Automatic Singularity Test for Motion Analysis by an Information Criterion

Kenichi Kanatani*

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

Abstract. The structure-from-motion algorithm from two views fails if the object is a planar surface or the camera motion is a pure rotation. This paper presents a new scheme for automatically detecting these anomalies without using any knowledge about the noise in the images. This judgment does not involve any empirically adjustable thresholds, either. The basic principle of our scheme is to choose a model that has "higher predicting capability" measured by the *geometric information criterion* (*geometric AIC*).

1 Introduction

The *structure-from-motion* algorithm from two views has been studied by many researchers in the past [3, 5, 11]. However, the algorithm fails if all the feature points are coplanar in the scene, because a planar surface is a degenerate *critical surface* that gives rise to ambiguity of 3-D interpretation [4, 9]. Hence, a different algorithm is necessary for a planar surface. The planar surface algorithm has also been studied by many researchers in the past [7, 8, 10]. However, both the general and the planar surface algorithms assume that the translation of the camera is not zero; if the camera motion is a pure rotation around the center of the lens, no 3-D information can be obtained. If follows that the structure-from-motion analysis must take the following steps:

1. Test if the translation is 0—we call this the *rotation test*. If so, output a warning message and stop.
2. Test if the object is a planar surface—we call this the *planarity test*. If so, apply the planar surface algorithm.
3. Else, apply the general algorithm.

In practice, however, the images have noise, and the general algorithm applied in the presence of noise produces some (unreliable) solution even when the camera motion is a pure rotation or the object is a planar surface. In the past, the above tests have been done by introducing an ad-hoc criterion and an empirically adjustable threshold. For example, based on the fact that the smallest eigenvalue

* This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research B (No. 07458067) and the Okawa Institute of Information and Telecommunication.

of a matrix involved in the analysis should be a multiple root in the absence of noise if the object is a planar surface, the object is judged as planar if its smallest two eigenvalues are close enough. However, ***how can we determine the threshold for such a judgement?***

- First of all, we need to know the accuracy of the detected feature points, because the threshold should be set high if the accuracy is low while it should be set low if the accuracy is high. However, the accuracy of the feature points detected by an image processing operation depends on not only the operation itself but also various imaging conditions such as the focus, the resolution, the lighting, the shape of the object, and its orientation and position. Hence, the accuracy is different from image to image, and it is almost impossible to predict it in advance.
- Even if the accuracy can be predicted, what can be obtained is a *probability* of the noise, since the noise is a random phenomenon. Following the statistical theory of testing of hypotheses, we can set the threshold in such a way that the probability (the *significance level*) that a planar surface is judged as non-planar is $a\%$. However, how can we set that significance level? The result of the judgment differs if the significance level is set different.

In the past, little attention has been paid to this problem. Sometimes, thresholds are adjusted so that the experiment in question works well. In this paper, we present a theoretical framework for doing the planarity and rotation tests *without knowing the magnitude of the image noise* and *without introducing any empirically adjustable thresholds*. What makes this possible is the introduction of the *geometric information criterion* (*geometric AIC*), which measures the *predictive capacity* of the model in statistical terms.

2 3-D Reconstruction from Two Views

Define an XYZ camera coordinate system in such a way that the origin O is at the center of the lens and the Z-axis is along the optical axis. If the distance between the origin O and the photo-cell array is taken as the unit of length, the image plane can be identified with $Z = 1$; the imaging geometry can be regarded as perspective projection onto it. Define an xy image coordinate system on the image plane $Z = 1$ in such a way that the origin o is on the optical axis and the x- and y-axes are parallel to the X- and Y-axes, respectively. Then, a point (x, y) on the image plane can be represented by vector $\boldsymbol{x} = (x, y, 1)^\top$ (the superscript $\top$ designates transpose).

Suppose the camera is rotated around the center of the lens by $\boldsymbol{R}$ and translated by $\boldsymbol{h}$ (Fig. 1(a)). We call $\{\boldsymbol{h}, \boldsymbol{R}\}$ the *motion parameters*. If we define the $X'Y'Z'$ camera coordinate system and the $x'y'$ image coordinate system with respect to the camera position after this motion, a point (x', y') on the image plane, which can be represented by vector $\boldsymbol{x}' = (x', y', 1)$ with respect to the $X'Y'Z'$ coordinate system, can be represented by vector $\boldsymbol{R}\boldsymbol{x}'$ with respect to the XYZ coordinate system. If follows that vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be images of

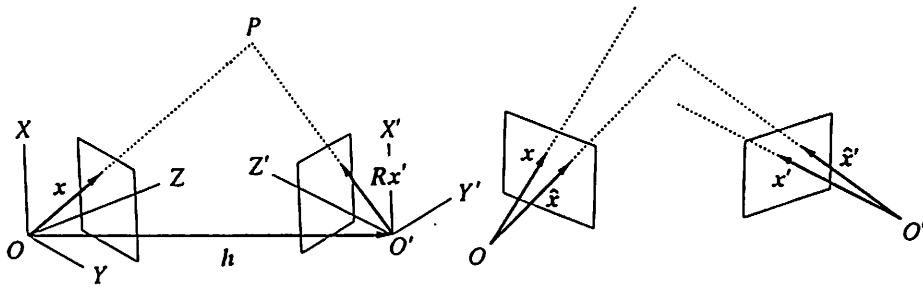


Fig. 1. (a) Geometry of camera motion. (b) Vectors $\mathbf{x}$ and $\mathbf{x}'$ are corrected so that their lines of sight meet in the scene.

the same point in the scene if and only if the following *epipolar equation* holds[2] [2, 4, 9, 12]:

$$|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{R}\boldsymbol{x}'| = 0. \tag{1}$$

This equation states that the *lines of sight* defined by vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ should meet in the scene. In the presence of image noise, vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ may not satisfy eq. (1) even if the motion parameters $\{\boldsymbol{h}, \boldsymbol{R}\}$ are correct. In order to determine a unique point in the scene, we must correct vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ so that their lines of sight meet in the scene (Fig. 1(b)).

3 General Model

Given N corresponding points (x_α, y_α), (x'_α, y'_α), $\alpha = 1, ..., N$, between the two images, we decompose the vectors $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$ that represent them into the form

$$\boldsymbol{x}_\alpha = \bar{\boldsymbol{x}}_\alpha + \Delta\boldsymbol{x}_\alpha, \qquad \boldsymbol{x}'_\alpha = \bar{\boldsymbol{x}}'_\alpha + \Delta\boldsymbol{x}'_\alpha, \tag{2}$$

where $\bar{\boldsymbol{x}}_\alpha$ and $\bar{\boldsymbol{x}}'_\alpha$ represent the "true" positions supposedly observed if noise did not exist (i.e., if the feature points detected from the gray-level images were accurate). We regard the noise terms $\Delta\boldsymbol{x}_\alpha$ and $\Delta\boldsymbol{x}'_\alpha$ as independent Gaussian random variables of mean 0 and respective covariance matrices[3] $V[\boldsymbol{x}_\alpha]$ $(= E[\Delta\boldsymbol{x}_\alpha \Delta\boldsymbol{x}_\alpha^\top])$ and $V[\boldsymbol{x}'_\alpha]$ $(= E[\Delta\boldsymbol{x}'_\alpha \Delta\boldsymbol{x}'^\top_\alpha])$ ($E[\,\cdot\,]$ denotes expectation). Although it is in general very difficult to predict the accuracy of feature point detection in advance, it is often possible to predict qualitative characteristics such as uniformity and isotropy. So, we assume that the covariance matrices $V[\boldsymbol{x}_\alpha]$ and $V[\boldsymbol{x}'_\alpha]$ are known *up to scale*. In other words, we assume that there exists an unknown constant ϵ, which we call the *noise level*, such that

$$V[\boldsymbol{x}_\alpha] = \epsilon^2 V_0[\boldsymbol{x}_\alpha], \qquad V[\boldsymbol{x}'_\alpha] = \epsilon^2 V_0[\boldsymbol{x}'_\alpha] \tag{3}$$

[2] $|\mathbf{a}, \mathbf{b}, \mathbf{c}|$ denotes the scalar triple product of vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$.

[3] Since the third components of vectors $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ are both 1, the covariance matrices $V[\mathbf{x}_\alpha]$ and $V[\mathbf{x}'_\alpha]$ are singular matrices of rank 2 whose ranges are the XY and $X'Y'$ planes, respectively.

for known matrices $V_0[x_\alpha]$ and $V_0[x_\alpha]$, which we call the *normalized covariance matrices*.

In the presence of noise, the corresponding points x_α and x'_α are not accurate, so they must be corrected so as to satisfy the epipolar equation (1). Infinitely many solutions exist for this correction. From among them, we choose the one that is statistically the most likely. This can be done by choosing the vectors $\hat{x}_\alpha$ and $\hat{x}'_\alpha$ that minimize the sum of the squared *Mahalanobis distances* [6]:

$$J = \sum_{\alpha=1}^{N}(x_\alpha - \hat{x}_\alpha, V_0[x_\alpha]^-(x_\alpha - \hat{x}_\alpha)) + \sum_{\alpha=1}^{N}(x'_\alpha - \hat{x}'_\alpha, V_0[x'_\alpha]^-(x'_\alpha - \hat{x}'_\alpha)). \quad (4)$$

Throughout this paper, we denote the inner product of vectors by $(\cdot, \cdot)$ and the *(Moore-Penrose) generalized inverse*[4] by $(\cdot)^-$. The first order solution of the above minimization is given as follows [6]:

$$\hat{x}_\alpha = x_\alpha - \frac{(x_\alpha, Gx'_\alpha)V_0[x_\alpha]Gx'_\alpha}{(\hat{x}'_\alpha, G^\top V_0[x_\alpha]G\hat{x}'_\alpha) + (\hat{x}_\alpha, GV_0[x'_\alpha]G^\top \hat{x}_\alpha)},$$
$$\hat{x}'_\alpha = x'_\alpha - \frac{(x_\alpha, Gx'_\alpha)V_0[x'_\alpha]G^\top x_\alpha}{(\hat{x}'_\alpha, G^\top V_0[x_\alpha]G\hat{x}'_\alpha) + (\hat{x}_\alpha, GV_0[x'_\alpha]G^\top \hat{x}_\alpha)}. \quad (5)$$

Here, the matrix G is defined by

$$G = h \times R, \quad (6)$$

and called the *essential matrix* [2, 4, 9, 12]; we define the product $a \times U$ of vector $a = (a_i)$ and matrix $U = (U_{ij})$ as a matrix whose (ij) element is[5] $\sum_{k,l=1}^{3} \varepsilon_{ikl}a_k U_{lj}$.

The minimum of the function J defined by eq. (4) is a function of the motion parameters $\{h, R\}$, so we write it as $J[h, R]$. Substituting eqs. (5) into eq. (4), we obtain the following expression [6]:

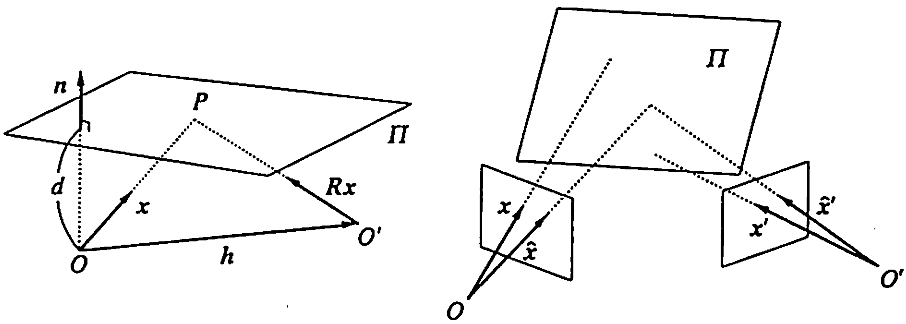
$$J[h, R] = \sum_{\alpha=1}^{N} \frac{(x_\alpha, Gx'_\alpha)^2}{(x'_\alpha, G^\top V_0[x_\alpha]Gx'_\alpha) + (x_\alpha, GV_0[x'_\alpha]G^\top x_\alpha)}. \quad (7)$$

The statistically most likely values of the motion parameters $\{h, R\}$ are those that minimize this function. From eq. (7), we can immediately see that the scale of the translation h is indeterminate[6], so we normalize it into $\|h\| = 1$ (we denote the norm of a vector by $\|\cdot\|$). The minimization of eq. (7) can be conducted accurately and effectively by a numerical technique called

[4] It is computed by applying the *spectral decomposition* [6] and replacing nonzero eigenvalues by their reciprocals.

[5] The symbol ε_{ijk} is the *Eddington epsilon*, taking values 1 and -1 if (iji) is an even and odd permutations of (123), respectively, and value 0 otherwise.

[6] This corresponds to the well known fact that a large camera motion relative to a large object in the distance is indistinguishable from a small camera motion relative to a small object near the camera as long as images are the only source of information.



(a) (b)

Fig. 2. (a) Geometry of camera motion relative to a planar surface. (b) Vectors $\mathbf{x}$ and $\mathbf{x}'$ are corrected so that their lines of sight meet on the surface Π.

renormalization [5]. According to our statistical model, the motion parameters $\{\hat{\boldsymbol{h}}, \hat{\boldsymbol{R}}\}$ that minimize eq. (7) and the corrected positions $\hat{\boldsymbol{x}}_\alpha$ and $\hat{\boldsymbol{x}}'_\alpha$ are the *maximum likelihood estimators* of $\{\boldsymbol{h}, \boldsymbol{R}\}$, $\bar{\boldsymbol{x}}_\alpha$, and $\bar{\boldsymbol{x}}'_\alpha$, respectively.

4 Planar Surface Model

Suppose the object is a planar surface Π. Let $\boldsymbol{n}$ be its unit surface normal with respect to the XYZ coordinate system, and d its distance (positive in the direction $\boldsymbol{n}$) from the origin O; we call $\{\boldsymbol{n}, d\}$ the *surface parameters* (Fig. 2(a)). It can be easily shown that vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be images of the same point on the surface Π if and only if the following equation holds [2, 4, 8, 9, 10, 12]:

$$\boldsymbol{x}' \times \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}. \tag{8}$$

Here, $\boldsymbol{A}$ is a matrix defined by

$$\boldsymbol{A} = \boldsymbol{R}^\top(\boldsymbol{h}\boldsymbol{n}^\top - d\boldsymbol{I}), \tag{9}$$

where $\boldsymbol{I}$ is the unit matrix. Eq. (8) states that the lines of sights defined by $\boldsymbol{x}$ and $\boldsymbol{x}'$ should meet on the surface Π; this is a stronger condition than the epipolar equation (1), which is automatically implied. However, eq. (8) may not exactly hold in the presence of noise, so we must correct $\boldsymbol{x}$ and $\boldsymbol{x}'$ so as to satisfy it (Fig. 2(b)).

Given N corresponding points $\boldsymbol{x}_\alpha$, $\boldsymbol{x}'_\alpha$, $\alpha = 1, ..., N$, the maximum likelihood estimators $\hat{\boldsymbol{x}}_\alpha$ and $\hat{\boldsymbol{x}}'_\alpha$ of $\bar{\boldsymbol{x}}_\alpha$ and $\bar{\boldsymbol{x}}'_\alpha$ for fixed surface and motion parameters $\{\boldsymbol{n}, d\}$ and $\{\boldsymbol{h}, \boldsymbol{R}\}$ are the vectors that minimize the function J given by eq. (4) under the constraint that they satisfy eq. (8). The first order solution is given as follows [6]:

$$\begin{aligned} \hat{\boldsymbol{x}}_\alpha &= \boldsymbol{x}_\alpha - (\boldsymbol{x}'_\alpha \times \boldsymbol{A}V_0[\boldsymbol{x}_\alpha])^\top \boldsymbol{W}_\alpha(\boldsymbol{x}'_\alpha \times \boldsymbol{A}\boldsymbol{x}_\alpha), \\ \hat{\boldsymbol{x}}'_\alpha &= \boldsymbol{x}'_\alpha + ((\boldsymbol{A}\boldsymbol{x}_\alpha) \times V_0[\boldsymbol{x}'_\alpha])^\top \boldsymbol{W}_\alpha(\boldsymbol{x}'_\alpha \times \boldsymbol{A}\boldsymbol{x}_\alpha), \end{aligned} \tag{10}$$

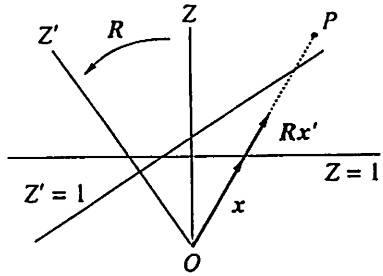


Fig. 3. Pure rotation of the camera.

$$W_\alpha = \left(x'_\alpha \times AV_0[x_\alpha]A^\top \times x'_\alpha + (Ax_\alpha) \times V_0[x'_\alpha] \times (Ax_\alpha)\right)^-_2. \tag{11}$$

The product $a \times A \times b$ of vectors $a = (a_i)$ and $b = (b_i)$ and a matrix $A = (A_{ij})$ is defined to be a matrix whose (ij) element is $\sum_{k,l,m,n=1}^{3} \varepsilon_{ikm}\epsilon_{jln}a_k b_l A_{mn}$. The symbol $(\cdot)^-_r$ denotes the *rank-constrained generalized inverse*[7] [6]. The minimum of the function J is a function of the surface and motion parameters $\{n, d\}$ and $\{h, R\}$, so we write it as $J[n, d, h, R]$. Substituting eqs. (10) into eq. (4), we obtain the following expression [6]:

$$J[n, d, h, R] = \sum_{\alpha=1}^{N} (x'_\alpha \times Ax_\alpha, W_\alpha(x'_\alpha \times Ax_\alpha)). \tag{12}$$

The maximum likelihood estimators $\{\hat{n}, \hat{d}\}$ and $\{\hat{h}, \hat{R}\}$ of the surface and motion parameters $\{n, d\}$ and $\{h, R\}$ are the values that minimize eq. (12). This minimization can be conducted accurately and effectively by a numerical technique called *renormalization* [7].

5 Pure Rotation Model

Let x and x' be the images of the same point in the scene. From Fig. 3, it is easily seen that the camera motion is a pure rotation if and only if there exists a rotation matrix R such that

$$x \times Rx' = 0. \tag{13}$$

Given N corresponding points x_α, x'_α, $\alpha = 1, \ldots, N$, the maximum likelihood estimators $\hat{x}_\alpha$ and $\hat{x}'_\alpha$ of $\bar{x}_\alpha$ and $\bar{x}'_\alpha$ for a fixed rotation R are the vectors that

[7] The matrix obtained by applying the spectral decomposition, replacing all but the largest r eigenvalues by 0, and computing the generalized inverse. This operation is necessary to prevent numerical instability of the computation; the expression inside the parentheses becomes singular in the limit as $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ approach $\bar{\mathbf{x}}_\alpha$ and $\bar{\mathbf{x}}'_\alpha$, respectively.

minimize the function J given by eq. (4) under the constraint that they satisfy eq. (13). The first order solution is given as follows [6]:

$$\hat{\boldsymbol{x}}_\alpha = \boldsymbol{x}_\alpha + ((\boldsymbol{R}\boldsymbol{x}'_\alpha) \times V_0[\boldsymbol{x}_\alpha])^\top \boldsymbol{W}_\alpha(\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha),$$
$$\hat{\boldsymbol{x}}'_\alpha = \boldsymbol{x}'_\alpha - (\boldsymbol{x}_\alpha \times \boldsymbol{R}V_0[\boldsymbol{x}'_\alpha])^\top \boldsymbol{W}_\alpha(\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha), \qquad (14)$$

$$\boldsymbol{W}_\alpha = \Big((\boldsymbol{R}\boldsymbol{x}'_\alpha) \times V_0[\boldsymbol{x}_\alpha] \times (\boldsymbol{R}\boldsymbol{x}'_\alpha) + \boldsymbol{x}_\alpha \times \boldsymbol{R}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{R}^\top \times \boldsymbol{x}_\alpha\Big)_2^- . \qquad (15)$$

The minimum of the function J is a function of the rotation $\boldsymbol{R}$, so we write it as $J[\boldsymbol{R}]$. Substituting eqs. (14) into eq. (4), we obtain the following expression [6]:

$$J[\boldsymbol{R}] = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha, \boldsymbol{W}_\alpha(\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha)). \qquad (16)$$

The maximum likelihood estimators $\hat{\boldsymbol{R}}$ of the rotation $\boldsymbol{R}$ is the matrix that minimizes eq. (16). This minimization can be conducted numerically by steepest descent. A good approximation[8] of $\hat{\boldsymbol{R}}$ is analytically obtained by applying the singular value decomposition [4].

6 Geometric Model

Regardless of the shape of the object and the motion of the camera, the problem can be generalized in abstract terms as follows. A pair of corresponding vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be identified with a *six*-dimensional direct sum vector $\boldsymbol{x} \oplus \boldsymbol{x}'$. The third components of $\boldsymbol{x}$ and $\boldsymbol{x}'$ are both 1, so vector $\boldsymbol{x} \oplus \boldsymbol{x}'$ is constrained to be in the four-dimensional affine subspace

$$\mathcal{X} = \{(x, y, 1, x', y', 1)^\top | x, y, x', y' \in \mathcal{R}\}, \qquad (17)$$

which we call the *data space* ($\mathcal{R}$ denotes the set of all real numbers). Observing N corresponding points $\boldsymbol{x}_\alpha$ and $\boldsymbol{x}'_\alpha$ is equivalent to observing N vectors $\boldsymbol{x}_\alpha \oplus \boldsymbol{x}'_\alpha$ in the data space $\mathcal{X}$.

Suppose there exists a constraint on the shape of the object and/or the motion of the camera that can be expressed as L equations parameterized by an n-dimensional vector $\boldsymbol{u}$ in the form

$$F^{(k)}(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{u}) = 0, \qquad k = 1, ..., L. \qquad (18)$$

These L equations need not be algebraically independent[9] as equations of $\boldsymbol{x}$ and $\boldsymbol{x}'$. We call the number r of independent equations the *rank* of the constraint.

[8] As shown shortly, what we actually need is not the estimate $\hat{\mathbf{R}}$ itself but the residual $J[\hat{\mathbf{R}}]$. Since $\hat{\mathbf{R}}$ minimizes $J[\mathbf{R}]$, the residual $J[\hat{\mathbf{R}}]$ can be accurately computed even from an approximate value of $\hat{\mathbf{R}}$.

[9] In order to avoid pathological cases, we assume that each of the L equations defines a manifold of *codimension 1* in the data space $\mathcal{X}$ such that the L manifolds intersect each other *transversally* [6].

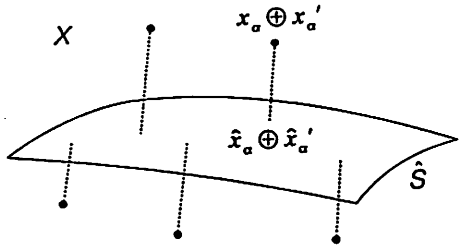


Fig. 4. The model $\mathcal{S}$ is optimally fitted to the data points, and the data points are optimally projected onto it.

Eq. (18) then defines a manifold $\mathcal{S}$ of *codimension*[10] r in the data space $\mathcal{X}$. We call $\mathcal{S}$ the (*geometric*) *model*. The domain $\mathcal{U}$ of the vector u that parameterizes the constraint is called the *parameter space*. If it is an n'-dimensional manifold in $\mathcal{R}^n$, the model $\mathcal{S}$ is said to have n' *degrees of freedom*.

Given N corresponding points x_α, x'_α, $\alpha = 1, ..., N$, in the presence of noise, there may not exist an instance of $\mathcal{S}$ that exactly passes through all the N points $x \oplus x'_\alpha \in \mathcal{X}$. So, we optimally *fit* the model $\mathcal{S}$ by adjusting the parameter $u \in \mathcal{U}$ in the sense of *maximum likelihood estimation*. Let $\hat{\mathcal{S}}$ be the resulting optimal fit. We then optimally *project* each point $x_\alpha \oplus x'_\alpha \in \mathcal{X}$ onto $\hat{\mathcal{S}}$ in the sense of *maximum likelihood estimation* (Fig. 4). Let $\hat{x}_\alpha \oplus \hat{x}'_\alpha$ be the resulting optimal projection; $\hat{x}_\alpha$ and $\hat{x}'_\alpha$ are the maximum likelihood estimators of $\bar{x}_\alpha$ and $\bar{x}'_\alpha$, respectively. Let $\hat{J}$ be the residual of the function J for the maximum likelihood estimators $\hat{x}_\alpha$ and $\hat{x}'_\alpha$. It can be proved that $\hat{J}/\epsilon^2$ is subject to a χ^2 distribution with $rN - n'$ degrees of freedom in the first order [6]. Hence, an unbiased estimator of the squared noise level ϵ^2 is obtained in the following form:

$$\hat{\epsilon}^2 = \frac{\hat{J}}{rN - n'} \tag{19}$$

7 Geometric Information Criterion

The "goodness" of a model can be measured by its "predicting capability" [1]. Let x^*_α and $x^{*\prime}_\alpha$ be *future data* that have the same probability distribution as the current data x_α and x'_α and are independent of x_α and x'_α. The residual of model $\mathcal{S}$ for the maximum likelihood estimators $\hat{x}_\alpha$ and $\hat{x}'_\alpha$, which are computed from the current data x_α and x'_α, *with respect to the future data* x^*_α *and* $x^{*\prime}_\alpha$ is

$$\hat{J}^* = \sum_{\alpha=1}^{N}(x^*_\alpha - \hat{x}_\alpha, V_0[x_\alpha]^-(x^*_\alpha - \hat{x}_\alpha)) + \sum_{\alpha=1}^{N}(x^{*\prime}_\alpha - \hat{x}'_\alpha, V_0[x'_\alpha]^-(x^{*\prime}_\alpha - \hat{x}'_\alpha)). \tag{20}$$

[10] The difference between the dimension of the space (in this case $\mathcal{X}$) and the dimension of the manifold.

This is a random variable; its expectation is

$$I(\mathcal{S}) = E^*[E[\hat{J}^*]], \tag{21}$$

where $E[\cdot]$ and $E^*[\cdot]$ denotes expectation with respect to the current data x_α and x'_α and the future data x^*_α and $x^{*\prime}_\alpha$, respectively. We call $I(\mathcal{S})$ the *expected residual*. A model with a small expected residual is expected to have high predicting capability [1]. It can be proved [6] that an unbiased estimator of the expected residual is given by

$$AIC(\mathcal{S}) = \hat{J} + 2(pN + n')\epsilon^2, \tag{22}$$

where $p = 4 - r$ is the dimension of the manifold $\mathcal{S}$. We call eq. (22) the *geometric information criterion*, or *geometric AIC* for short.

8 Automatic Model Selection

Let $\mathcal{S}_1$ be a model of dimension p_1 and codimension r_1 with n'_1 degrees of freedom, and $\mathcal{S}_2$ a model of dimension p_2 and codimension r_2 with n'_2 degrees of freedom. Suppose model $\mathcal{S}_2$ is obtained by adding an additional constraint to model $\mathcal{S}_2$. We say that model $\mathcal{S}_2$ is *stronger* than model $\mathcal{S}_1$, or model $\mathcal{S}_1$ is *weaker* than model $\mathcal{S}_2$, and write

$$\mathcal{S}_2 \succ \mathcal{S}_1. \tag{23}$$

Let $\hat{J}_1$ and $\hat{J}_2$ be the residuals of $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively. If model $\mathcal{S}_1$ is correct, the squared noise level ϵ^2 is estimated by eq. (19). Substituting it into the expression for the geometric AIC, we obtain

$$AIC(\mathcal{S}_1) = \hat{J}_1 + \frac{2(p_1N + n'_1)}{r_1N - n'_1}\hat{J}_1, \qquad AIC(\mathcal{S}_2) = \hat{J}_2 + \frac{2(p_2N + n'_2)}{r_1N - n'_1}\hat{J}_1. \tag{24}$$

Recalling that the geometric AIC is an estimator of the expected residual (see eq. (20)), we put

$$K = \sqrt{\frac{AIC(\mathcal{S}_2)}{AIC(\mathcal{S}_1)}} = \sqrt{\frac{r_1N - n'_1}{(2p_1 + r_1)N + n'_1}\left(\frac{\hat{J}_2}{\hat{J}_1} + \frac{2(p_2N + n'_2)}{r_1N - n'_1}\right)}. \tag{25}$$

This quantity describes the ratio of the expected deviation from model $\mathcal{S}_2$ to the expected deviation from model $\mathcal{S}_1$. It follows that if

$$K < 1 \tag{26}$$

model $\mathcal{S}_2$ is *preferable* to $\mathcal{S}_1$ with regard to the predicting capability. This criterion requires no knowledge about the noise magnitude and involves no empirically adjustable thresholds.

9 Planarity and Rotation Tests

The epipolar equation (1) defines a model $\mathcal{S}$ of codimension 1 in the four-dimensional data space $\mathcal{X}$ parameterized by the motion parameters $\{\boldsymbol{h}, \boldsymbol{R}\}$, which have five degrees of freedom. Hence, its geometric AIC is

$$AIC(\mathcal{S}) = \hat{J} + (6N+10)\epsilon^2. \tag{27}$$

The three components of eq. (8) are algebraically independent, so the rank of the constraint that the object is a planar surface is 2. Hence, it defines a model $\mathcal{S}_\Pi$ of codimension 2 in $\mathcal{X}$ parameterized by the surface and motion parameters $\{\boldsymbol{n}, d\}$ and $\{\boldsymbol{h}, \boldsymbol{R}\}$, which have eight degrees of freedom. Its geometric AIC is

$$AIC(\mathcal{S}_\Pi) = \hat{J}_\Pi + (4N+16)\epsilon^2. \tag{28}$$

The three components of eq. (13) are algebraically independent, so the rank of the constraint that the camera motion is a pure rotation is 2. Hence, it defines a model $\mathcal{S}_R$ of codimension 2 in $\mathcal{X}$ parameterized by the rotation $\boldsymbol{R}$, which has three degrees of freedom. Its geometric AIC is

$$AIC(\mathcal{S}_R) = \hat{J}_\Pi + (4N+6)\epsilon^2. \tag{29}$$

From eqs. (1), (8), and (13), we observe the following order of strength:

$$\mathcal{S}_R \succ \mathcal{S}_\Pi \succ \mathcal{S}. \tag{30}$$

Since the general model $\mathcal{S}$ is also valid for the planar surface model $\mathcal{S}_\Pi$ and the pure rotation model $\mathcal{S}_R$, the object is judged to be a planar surface if

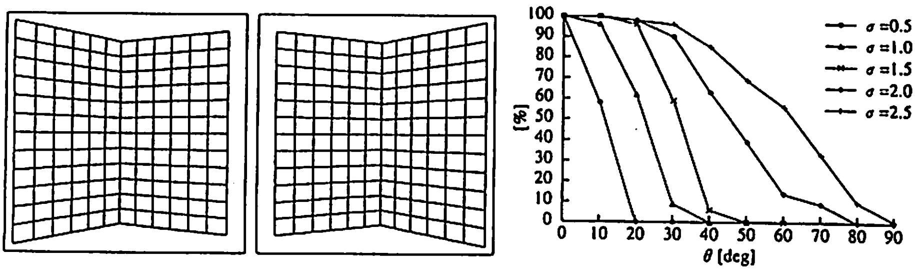
$$K_\Pi = \sqrt{\frac{N-5}{7N+5}\left(\frac{\hat{J}_\Pi}{\hat{J}} + \frac{4N+16}{N-5}\right)} < 1, \tag{31}$$

and the camera motion is judged to be a pure rotation if

$$K_R = \sqrt{\frac{N-5}{7N+5}\left(\frac{\hat{J}_R}{\hat{J}} + \frac{4N+6}{N-5}\right)} < 1. \tag{32}$$

10 Examples

Two planar grids hinged together with angle $\pi - \theta$ were defined in the scene, and images that simulate two views from different camera positions were generated. The image size and the focal length were assumed to be 512×512 (pixels) and 600 (pixels), respectively. Fig. 5(a) shows the images for $\theta = 50°$. The x- and y-coordinates of each grid point were perturbed by independent random Gaussian noise of mean 0 and standard deviation σ (pixels). Using the grid points as feature points, we conducted the planarity test 100 times, each time using



(a) (b)

Fig. 5. (a) Images of two planar grids hinged together in the scene ($\theta = 50°$). (b) The percentage of the instances judged as planar.

different noise. Fig. 5(b) shows, for various values of θ and for $\sigma = 0.5 \sim 3.0$, the percentage of the instances for which the object is judged as planar. We see that the threshold for the test is automatically adjusted to the noise.

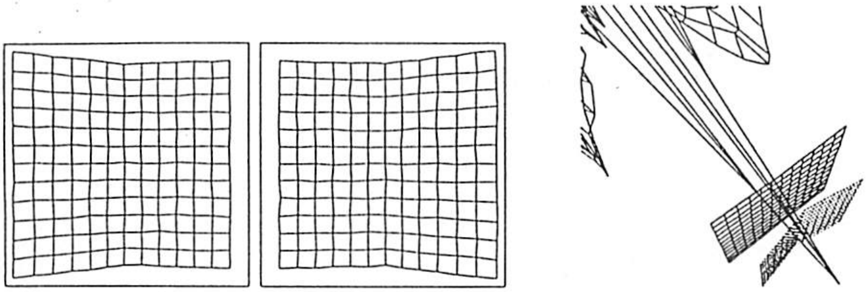
If $\sigma = 1.0$, the percentage is approximately 50% for $\theta = 22°$. Fig. 6(a) shows one instance for which the object is judged as planar ($K_{\Pi} = 0.87$), and Fig. 6(b) shows the 3-D shapes reconstructed from the general and planar surface models. The true shape is superimposed in dashed lines. Fig. 7(a) shows a one instance for which the object is judged as non-planar ($K_{\Pi} = 1.13$). From these, we see that although the perturbed images look almost the same, the reconstructed shape from the general model has little sense if the object is judged as planar, while the non-planar shape can be reconstructed fairly well if the object is judged as non-planar.

11 Concluding Remarks

We have presented a scheme for automatically testing if the object can be regarded as a planar surface or if the camera motion is a pure rotation without using any knowledge about the noise in the images. This judgment does not involve any empirically adjustable thresholds, either. The basic principle of our scheme is to choose a model that has "higher predicting capability" measured by the *geometric information criterion* (*geometric AIC*). Our approach presents a new paradigm for geometric model selection in a wide range of problems of robotics and computer vision.

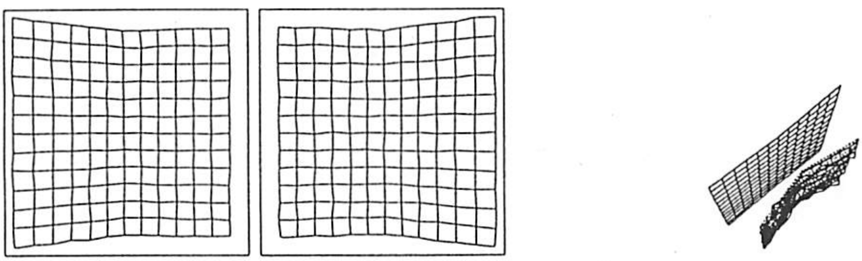
References

1. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Automation Control 19-6 (1974) 776–723
2. Faugeras, O.: Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge, MA, 1993
3. Faugeras, O. D. and S. Maybank, S.: Motion from point matches: Multiplicity of solutions. Int. J. Comput. Vision 4-3 (1990) 225–246



(a) (b)

Fig. 6. (a) An instance for which the object is judged as planar. (b) Reconstructed 3-D shapes and the true shape.



(a) (b)

Fig. 7. (a) An instance for which the object is judged as non-planar. (b) Reconstructed 3-D shapes and the true shape.

4. Kanatani, K.: Geometric Computation for Machine Vision. Oxford University Press, Oxford, 1993
5. Kanatani, K.: Renormalization for motion analysis: Statistically optimal algorithm. IEICE Trans. Inf. & Sys. **E77-D**-11 (1994) 1233–1239
6. Kanatani, K.: Statistical Optimization for Geometric Computation: Theory and Practice. Elsevier Science, Amsterdam, 1996
7. Kanatani, K. and Takeda, S.: 3-D motion analysis of a planar surface by renormalization. IEICE Trans. Inf. & Syst. **E78-D**-8 (1995) 1074–1079
8. Longuet-Higgins, L. C.: The reconstruction of a plane surface from two perspective projections. Proc. Roy. Soc. Lond. B**227** (1986) 399–410
9. Maybank, S.: Theory of Reconstruction from Image Motion. Springer, Berlin, 1993
10. Weng, J., Ahuja, N. and Huang, T. S.: Motion and structure from point correspondences with error estimation: Planar surfaces. IEEE Trans. Sig. Proc. **39**-12 (1991) 2691–2717
11. Weng, J., Ahuja, N. and Huang, T. S.: Optimal motion and structure estimation. IEEE Trans. Patt. Anal. Mach. Intell. **15**-9 (1993) 864–884
12. Weng, J., Huang, T. S. and Ahuja, N.: Motion and Structure from Image Sequences. Springer, Berlin, 1993