1}{\sin\theta}+O(\frac{d}{r}). \tag{52}$$

From the definition of the deviation angle ϕ_0, we have $\sin\phi_0 = \|\boldsymbol{n}_1\times\boldsymbol{n}_0\|$. From (50) and (51), we have

$$\boldsymbol{n}_1\times\boldsymbol{n}_0 = \frac{cd}{r}\frac{(\boldsymbol{m}\times\boldsymbol{m}')\times(\boldsymbol{m}\times\boldsymbol{k})}{\sin\theta} = \frac{cd}{r}\frac{|\boldsymbol{m},\boldsymbol{m}',\boldsymbol{k}|}{\sin\theta}\boldsymbol{m}. \tag{53}$$

From (52), we obtain

$$\sin\phi_0 = \frac{d}{r}\frac{|\boldsymbol{m},\boldsymbol{m}',\boldsymbol{k}|}{\sin^2\theta}+O((\frac{d}{r})^2). \tag{54}$$

We also obtain a similar expression for $\sin\phi'_0$.

Three vectors $\{\boldsymbol{m}, \boldsymbol{m}', \boldsymbol{m}\times\boldsymbol{m}'\}$ form an orthonormal set. Since $\boldsymbol{k}$ makes angles θ and θ' from $\boldsymbol{m}$ and $\boldsymbol{m}'$, respectively, vector $\boldsymbol{k}$ can be expressed in the form

$$\boldsymbol{k} = \cos\theta\boldsymbol{m}+\cos\theta'\boldsymbol{m}'+C\boldsymbol{m}\times\boldsymbol{m}' \tag{55}$$

so $|\boldsymbol{m},\boldsymbol{m}',\boldsymbol{k}| = (\boldsymbol{m}\times\boldsymbol{m}',\boldsymbol{k}) = C$. Since $\boldsymbol{k}$ is a unit vector, we have $C^2 = 1-\cos^2\theta-\cos^2\theta'$. Hence,

$$|\boldsymbol{m},\boldsymbol{m}',\boldsymbol{k}|^2 = 1-\cos^2\theta-\cos^2\theta'. \tag{56}$$

Thus, we obtain

$$V[f] \approx \frac{18\kappa r^5}{N(N^2-1)fd^2l^3(1-\cos^2\theta-\cos^2\theta')^2}\left(\frac{\sin\theta}{\cos^2\theta}+\frac{\sin\theta'}{\cos^2\theta'}\right). \tag{57}$$

Proposition 5: The variance $V[f]$ is minimized when

$$\theta = \theta' = \sin^{-1}\sqrt{\frac{3+\sqrt{33}}{12}} \approx 58.6^\circ \tag{58}$$

for which the two sets of grid lines make angle

$$\gamma = \pi-\cos^{-1}\frac{\sqrt{33}-5}{2} \approx 111.9^\circ. \tag{59}$$

Proof: The right-hand side of (57) is written as const. $\times F(\sin\theta,\sin\theta')$ if we define the function $F(x,y)$ by

$$F(x,y) = \frac{1}{x^2+y^2-1}\left(\frac{1}{1-x^2}+\frac{1}{1-y^2}\right). \tag{60}$$

This function attains its minimum for $0\le x$, $0\le y$, and $x^2+y^2\ge 1$ when $x=y=\sqrt{(3+\sqrt{33})/12}$. If γ is the angle made by the two grid lines at the image origin, we have $\cos\gamma = \cos(\phi-\phi') = -\cos^2\theta/\sin^2\theta$ (see (44)). Thus, we obtain the assertion. ■

In intuitive terms, the planar surface must be inclined to make a large angle from the image plane so that the effect of foreshortening is large. However, too much inclination compresses the resulting image of the grid pattern, decreasing the lengths of the edge segments and the separations between them, thereby decreasing the accuracy of the detected vanishing points. The optimal balance is attained at the values given in Proposition 5.

VIII. Optimal Estimation of Focal Length

A well-known strategy to enhance the accuracy of a measurement is repeating it and averaging the results. However, a simple average can be greatly distorted by a small number of very disturbed values (so-called *outliers*). Hence, a weighted average should be taken so that unreliable values are given small weights while reliable values are given large weights. Let f_α, $\alpha = 1,\cdots,N$, be estimates of the focal length f obtained from N different images. Consider the weighted average in the form

$$\bar{f} = \sum_{\alpha=1}^{N}W_\alpha f_\alpha, \qquad \sum_{\alpha=1}^{N}W_\alpha = 1. \tag{61}$$

The accuracy of $\bar{f}$ is maximized by minimizing the variance $V[\bar{f}]$. Let us call the weights W_α that minimize $V[\bar{f}]$ the *optimal weights*, and the resulting estimate $\bar{f}$ the *optimal estimate*. The following is well known in statistics [4]:

Proposition 6: If f_α is independent with positive variances $V[f_\alpha]$, $\alpha =,\cdots,N$, the optimal weights W_α are given by

$$W_\alpha = \frac{1}{V[f_\alpha]}\Big/\sum_{\beta=1}^{N}\frac{1}{V[f_\beta]}. \tag{62}$$

Proof: If each estimate is independent, the variance of $\bar{f}$ is

$$V[\bar{f}] = \sum_{\alpha=1}^{N}W_\alpha{}^2V[f_\alpha]. \tag{63}$$

If we introduce a Lagrange multiplier for the constraint $\sum_{\alpha=1}^{N}W_\alpha = 1$, it is easy to see that (63) is minimized by (62). ■

Substituting (62) into (63), we obtain the following:

Proposition 7: The variance of the optimal estimate is given by

$$V[\bar{f}] = 1\Big/\sum_{\alpha=1}^{N}\frac{1}{V[f_\alpha]}. \tag{64}$$

Once the variance $V[f]$ is evaluated, the *confidence interval* can be computed by employing the Gaussian approximation. Namely, if $\bar{f}$ is the computed optimal estimate for the true value f, the statistic $u =$

$(f-\bar{f})/\sqrt{V[\bar{f}]}$ obeys the standard normal distribution (of mean 0 and variance 1), so we have $-\lambda_a \leq u \leq \lambda_a$ with $(100-a)\%$ confidence, where λ_a is the $a\%$ point of the standard normal distribution (e.g., $\lambda_5 = 1.96$). Hence, with $(100-a)\%$ confidence, the true value f is inferred to be in the interval

$$\left[\bar{f} - \lambda_a\sqrt{V[\bar{f}]}, \bar{f} + \lambda_a\sqrt{V[\bar{f}]}\right]. \tag{65}$$

In order to compute the optimal estimate $\bar{f}$ by (61), the optimal weights W_α must be computed by (62), which involves the variance $V[f_\alpha]$ of each estimate f_α. The variance $V[f_\alpha]$ is computed by (35), which involves the covariance matrices $V[\boldsymbol{m}]$ and $V[\boldsymbol{m}']$ of the vanishing points. They are computed by (20), which involves optimal weights for individual image lines. They are computed by (14), which involve the covariance matrices $V[\boldsymbol{n}_\alpha]$ of the individual image lines. They are computed by (8), which involves the image resolution κ.

Note that the optimal weights W_α of (62) are not affected by multiplication of $V[f_\alpha]$ by a constant. It is also easily confirmed that the optimally estimated vanishing points and the optimal estimates f_α are not affected by the image resolution κ. It follows that the optimal estimate $\bar{f}$ is not affected by κ, for which an arbitrary value can be assumed.

However, the image resolution κ directly affects all the covariance matrices and variances, and thereby the confidence interval. If the image resolution κ is difficult to estimate *a priori*, we can do without its value by the following statistical technique:

Proposition 8: Let

$$s = \sqrt{\sum_{\alpha=1}^{N} W_\alpha (f_\alpha - \bar{f})^2}. \tag{66}$$

If f is the true value of the focal length, then

$$t = \sqrt{N-1}\frac{\bar{f}-f}{s} \tag{67}$$

is a statistic obeying the Student distribution with $N-1$ degrees of freedom.

Proof: Since $(f_\alpha - \bar{f})/\sqrt{V[f_\alpha]}$ obeys the standard normal distribution (of mean 0 and variance 1), the statistic

$$\chi^2 = \sum_{\alpha=1}^{N} \frac{(f_\alpha - \bar{f})^2}{V[f_\alpha]} = \sum_{\alpha=1}^{N} W_\alpha (f_\alpha - \bar{f})^2 \sum_{\beta=1}^{N} \frac{1}{V[f_\beta]} \tag{68}$$

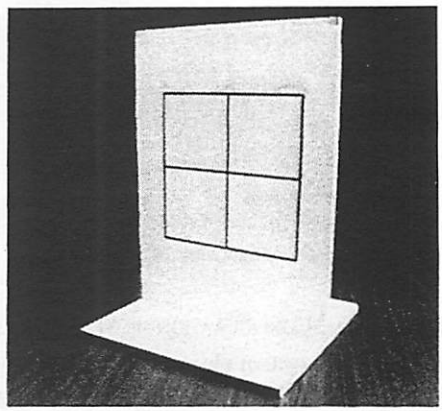
obeys the χ^2 distribution with $N-1$ degrees of freedom, where Proposition 7 has been used. On the other hand, the statistic $u = (\bar{f}-f)/\sqrt{V[\bar{f}]}$ obeys the standard normal distribution, so the statistic

$$t = \sqrt{N-1}\frac{u}{\chi} = \frac{\sqrt{N-1}(\bar{f}-f)\sqrt{\sum_{\alpha=1}^{N} 1/V[f_\alpha]}}{\sqrt{\sum_{\beta=1}^{N} W_\alpha (f_\beta - \bar{f})^2}\sqrt{\sum_{\gamma=1}^{N} 1/V[f_\gamma]}}$$
$$= \sqrt{N-1}\frac{\bar{f}-f}{s} \tag{69}$$

obeys the Student distribution with $N-1$ degrees of freedom, where Proposition 6 has been used. ■

The statistic t is no longer affected by multiplication of $V[f_\alpha]$ by a constant, so the image resolution κ can be chosen arbitrarily. We now have $-t_{a,N} \leq t \leq t_{a,N}$ with $(100-a)\%$ confidence, where $t_{a,N}$ is the $a\%$ point of the Student distribution with N degrees of freedom (e.g., $t_{5,9} = 2.262$). Hence, with $(100-a)\%$ confidence, the true value f is inferred to be in the interval

$$\left[\bar{f} - t_{a,N-1}\frac{s}{\sqrt{N-1}}, \bar{f} + t_{a,N-1}\frac{s}{\sqrt{N-1}}\right]. \tag{70}$$



(a)

α	f_α	$V[f_\alpha]$	W_α
1	206.942	1091.712	0.000303
2	522.662	24.635	0.013423
3	551.018	9.621	0.034369
4	575.322	1.057	0.312835
5	588.870	0.733	0.451114
6	665.852	3.679	0.089880
7	675.818	5.501	0.060110
8	680.580	10.368	0.031893
9	722.831	56.581	0.005844
10	925.895	1447.349	0.000228

(b)

Fig. 7. (a) A real image of a planar board. A square grid pattern is drawn on it. (b) Focal length f_α, variance $V[f_\alpha]$, and optimal weight W_α for ten trials.

Example: The planar grid pattern of Fig. 7(a) was placed in various locations and orientations in the scene. Edges were detected, and lines were fitted by least squares. Then, the vanishing points were computed by the optimal least-squares method described in Section IV, and the true focal length was estimated by the method described in Section V. At the same time, the covariance matrix of each estimate was computed.

Fig. 7(b) shows the estimated focal lengths f_α, their variances $V[f_\alpha]$, and the corresponding optimal weights W_α for ten different positions and orientations of the pattern. The variances $V[f_\alpha]$ were computed by assuming $\kappa = 1$. Trials 1 and 10 correspond to the case where the board is nearly parallel to the image plane. The corresponding variances are very large, meaning that the reliability is very low for such configurations. This is obvious because foreshoretening effect is very small if the pattern is nearly parallel to the image plane. As a result, the computed estimates are given very small weights. Applying the optimally weighted average of (61), we obtain the optimal estimate of f to be $\bar{f} = 598.257$. The 95% confidence interval given by (70) is [568.979, 627.534].

In each trial, the square grid pattern was projected onto a large area in the image, so roughly $l/r \approx 2/3$ and $d/r \approx 1/3$. For these values, the minimum of $V[f]$ given by Proposition 5 is 0.5223.... Hence, trial 5 is close to the optimal configuration. Indeed, we confirm that the grid lines intersect at an angle very close to the optimal value.

In computing the optimal weights by (14), the N-vector $\boldsymbol{m}$ was approximated by the value obtained by using uniform weights. This process could be iterated. However, such iterations cause a change of $\boldsymbol{m}$ of about $0.0001°$ in orientation, resulting in changes of each f_α by about 0.001% and of $\bar{f}$ by about 0.0001%. The upper and lower bounds of the 95% confidence interval change only by about 0.03%. Hence, no further iterations are necessary.

IX. Concluding Remarks

In this paper, we have presented a theory of evaluating the reliability of camera focal length calibration. Although only the focal length was studied here, many other factors exist that must be considered in real situations. For example, straight lines may not be projected to straight lines due to optical distortion of the lens (*aberration*). So, an appropriate mapping must be applied to remove such distortions (*geometric correction*). Another important parameter is the *aspect ratio*—the ratio of the horizontal scale to the vertical scale. Then the image origin must be located. This is theoretically possible by image analysis alone [2], [10], [11], but the procedure is very sensitive to noise. The most reliable way is probably the use of a "mechanical" method. For example, we can physically locate the center of the lens [5]. If a face of the camera body parallel to the optical axis is identified, we can move the camera along the assumed optical axis. Then, image origin is located by detecting the "focus of expansion" on the image plane [14].

In the past, many researchers have tried to estimate all the parameters by, say, "across-the-board fitting":

1) Set, in the scene, multiple reference points whose scene coordinates are known.
2) Locate their images on the video display.
3) Construct a parameterized camera model that incorporates all conceivable factors (lens aberration, focal length, aspect ratio, raster scanning distortion, 3-D camera position, etc.), and express the image coordinates of the reference points in terms of the model parameters.
4) Adjust the parameters by minimizing, say by least squares, the discrepancies between the image coordinates of the observed reference points and their predicted locations.

The quantity to be minimized is usually a complicated nonlinear function of the model parameters, so numerical iterations are necessary. If lens and image distortions are neglected, the equations can be "linearized" by introducing auxiliary variables ("homogeneous coordinates"), and *in appearance* the optimization reduces to solving simultaneous linear equations. However, what is actually minimized is not clear if such artificial linearization is involved. The least-squares minimization makes sense only when error behaviors are well understood.

On the other hand, if parameters that attain the minimum are found, this does not necessarily mean that each parameter is reliable. Suppose, for example, quantity J is to be minimized but it is not so very sensitive to one parameter, say α, as compared with another parameter, say β, near the optimum: $|\partial J/\partial \alpha| \ll |\partial J/\partial \beta|$. Then, the estimate of α may be largely disturbed to compensate for the error in β. This typically occurs when parameters of different geometric origins, such as the focal length and the 3-D camera position, are incorporated at the same time.

This is because the mechanism of estimation is different from parameter to parameter. As pointed out earlier, the focal length cannot be detected accurately unless the effect of foreshortening is strong, because assuming different focal lengths does not affect the resulting 3-D interpretation very much if foreshortening is not apparent. Hence, if a planar pattern is used, it must be placed so that it makes a large angle with the image plane, whereas if the 3-D camera position is to be computed by using the same pattern, a reliable estimate is obtained when it is placed nearly parallel to the image plane.

Since each parameter has a different configuration in maximizing the reliability, the calibration procedure should be decomposed into separate modules corresponding to individual parameters. Then each module should be designed so that its reliability is maximized, and the reliability of the resulting estimate must be evaluated in quantitative terms. One of the main purposes of this paper is to emphasize the effectiveness of this approach.

Acknowledgment

The author thanks M. Brady, A. Blake, and A. Zisserman of the University of Oxford for helpful comments during his stay at Oxford. He also thanks K. Urasawa of Gunma University for doing the real image experiments.

References

[1] B. Brillault-O'Mahony, "New method for vanishing point detection," *CVGIP: Image Understanding*, vol. 54, pp. 289–300, 1991.
[2] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *Int. J. Comput. Vision*, vol. 4, pp. 127–140, 1990.
[3] R. T. Collins and R. S. Weiss, "Vanishing point calculation as a statistical data," in *Proc. 3rd Int. Conf. Comput. Vision* (Osaka, Japan), Dec. 1990, pp. 400–403.
[4] H. Cramér, *The Elements of Probability Theory and Some of Its Applications*. Stockholm: Almqvist & Wiksell/Gebers, 1955.
[5] T. Echigo, "A camera calibration technique using three sets of parallel lines," *Machine Vision Appl.*, vol. 3, pp. 159–167, 1990.
[6] W. I. Grosky and L. A. Tamburino, "A unified approach to the linear camera calibration problem," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 663–671, 1990.
[7] K. Kanatani, *Group-Theoretical Methods in Image Understanding*. Berlin: Springer, 1990.
[8] ——, "Computational projective geometry," *CVGIP: Image Understanding*, vol. 54, pp. 333–348, 1991.
[9] ——, "Statistical analysis of geometric computation," *CVGIP: Image Understanding*, submitted for publication.
[10] ——, *Geometric Computation for Machine Vision*. Oxford, U.K.: Oxford University Press, 1993.
[11] K. Kanatani and Y. Onodera, "Anatomy of camera calibration using vanishing points," *IEICE Trans. Info. Syst.*, vol. 74, no. 10, pp. 3369–3378, 1991.
[12] K. Kanatani and K. Watanabe, "Reconstruction of 3-D road geometry from images for autonomous land vehicles," *IEEE Trans. Robotics Automat.*, vol. 6, pp. 127–132, 1990.
[13] ——, "Road shape reconstruction by local flatness approximation," *Advanced Robotics*, vol. 6, 1992.
[14] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high-accuracy 3-D machine vision metrology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 713–720, 1988.
[15] R. K. Lenz and R. Y. Tsai, "Calibrating a Cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 916–928, 1989.
[16] S.-P. Liou and R. C. Jain, "Road following using vanishing points," *Comput. Vision, Graphics, Image Process.*, vol. 39, pp. 116–130, 1987.
[17] M. J. Magee and J. K. Aggarwal, "Determining vanishing points from perspective images," *Comput. Vision, Graphics, Image Process.*, vol. 26, pp. 256–267, 1984.
[18] D. G. Morgenthaler, S. Hennessy, and D. DeMenthon, "Range-video fusion and comparison of inverse perspective algorithms in static images," *IEEE Trans. Syst. Man Cybern.*, vol. 20, pp. 1301–1312, 1990.
[19] M. A. Penna, "Camera calibration: A quick and easy way to determine the scale factor," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 1240–1245, 1991.
[20] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robotics Automat.*, vol. 5, pp. 345–358, 1989.
[21] L.-L. Wang and W.-H. Tsai, "Camera calibration by vanishing lines for 3-D computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 370–376, 1991.
[22] R. Weiss, H. Nakatani, and M. Riseman, "An error analysis for surface orientation from vanishing points," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1179–1185, 1991.