PAPER

Infinity and Planarity Test for Stereo Vision

Yasushi KANAZAWA† *and* **Kenichi KANATANI**††, *Members*

SUMMARY Introducing a mathematical model of noise in stereo images, we propose a new criterion for intelligent statistical inference about the scene we are viewing by using the *geometric information criterion* (*geometric AIC*). Using synthetic and real-image experiments, we demonstrate that a robot can test whether or not the object is located very far away or the object is a planar surface without using any knowledge about the noise magnitude or any empirically adjustable thresholds.
key words: AIC, testing of hypotheses, model selection, stereo vision, infinity test, planarity test.

1. Introduction

Stereo vision is one of the most widely used means of sensing for autonomous robot navigation, and various techniques have been proposed for correspondence detection between the two images [3], [13]. The reliability of the reconstructed 3-D shape has also been studied by introducing a statistical model of image noise [2], [8]–[11]. However, no attempts seem to have been made at intelligent statistical inference about the scene based on stereo images.

For example, the 3-D shape of an object can be accurately reconstructed from stereo images if the object is known to be a planar surface [8]. Hence, if a robot can infer that the object is a planar surface, it can produce an accurate planar surface by using that knowledge. This is very important for robots working indoors, since indoor scenes have many planar objects (walls, ceilings, floors, tables, etc.).

On the other hand, the 3-D reconstruction is unreliable if the object is very far away. If a robot can infer that the object is too far away, it can output, say, a warning message, telling us that the disparity is too short for reliable 3-D reconstruction. This is very important for a stereo system with a short baseline, e.g., when small cameras are attached to a robot hand.

A naive solution to these problems is first reconstructing the 3-D shape and deciding that the object is planar if the reconstructed shape is approximately planar within a specified threshold or the object is too away if the reconstructed position is farther a way than a specified threshold. But *how can we set these thresholds?*

The thresholds should be high if the images are accurate and low if the images are not accurate. They also depend on the resolution of the cameras, the shape of the objects, the lighting condition, the number of corresponding points, the accuracy of the image processing techniques involved, and many other factors. Hence, even if we find an appropriate threshold value after repeating many trial-and-error experiments, that value becomes meaningless in a new environment.

In this paper, we present an *environment-independent* criterion for intelligent statistical inference *without introducing any thresholds to be adjusted empirically*. It is obtained by applying the *geometric information criterion* (or *geometric AIC* for short) [5], [7]. We show synthetic and real-image examples to illustrate our method.

2. Statistical Model of Image Noise

We 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 in the direction of the optical axis. With an appropriate scaling, the image plane can be identified with the plane $Z = 1$, on which an xy image coordinate system is defined in such a way that the origin o is on the Z-axis and the x- and y-axes are parallel to the X- and Y-axes, respectively.

For a stereo system, we define a reference coordinate system with respect to the first camera and place the second camera in a position obtained by translating the first camera by vector $\boldsymbol{h}$ and rotating it around the center of the lens by matrix $\boldsymbol{R}$. We call $\{\boldsymbol{h}, \boldsymbol{R}\}$ the *motion parameters* (Fig. 1).

Let (x, y) be the image coordinates of a feature point projected onto the image plane of the first camera, and (x', y') those for the second camera. We use the following three-dimensional vectors to represent them (the superscript $\top$ denotes transpose):

$$\boldsymbol{x} = (x, y, 1)^\top, \quad \boldsymbol{x}' = (x', y', 1)^\top. \tag{1}$$

Let $\{\boldsymbol{x}, \boldsymbol{x}'\}$ be a corresponding pair. Their uncertainty can be described by their *covariance matrices* $V[\boldsymbol{x}]$ and $V[\boldsymbol{x}']$ by regarding $\boldsymbol{x}$ and $\boldsymbol{x}'$ as independent

Manuscript received February 29, 1996.
Manuscript revised December 16, 1996.
†The author is with the Department of Information and Computer Engineering, Gunma College of Technology, Maebashi-shi, 371 Japan.
††The author is with the Department of Computer Science, Gunma University, Kiryu-shi, 376 Japan.

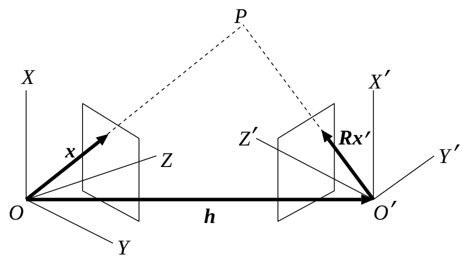

Fig. 1 Camera imaging geometry and the motion parameters of a stereo system.

Gaussian random variables. The form of the covariance matrices depend on the matching method of the two images. Matthies and Shafer [10] discussed this problem in detail and gave uncertainty models for various types of matching methods. In general, the absolute magnitude of the uncertainty is difficult to predict a priori, but its geometric characteristics, such as homogeneity/inhomogeneity and isotropy/anisotropy can be relatively easily predicted. Here, we assume that the covariance matrices are known only *up to scale* and write

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

The constant ϵ, which indicates the average magnitude of the image noise, is assumed unknown; we call it the *noise level*. The matrices $V_0[\boldsymbol{x}]$ and $V_0[\boldsymbol{x}']$ indicate in which orientation the deviation is likely to occur; they are assumed known and called the *normalized covariance matrices*. Since the third components of $\boldsymbol{x}$ and $\boldsymbol{x}'$ are 1, they are singular matrices of rank 2.

3. General Stereo Vision

Ideally, vectors $\boldsymbol{x}$, $\boldsymbol{R}\boldsymbol{x}'$, and $\boldsymbol{h}$ should be coplanar (Fig. 1), so they should satisfy the following *epipolar equation* [3], [4], [13]:

$$|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{R}\boldsymbol{x}'| = 0. \tag{3}$$

In this paper, $|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}|$ denotes the scalar triple product of vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.

Let $\{\boldsymbol{x}_\alpha,\ \boldsymbol{x}'_\alpha\}$, $\alpha = 1, ..., N$, represent the corresponding points, and let $V_0[\boldsymbol{x}_\alpha]$ and $V_0[\boldsymbol{x}'_\alpha]$ be their normalized covariance matrices. Since they do not necessarily satisfy Eq. (3), we correct them so that the corrected positions $\{\hat{\boldsymbol{x}}_\alpha,\ \hat{\boldsymbol{x}}'_\alpha\}$ satisfy Eq. (3). An optimal correction in the sense of maximum likelihood estimation is obtained by minimizing the sum of the squared *Mahalanobis distances*

$$J = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^-(\boldsymbol{x}_\alpha - \hat{\boldsymbol{x}}_\alpha)) + \sum_{\alpha=1}^{N}(\boldsymbol{x}'_\alpha - \hat{\boldsymbol{x}}'_\alpha, V_0[\boldsymbol{x}'_\alpha]^-(\boldsymbol{x}'_\alpha - \hat{\boldsymbol{x}}'_\alpha)) \to \min \tag{4}$$

under the constraint that $\{\hat{\boldsymbol{x}}_\alpha,\ \hat{\boldsymbol{x}}'_\alpha\}$ satisfy Eq. (3). Here, $V_0[\boldsymbol{x}_\alpha]^-$ and $V_0[\boldsymbol{x}'_\alpha]^-$ are the (Moore-Penrose) generalized inverses of $V_0[\boldsymbol{x}_\alpha]$ and $V_0[\boldsymbol{x}'_\alpha]$, respectively. In this paper, $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

It can be shown that the minimum value of J, which we call simply the *residual*, is given as follows [5]:

$$\hat{J} = \sum_{\alpha=1}^{N}\frac{(\boldsymbol{x}_\alpha, \boldsymbol{G}\boldsymbol{x}'_\alpha)^2}{(\boldsymbol{x}'_\alpha, \boldsymbol{G}^\top V_0[\boldsymbol{x}_\alpha]\boldsymbol{G}\boldsymbol{x}'_\alpha) + (\boldsymbol{x}_\alpha, \boldsymbol{G}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{G}^\top \boldsymbol{x}_\alpha)}. \tag{5}$$

Here, $\boldsymbol{G}$ is the *essential matrix* [3], [4], [13] defined by

$$\boldsymbol{G} = \boldsymbol{h} \times \boldsymbol{R}, \tag{6}$$

where the right-hand side denotes the matrix defined by the vector product of $\boldsymbol{h}$ and each column of $\boldsymbol{R}$.

4. Infinitely Far Away Scene

Suppose $\{\boldsymbol{x},\ \boldsymbol{x}'\}$ are images of a feature point that belongs to an object located practically infinitely far away (e.g., a mountain, a boat on the sea, or an airplane in the sky). Ideally, the following equation must be satisfied [5]:

$$\boldsymbol{x} \times \boldsymbol{R}\boldsymbol{x}' = \boldsymbol{0}. \tag{7}$$

The observed positions $\{\boldsymbol{x}_\alpha,\ \boldsymbol{x}'_\alpha\}$ do not necessarily satisfy this equation. So, we correct them so that the corrected positions $\{\hat{\boldsymbol{x}}_\alpha,\ \hat{\boldsymbol{x}}'_\alpha\}$ satisfy Eq. (7). An optimal correction in the sense of maximum likelihood estimation is obtained by the minimization (4) under the constraint that $\{\hat{\boldsymbol{x}}_\alpha,\ \hat{\boldsymbol{x}}'_\alpha\}$ satisfy Eq. (7), and the residual is given as follows [5]:

$$\hat{J}_\infty = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha, \boldsymbol{W}_\alpha(\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha)). \tag{8}$$

Here, $\boldsymbol{W}_\alpha$ is a matrix defined by

$$\boldsymbol{W}_\alpha = \Big((\boldsymbol{R}\boldsymbol{x}'_\alpha) \times V_0[\boldsymbol{x}_\alpha] \times (\boldsymbol{R}\boldsymbol{x}'_\alpha) + \boldsymbol{x}_\alpha \times \boldsymbol{R}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{R}^\top \times \boldsymbol{x}_\alpha\Big)_2^-. \tag{9}$$

For a vector $\boldsymbol{a}$ and a matrix $\boldsymbol{U}$, the product $\boldsymbol{a} \times \boldsymbol{U} \times \boldsymbol{a}$ is an abbreviation of $(\boldsymbol{a} \times \boldsymbol{U})(\boldsymbol{a} \times \boldsymbol{I})^\top$, where $\boldsymbol{I}$ is the unit matrix. The symbol $(\,\cdot\,)_2^-$ denotes generalized inverse computed after projecting the matrix onto a matrix of rank 2 by ignoring the smallest eigenvalue [5]. This operation is necessary to prevent numerical instability of the computation [5].

5. Planar Surface Scene

Suppose $\{\boldsymbol{x},\ \boldsymbol{x}'\}$ are images of a feature point on a planar surface. Let $\boldsymbol{n}$ be the unit normal to the surface, and d its distance (positive in the direction $\boldsymbol{n}$) from the origin O. We call $\{\boldsymbol{n}, d\}$ the *surface parameters*. Ideally, the following equation must be satisfied [4], [5], [8]:

$$\boldsymbol{x}' \times \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \qquad \boldsymbol{A} = \boldsymbol{R}^\top(\boldsymbol{h}\boldsymbol{n}^\top - d\boldsymbol{I}). \tag{10}$$

So, we correct the observed positions $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$ in such a way that the corrected positions $\{\hat{\boldsymbol{x}}_\alpha, \hat{\boldsymbol{x}}'_\alpha\}$ satisfy Eqs. (10) for *some* surface parameters $\{\boldsymbol{n}, d\}$. An optimal correction of $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$ and estimation of $\{\boldsymbol{n}, d\}$ in the sense of maximum likelihood estimation can be done by the minimization (4). This computation can be efficiently done by an iterative procedure called *renormalization* [8]. The residual of the optimization (4) is given as follows [5]:

$$\hat{J}_\Pi = \sum_{\alpha=1}^{N}(\boldsymbol{x}'_\alpha \times \hat{\boldsymbol{A}}\boldsymbol{x}_\alpha, \boldsymbol{W}_\alpha(\boldsymbol{x}'_\alpha \times \hat{\boldsymbol{A}}\boldsymbol{x}_\alpha)). \tag{11}$$

Here, $\hat{\boldsymbol{A}}$ is the matrix obtained by substituting the surface parameters $\{\hat{\boldsymbol{n}}\ \hat{d}\}$ computed by renormalization into the second of Eqs. (10), and $\boldsymbol{W}_\alpha$ is a matrix defined by

$$\boldsymbol{W}_\alpha = \Big(\boldsymbol{x}'_\alpha \times \hat{\boldsymbol{A}}V_0[\boldsymbol{x}_\alpha]\hat{\boldsymbol{A}}^\top \times \boldsymbol{x}'_\alpha + (\hat{\boldsymbol{A}}\boldsymbol{x}_\alpha) \times V_0[\boldsymbol{x}'_\alpha] \times (\hat{\boldsymbol{A}}\boldsymbol{x}_\alpha)\Big)^-_2. \tag{12}$$

6. Geometric Model

3-D reconstruction by stereo vision can be generalized in abstract terms as follows. A pair of corresponding vectors $\{\boldsymbol{x}, \boldsymbol{x}'\}$ can be identified with a six-dimensional direct sum vector $\boldsymbol{x} \oplus \boldsymbol{x}'$. Since the third components of $\boldsymbol{x}$ and $\boldsymbol{x}'$ are both 1, the vector $\boldsymbol{x} \oplus \boldsymbol{x}'$ is constrained to be in the four-dimensional affine subspace

$$\mathcal{X} = \{(x, y, 1, x', y', 1)^\top | x, y, x', y' \in \mathcal{R}\} \subset \mathcal{R}^6, \tag{13}$$

which we call the *data space* ($\mathcal{R}$ denotes the set of real numbers).

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

$$F^{(k)}(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{u}) = 0, \qquad k = 1, ..., L. \tag{14}$$

As equations of $\{\boldsymbol{x}, \boldsymbol{x}'\}$, these L equations need not be algebraically independent†. We call the number r of independent equations the *rank* of the constraint. Equation (14) then defines a manifold $\mathcal{S}$ of *codimension* (= the difference between the dimension of the space, in this case $\mathcal{X}$, and the dimension of the manifold) r in the data space $\mathcal{X}$. We call $\mathcal{S}$ the (*geometric*) *model*. The domain $\mathcal{U}$ of the vector $\boldsymbol{u}$ that parameterizes the constraint is called the *parameter space*. If the parameter space $\mathcal{U}$ is an n'-dimensional manifold in $\mathcal{R}^n$, we say that the model $\mathcal{S}$ has n' *degrees of freedom*.

†In order to avoid pathological cases, we need to assume that each of the L equations defines a manifold of *codimension 1* in the data space $\mathcal{X}$ in such a way that the L manifolds intersect each other *transversally*. See [5] for the details.

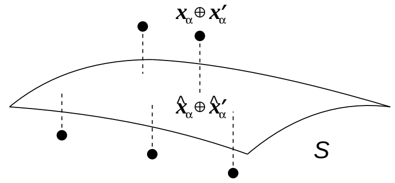

Fig. 2 The model $\mathcal{S}$ is optimally fitted to the data points, and the data points are optimally projected onto it.

Given N corresponding points $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$, $\alpha = 1$, ..., N, in the presence of noise, there may not exist any instance of $\mathcal{S}$ that exactly passes through all the N points $\boldsymbol{x}_\alpha \oplus \boldsymbol{x}'_\alpha \in \mathcal{X}$. The 3-D reconstruction by stereo vision can be formally stated as the problem of estimating the true positions $\{\bar{\boldsymbol{x}}_\alpha, \bar{\boldsymbol{x}}'_\alpha\}$, $\alpha = 1$, ..., N, and the true value $\bar{\boldsymbol{u}}$ of the parameter that satisfy

$$F^{(k)}(\bar{\boldsymbol{x}}_\alpha, \bar{\boldsymbol{x}}'_\alpha; \bar{\boldsymbol{u}}) = 0, \qquad k = 1, ..., L, \tag{15}$$

for $\alpha = 1$, ..., N from the noisy data $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$.

If we write the positions into which $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$ are to be corrected as $\{\hat{\boldsymbol{x}}_\alpha, \hat{\boldsymbol{x}}'_\alpha\}$, the maximum likelihood solution of the above problem is obtained by the optimization (4) under the constraint that $\boldsymbol{x}_\alpha \oplus \boldsymbol{x}'_\alpha \in \mathcal{S}$. In geometric terms, this is equivalent to *fitting* the model $\mathcal{S}$ optimally by adjusting the parameter $\boldsymbol{u} \in \mathcal{U}$ (Fig. 2).

Let $\hat{\mathcal{S}}$ be the resulting optimal fit, which is the maximum likelihood estimator of the model $\mathcal{S}$. For a fixed parameter value $\boldsymbol{u}$, the maximum likelihood estimators of $\{\bar{\boldsymbol{x}}_\alpha, \bar{\boldsymbol{x}}'_\alpha\}$ are obtained by *projecting* each direct sum point $\boldsymbol{x}_\alpha \oplus \boldsymbol{x}'_\alpha \in \mathcal{X}$ onto $\mathcal{S}$. Let $\hat{\boldsymbol{x}}_\alpha \oplus \hat{\boldsymbol{x}}'_\alpha$ be the resulting optimal projection, and let $\hat{J}$ be the residual of the function J obtained by substituting $\{\hat{\boldsymbol{x}}_\alpha, \hat{\boldsymbol{x}}'_\alpha\}$ for $\{\bar{\boldsymbol{x}}_\alpha, \bar{\boldsymbol{x}}'_\alpha\}$. 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 [5]. Hence, an unbiased estimator of the squared noise level ϵ^2 is obtained in the following form:

$$\hat{\epsilon}^2 = \frac{\hat{J}}{rN - n'} \tag{16}$$

7. Geometric Information Criterion

A good model should explain the observed data $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$, $\alpha = 1$, ..., N, well, which implies that the residual $\hat{J}$ should be small. However, since $\hat{J}/\epsilon^2$ is subject to a χ^2 distribution with $rN - n'$ degrees of freedom, the residual $\hat{J}$ becomes smaller as r decreases and n' increases. In other words, the residual $\hat{J}$ can be made arbitrarily small by reducing the codimension r (i.e., increasing the dimension) of $\mathcal{S}$ and increasing the number of free parameters n' of $\mathcal{S}$. Such a model can only explains the *current data* which happen to be observed; there is no guarantee that it could explain the data if the noise occurred differently.

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

$$\hat{J}^* = \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha^* - \hat{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^-(\boldsymbol{x}_\alpha^* - \hat{\boldsymbol{x}}_\alpha)) + \sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha^{*\prime} - \hat{\boldsymbol{x}}'_\alpha, V_0[\boldsymbol{x}'_\alpha]^-(\boldsymbol{x}_\alpha^{*\prime} - \hat{\boldsymbol{x}}'_\alpha)). \quad (17)$$

It can be shown that $\hat{J}$ is smaller than $\hat{J}^*$ by $2(pN+n')$ *in expectation*, where $p = 4 - r$ is the dimension of the manifold $\mathcal{S}$ [5]. Hence, the *geometric information criterion* (or *geometric AIC* for short) is defined as follows [5]:

$$AIC(\mathcal{S}) = \hat{J} + 2(pN + n')\epsilon^2. \quad (18)$$

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 $\mathcal{S}_1$. 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

$$\mathcal{S}_2 \succ \mathcal{S}_1. \quad (19)$$

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. (16). Substituting it into the expression for the geometric AIC, we obtain

$$AIC(\mathcal{S}_1) = \hat{J}_1 + \frac{2(p_1N + n'_1)}{r_1N - n'_1}\hat{J}_1, \quad (20)$$

$$AIC(\mathcal{S}_2) = \hat{J}_2 + \frac{2(p_2N + n'_2)}{r_1N - n'_1}\hat{J}_1. \quad (21)$$

If $AIC(\mathcal{S}_2) < AIC(\mathcal{S}_1)$, the predicting capability is expected to increase by replacing the general model $\mathcal{S}_1$ by the strong model $\mathcal{S}_2$. Recalling that the geometric AIC is an estimator of the expected residual (see Eq. (17)), we compute the square root of the ratio of Eqs. (20) and (21):

$$K = \sqrt{\frac{r_1N - n'_1}{(2p_1 + r_1)N + n'_1}\left(\frac{\hat{J}_2}{\hat{J}_1} + \frac{2(p_2N + n'_2)}{r_1N - n'_1}\right)}. \quad (22)$$

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 $K < 1$, 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.

8. Model Selection for Stereo Vision

We now consider the following three models:

1. *General model*: If we observe N corresponding points $\{\boldsymbol{x}_\alpha, \boldsymbol{x}'_\alpha\}$ and if we do not have any knowledge about the structure of the scene, the constraint is solely the epipolar equation (3), which defines a *three*-dimensional manifold $\mathcal{S}$ in the *four*-dimensional data space $\mathcal{X}$. Since no free parameters are involved, model $\mathcal{S}$ has *zero* degrees of freedom.

2. *Infinitely far away model*: If all the feature points are located infinitely far away, the constraint is given by Eq. (7). Only two of its three component equations are independent, so it defines a *two*-dimensional manifold $\mathcal{S}_\infty$ in $\mathcal{X}$. Since no free parameters are involved, model $\mathcal{S}_\infty$ has *zero* degrees of freedom.

3. *Planar surface model*: If all the feature points are coplanar in the scene, the constraint is given by Eqs. (10). Only two of its three component equations are independent, so it defines a *two*-dimensional manifold $\mathcal{S}_\Pi$ in $\mathcal{X}$. The unknown surface parameters $\{\boldsymbol{n}, d\}$ have *five* degrees of freedom.

Since the general model $\mathcal{S}$ applies irrespective of the structure of the scene, the object is judged to be located infinitely far away if

$$K_\infty = \sqrt{\frac{1}{7}\left(\frac{\hat{J}_\infty}{\hat{J}} + 4\right)} < 1, \quad (23)$$

and to be planar if

$$K_\Pi = \sqrt{\frac{1}{7}\left(\frac{\hat{J}_\Pi}{\hat{J}} + \frac{4N+6}{N}\right)} < 1. \quad (24)$$

The procedure for automatic inference about the scene is summarized as follows:

1. Detect corresponding feature points over the two images by a conventional method.

2. Compute the residual $\hat{J}$ by Eq. (5).

3. Compute the residual $\hat{J}_\infty$ by Eq. (8), and judge

Fig. 3 Real stereo images.

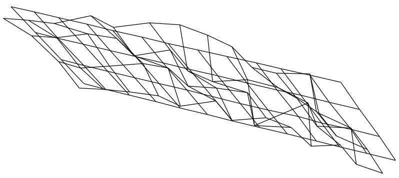

Fig. 4 Reconstructed planar and non-planar shapes.

that the object is located infinitely far away if $K_\infty < 1$.

4. Compute the surface parameters $\{\hat{\boldsymbol{n}}, d\}$ of an optimally fitted plane by renormalization [8].
5. Compute the residual $\hat{J}_\Pi$ by Eq. (11), and judge that the object is planar if $K_\Pi < 1$.

9. Examples

Figure 3 shows real stereo images. By using the corners of the windows as feature points, we can judge that the object is not infinitely far away ($K_\infty = 3.09$) but has a planar shape ($K_\Pi = 0.86$). The 3-D shapes reconstructed by using the general model and the planar surface model are superimposed in Fig. 4. The processing time for the planarity test and the infinity test was 0.7 seconds and 0.3 seconds, respectively, on JCC JS5/70 (SUN SPARC station 5 compatible).

Placing a planar grid in front of the first camera, we simulated a stereo system by translating and rotating the second camera in such a way that the grid was always viewed in the center of the frame (Fig. 5). Gaussian random noise of standard deviation σ pixels was added to the x and y coordinates of each grid point independently and conducted the infinity test 100 times, each time using different noise. Figure 6 shows the percentage of the instances for which the grid is judged to be infinitely far away. We can see that the threshold of the judgment is automatically adjusted to the noise.

Next, we fixed the baseline and defined two planar grids hinged together with angle $\pi - \theta$ (Fig. 7). We conducted the planarity test 100 times for each angle θ, adding random noise in the same way. Figure 8 shows the percentage of the instances for which the grid is

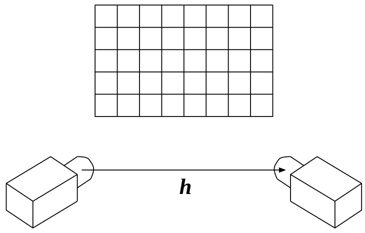

Fig. 5 Infinity test

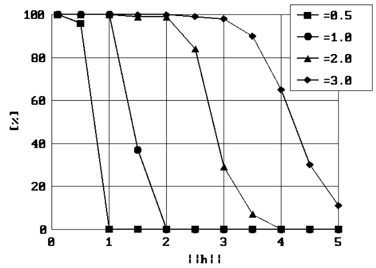

Fig. 6 Percentage of the instances judged to be infinitely far away.

judged to be planar. Again, the threshold is automatically adjusted to the noise.

Figure 9 shows one example of noisy stereo images for $\theta = 20°$ and $\sigma = 1$ pixel, and the object was judged to be planar ($K_\Pi = 0.94$). Figures 10(a) and (b) show the reconstructed shapes by the usual method; Figures 10(c) and (d) show the reconstructed shapes by our method.

10. Concluding Remarks

Introducing a mathematical model of noise in stereo images, we have proposed a new criterion for intelligent statistical inference about the scene we are viewing by using the *geometric information criterion* (*geometric AIC*). Using synthetic and real-image experiments, we have demonstrated that a robot can test whether or not the object is located very far away or the object is a planar surface without using any knowledge about the noise magnitude or any empirically adjustable thresholds.

In this paper, we have considered only the infinitely far away model and the planar surface model as alternatives to the general model, but we can consider various other models such as a quadratic surface model and a polyhedron model.

The principle described here can be applied to *active stereo vision* for automatically adjusting the baseline length so that the resulting disparity is sufficient for reliable 3-D reconstruction. It can also be applied to automatic singularity detection in motion analysis [6], intelligent CAD interface [12], segmentation of curves and surfaces, and many other problems where geomet-

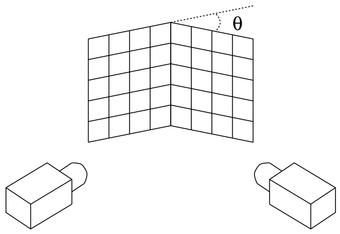

Fig. 7 Planarity test.

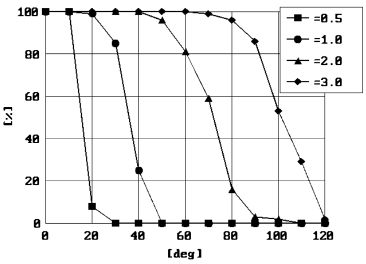

Fig. 8 Percentage of the instances judged to be planar.

ric inference is involved.

References

[1] H. Akaike, "A new look at the statistical model identification," IEEE Trans. Automation Control, vol.19, no.6, pp.176–723, 1974.

[2] N. Ayache and O. Faugeras, "Building, registrating, and fusing noisy visual maps," Int. J. Robotics Res., vol.7, no.6, pp.45–65, 1988.

[3] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Viewpoint, MIT Press, Cambridge, MA, 1993.

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

[5] K. Kanatani, Statistical Optimization for Geometric Computation: Theory and Practice, Elsevier Science, Amsterdam, 1996.

[6] K. Kanatani, "Automatic singularity test for motion analysis by an information criterion," Proc. 4th European Conference on Computer Vision, April, 1996, Cambridge, U.K., pp. 697–708,

[7] K. Kanatani, "Geometric information criterion for model selection," Int. J. Comput. Vis., to appear.

[8] Y. Kanazawa and K. Kanatani, "Direct reconstruction of planar surfaces by stereo vision," IEICE Trans. Inf. & Syst., vol.E78-D, no.7, pp.917–922, 1995.

[9] Y. Kanazawa and K. Kanatani, "Reliability of 3-D reconstruction by stereo vision," IEICE Trans. Inf. & Syst., vol.E78-D, no.10, pp.1301–1306, 1995.

[10] L. Matthies and S. A. Shafer, "Error modeling in stereo navigation," IEEE J. Robot. Autom., vol.3, no.3, pp.239–248, 1987.

[11] L. Matthies, "Stereo vision for planetary rovers: stochastic modeling to near real-time implementation," Int. J. Comput. Vis., vol.8, no.1, pp.71–91, 1992.

[12] I. Triono and K. Kanatani, "Automatic recognition of regular figures by geometric AIC," Proc. IAPR Workshop on Machine Vision Applications, November 1996, Tokyo, Japan, pp.393—396.

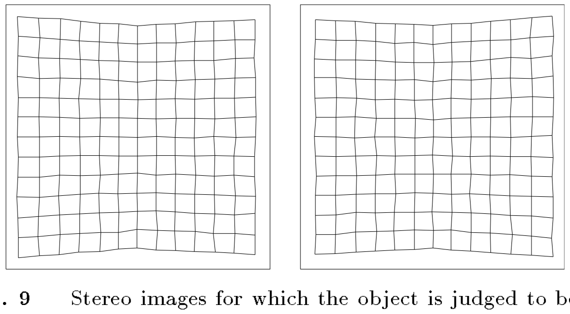

Fig. 9 Stereo images for which the object is judged to be a planar surface.

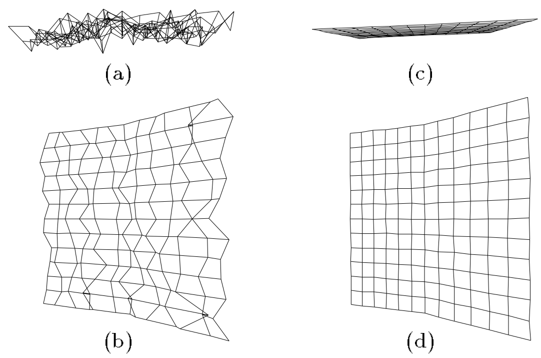

Fig. 10 3-D shape reconstructed from stereo images by the usual method: (a) top view, (b) front view. 3-D shape reconstructed from stereo images by our method: (c) top view, (d) front view.

[13] J. Weng, T. S. Huang and N. Ahuja, Motion and Structure from Image Sequences, Springer, Berlin, 1993.

Yasushi Kanazawa received his B.E. and M.E. degrees in information and computer science from Toyohashi University of Technology, Aichi, Japan, in 1985 and 1987, respectively. He is currently Lecturer of information and computer engineering at Gunma College of Technology. His research interests include image processing and computer vision.

Kenichi Kanatani received his Ph.D. in applied mathematics from the University of Tokyo in 1979. He is currently Professor of computer science at Gunma University. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford University Press, 1993) and *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier Science, 1996).