International Journal of Computer Vision, 26, 1–21 (1998)
© 1998 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Geometric Information Criterion for Model Selection

KENICHI KANATANI

kanatani@cs.gunma-u.ac.jp

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

Received December 21, 1995. Revised March 12, 1996.

Abstract. In building a 3-D model of the environment from image and sensor data, one must fit to the data an appropriate class of models, which can be regarded as a parameterized manifold, or *geometric model*, defined in the data space. In this paper, we present a statistical framework for detecting *degeneracies* of a geometric model by evaluating its predictive capability in terms of the *expected residual* and derive the *geometric AIC*. We show that it allows us to detect singularities in a structure-from-motion analysis without introducing any empirically adjustable thresholds. We illustrate our approach by simulation examples. We also discuss the application potential of this theory for a wide range of computer vision and robotics problems.

Keywords: model selection, degeneracy detection, statistical estimation, AIC, maximum likelihood estimation, structure-from-motion

1. Introduction

1.1. Model Selection and Degeneracy

In order for a robot to operate and navigate autonomously, it must build a 3-D model of the environment from image and sensor data. However, image and sensor data are not necessarily accurate. A basic approach to cope with this difficulty is *fitting* a two- or three-dimensional model consisting of primitives, such as lines, circles, planes, cylinders, and spheres, to the observed data. This requires the following two stages:

- selecting an appropriate class of models;
- fitting it to the data optimally.

These are also necessary in a more abstract level. For example, we can analyze the 3-D structure from motion images by assuming that the images are views of a moving object observed by a stationary camera or a stationary object observed by a moving camera. This assumption is also a "model", which imposes a strong constraint, called the *epipolar constraint*, on the observed image motion. Since the observed images may not completely agree with this model, we optimally *fit* this model to the images.

In the past, the fitting stage has been extensively studied, and many types of fitting schemes have been proposed for various purposes. The structure-from-motion algorithm has also been studied by many researchers. However, a serious difficulty is hidden in model fitting in general—occurrence of *degeneracy*.

Example 1. Suppose we want to fit an ellipse to a sequence of points in the image. If the points are nearly collinear, ellipse fitting breaks down due to numerical instability. So, if the points are nearly collinear, we must decide that they are collinear and fit a line. A line can be fitted robustly unless all the point are very close to each other. If the points are sufficiently clustered, we must decide that they are identical and robustly represent them by, say, their center of gravity (i.e., "point" fitting). However, how can we decide that the points are "nearly" collinear or "closely" clustered? A naive idea is to compare the residuals of least squares. For any point set, however, the residual of ellipse fitting is always not more than the residual of line fitting, which is always

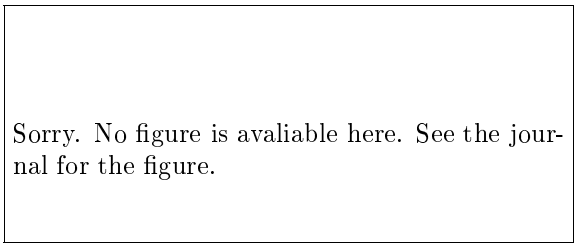

Fig. 1. Class A is never chosen whatever distance measure is used.

not more than the residual of point fitting. This is because *the discrepancy between the data and the model becomes larger as the model is more restricted*; in terms of the residual, an ellipse is always the best fit.

Example 2. As is well known, the algorithm of structure from motion fails if all the feature points are coplanar in the scene. So, we must apply a different algorithm for a planar surface scene. In practice, however, the images have noise, and the general algorithm applied in the presence of noise produces some (unreliable) solution even when the object is a planar surface. How can we detect this degeneracy? Again, the residual does not work; the discrepancy of the data from the general epipolar constraint is always no more than their discrepancy from the planar surface constraint. This is because planar surface constraint *implies* the epipolar constraint and hence is stronger than the epipolar constraint.

In general terms, the problem can be viewed as finding an optimal class for a given input pattern. This is the goal of *pattern recognition*, and usually the one that is the "closest" to the input measured by an appropriate metric is chosen. However, it is tacitly assumed in pattern recognition that the classes into which the input is to be classified are *disjoint*. If class A is *included* in class B, class A is never chosen whatever distance measure is used, because the distance to class A is always not more than that to class B (Fig. 1).

Although model selection that admits degeneracy has been studied by many researchers in the past (Rosin and West 1995; Zabrodsky et al. 1995), this theoretical difficulty has seldom been noticed (Kanatani 1997a, 1997b); usually an ad-hoc switching criterion with empirically adjustable thresholds is introduced. In this paper, we analyze statistical behavior of residuals very carefully and derive a criterion, which we call the *geometric AIC*, for detecting degeneracy in geometric inference for computer vision and robotics applications. Then, we show numerical simulations to illustrate the implications and effectiveness of our approach.

1.2. Geometric Inference vs. Statistics

Model selection is one of the central subjects of statistics. However, there is a wide gap between traditional statistical inference and the geometric inference we are concerned with. The main objective of traditional statistics is to infer the (unknown) *structure* of a random phenomenon by observing multiple data in domains that involve a large degree of uncertainty, such as medicine, biology, agriculture, economics, and politics. Usually, we first assume multiple plausible models as candidates and choose one according to some criterion.

One of the widely adopted criterion is the *MDL* (*minimum description length*) (Rissanen 1984, 1987). The underlying logic is as follows. If the candidate models are parameterized, all models may explain the data to some extent if the parameters are appropriately adjusted. The MDL principle demands that the model should explain the data very well and at the same time have a *simple structure*; if we do not know the true structure and if a simple model can explain the phenomenon, why should we bother to add complexity? Also, the MDL has been found to possess nice asymptotic properties for linear regression models (Barron and Cover 1991, Zhang 1992). Hence, it suits structure-guessing problems such as image segmentation (Gu et al. 1996, Leclerc 1989). However, the resulting model is *hypothetical* whichever model is chosen.

In the geometric inference we are concerned with, the model is not hypothetical; it is *definitive* and determined by a geometrical law (such as the epipolar constraint). What we are interested in is whether or not degeneracy has occurred. For this type of problem, the MDL may not be a good criterion because we need not "guess" an unknown structure. Since our attention is the existence of a *constraint* on the parameters, it is natural to

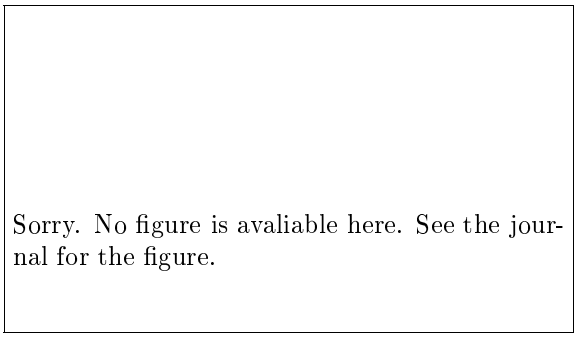

Fig. 2. Estimating a manifold $\mathcal{S} \subset \mathcal{A}$ and $\{\bar{\mathbf{a}}_\alpha\} \in \mathcal{S}$ from the data $\{\mathbf{a}_\alpha\}$.

base the inference process primarily on the *residual*, i.e., the discrepancy between the data and the model.

In the domain of traditional statistical estimation, a *statistical model* is a combination of a deterministic mechanism and random noise; estimating the noise behavior is almost equivalent to estimating the structure of the phenomenon. In most cases, the parameters that characterize the noise and the parameters that specify the deterministic mechanism are estimated simultaneously. A *geometric model*, on the other hand, merely states the existence of a definitive and exact constraint; noise is a characteristic of the measurement devices and data processing operations. In other words, while noise is the heart of a statistical model to be inferred about, noise in geometric inference is an external agency independent of a geometric model. It follows that in comparing two geometric models we must take into account the fact that *the noise is identical (but unknown) and has the same characteristics for both models.*

From these observation, we adopt a residual-based analysis that underlies the *AIC* (*Akaike information criterion*) (Akaike 1974). Since the original AIC is defined in the framework of traditional statistical estimation, it cannot be used for the geometric inference we are concerned with in its original form. In this paper, we define a new version, which we call the *geometric AIC*, by following the reasoning laid out by Akaike. We illustrate our theory by simple examples and discuss its application potential for a wide range of computer vision and robotics problems.

2. Geometric Models

2.1. Definition

Let $\mathbf{a}_1$, ..., $\mathbf{a}_N$ be m-dimensional vector data sampled from an m'-dimensional manifold $\mathcal{A} \subset \mathcal{R}^m$ ($\mathcal{R}^m$ denotes the m dimensional Cartesian space, i.e., the set of all m-dimensional vectors); we call $\mathcal{A}$ the *data space*. We write

$$\mathbf{a}_\alpha = \bar{\mathbf{a}}_\alpha + \Delta\mathbf{a}_\alpha, \tag{1}$$

where $\bar{\mathbf{a}}_\alpha$ is the true position of datum $\mathbf{a}_\alpha$. The noise term $\Delta\mathbf{a}_\alpha$ is assumed to be an independent Gaussian random variable of mean $\mathbf{0}$ and covariance matrix $V[\mathbf{a}_\alpha]$. Since $\mathbf{a}_\alpha$ is constrained to be in $\mathcal{A}$, the covariance matrix $V[\mathbf{a}_\alpha]$ is generally singular. We assume that $V[\mathbf{a}_\alpha]$ has range $T_{\bar{\mathbf{a}}_\alpha}(\mathcal{A})$ (= the tangent space to $\mathcal{A}$ at $\bar{\mathbf{a}}_\alpha$) and hence has rank m'. We want to find a manifold $\mathcal{S} \subset \mathcal{A}$ such that the true values $\bar{\mathbf{a}}_\alpha$ are all in it. This problem is stated as follows (Fig. 2):

PROBLEM 1 *Estimate a manifold $\mathcal{S} \subset \mathcal{A}$ and m-dimensional vectors $\{\bar{\mathbf{a}}_\alpha\} \in \mathcal{S}$ from the data $\{\mathbf{a}_\alpha\}$.*

In order to solve this problem, we parameterize the manifold $\mathcal{S}$ by an n-dimensional vector $\mathbf{u}$ constrained to be in an n'-dimensional manifold $\mathcal{U} \subset \mathcal{R}^n$; we call $\mathcal{U}$ the *parameter space*. It follows that the manifold $\mathcal{S}$ has n' degrees of freedom. We call a parameterized manifold a (*geometric*) *model*. If it is a d-dimensional manifold in an m'-dimensional data space and has n' free parameters, we say that the model has *dimension* d, *codimension* $r = m' - d$, and n' *degrees of freedom*.

Example 3. Suppose points are observed in three dimensions. The following are typical models for them:

1. *Point model*: their true positions all coincide (Fig. 3(a)). This model consists of a single point P; it has dimension 0, codimension 3, and three degrees of freedom. If it is constrained to be at the coordinate origin, the constrained model P' has zero degrees of freedom.
2. *Line model*: their true positions are all collinear (Fig. 3(b)). This model consists of a line L; it has dimension 1, codimension 2, and four degrees of freedom. If it is constrained to pass

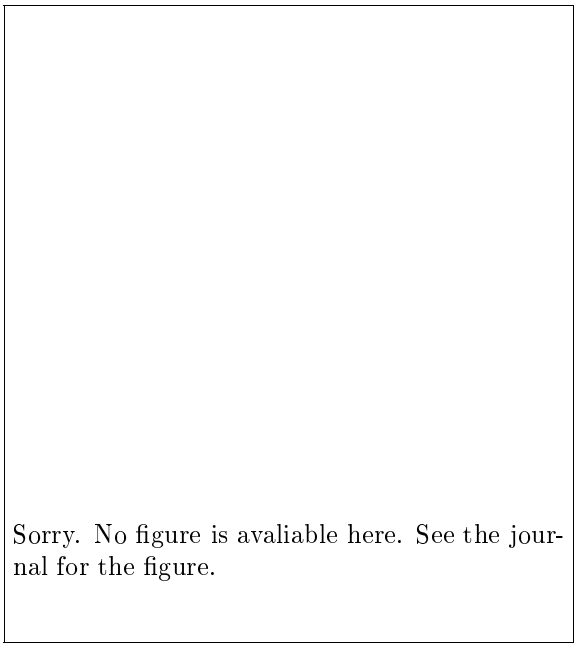

Fig. 3. (a) Point model P. (b) Line model L. (c) Plane model Π. (d) Space model $\mathcal{R}^3$.

through the coordinate origin, the constrained model L' has two degrees of freedom.

3. *Plane model*: their true positions are all coplanar (Fig. 3(c)). This model consists of a plane Π; it has dimension 2, codimension 1, and three degrees of freedom. If it is constrained to pass through the coordinate origin, the constrained model Π' has two degree of freedom.
4. *Space model*: no constraint is imposed on their true positions (Fig. 3(d)). This model consists of the entire space $\mathcal{R}^3$; it has dimension 3, codimension 0, and zero degrees of freedom.

2.2. Expected Residual

We define the *Mahalanobis norm* of an m-dimensional vector $\mathbf{v}$ with respect to $V[\mathbf{a}_\alpha]$ by

$$\|\mathbf{v}\|_\alpha = \sqrt{(\mathbf{v}, V[\mathbf{a}_\alpha]^-\mathbf{v})}, \tag{2}$$

where $(\cdot)^-$ denotes the (*Moore-Penrose*) *generalized inverse*. The proviso "with respect to $V[\mathbf{a}_\alpha]$" is omitted when understood.

Remark 1. A p-dimensional positive semi-definite symmetric matrix $\mathbf{A}$ can be decomposed into the form $\mathbf{A} = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$ (*spectral decomposition*), where $\{\mathbf{v}_i\}$ is an orthonormal system of the eigenvectors of $\mathbf{A}$ for eigenvalues $\{\lambda_i\}$. If $\lambda_1 \geq \cdots \geq \lambda_m > \lambda_{m+1} = \cdots = \lambda_p = 0$, the (*Moore-Penrose*) *generalized inverse* of $\mathbf{A}$ is given by $\mathbf{A}^- = \sum_{i=1}^{m} \mathbf{v}_i \mathbf{v}_i^\top / \lambda_i$.

Suppose a particular model $\mathcal{S}$ is given. Since the noise is assumed to be Gaussian, the *maximum likelihood estimator* of $\mathcal{S}$ and $\{\bar{\mathbf{a}}_\alpha\}$ for Problem 1 is obtained by minimizing $\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \mathbf{x}_\alpha\|_\alpha^2$ under the constraint that $\{\mathbf{x}_\alpha\} \in \mathcal{S}$. Let $\hat{\mathcal{S}}$ and $\{\hat{\mathbf{a}}_\alpha\}$ be the resulting maximum likelihood estimators of $\mathcal{S}$ and $\{\bar{\mathbf{a}}_\alpha\}$ ($\in \hat{\mathcal{S}}$), respectively. The residual

$$\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2 \tag{3}$$

measures how well the assumed model $\mathcal{S}$ fits the data $\{\mathbf{a}_\alpha\}$. This appears to be a suitable measure of the goodness of the model. However, because $\hat{\mathcal{S}}$ and $\{\hat{\mathbf{a}}_\alpha\}$ are determined so as to minimize the residual *for the current data* $\{\mathbf{a}_\alpha\}$, the residual can be made arbitrarily small by assuming a loose model. In fact, if the manifold $\mathcal{S}$ has a sufficient number of parameters, we can make $\mathcal{S}$ pass through all the data $\{\mathbf{a}_\alpha\}$ by adjusting the parameters. Such an artificial model may explain the current data very well but may be unable to predict the occurrence the data to be observed in the future.

Example 4. Recall Example 3. If we assume the space model $\mathcal{R}^3$, datum $\mathbf{a}_\alpha$ itself is an optimal estimator of $\bar{\mathbf{a}}_\alpha$, i.e., $\hat{\mathbf{a}}_\alpha = \mathbf{a}_\alpha$. Hence, $\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2 = 0$.

On the other hand, the square sum of estimation errors

$$\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2 \tag{4}$$

measures how accurate the estimators $\{\hat{\mathbf{a}}_\alpha\}$ are, *provided the model is correct*. This may be a suitable measure of the goodness of the "estimator" but cannot be used as a measure of the goodness of the assumed model. In fact, it can be made arbitrarily small by assuming a restrictive model; as

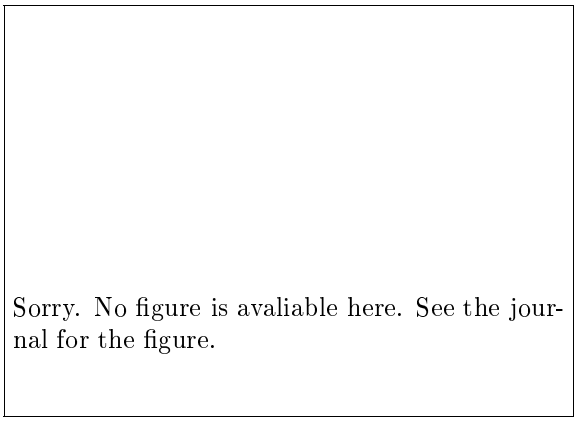

Fig. 4. The Mahalanobis projection of $\mathbf{a}_\alpha$ onto manifold $\mathcal{S}$ is the tangent point of the equilikelihood surface to $\mathcal{S}$.

the model is more restricted, the deviation $\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha$ is constrained in a smaller set.

Example 5. Recall Example 3. If we assume the constrained point model P', an optimal estimator of $\bar{\mathbf{a}}_\alpha$ is $\mathbf{0}$ irrespective of observed value, i.e., $\hat{\mathbf{a}}_\alpha = \mathbf{0}$. Hence, $\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2 = 0$.

Thus, the goodness of a model cannot be measured by the goodness of fitting or the accuracy of estimation. What we propose here is to evaluate the goodness of a model by its *predicting capacity*: we consider *future data* $\{\mathbf{a}_\alpha^*\}$ *that would be observed if the noise occurred differently under the same model*. In other words, we require that the residual for the future data

$$\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha^* - \hat{\mathbf{a}}_\alpha\|_\alpha^2 \tag{5}$$

be small. Since this is a random variable, we take expectation to define a definitive value for the model:

$$I(\mathcal{S}) = E^*[E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha^* - \hat{\mathbf{a}}_\alpha\|_\alpha^2]]. \tag{6}$$

Here, $E^*[\cdot]$ and $E[\cdot]$ denote expectation with respect to the future data $\{\mathbf{a}_\alpha^*\}$ and the current data $\{\mathbf{a}_\alpha\}$, respectively; $\{\mathbf{a}_\alpha^*\}$ and $\{\mathbf{a}_\alpha\}$ are assumed to be independent and subject to the same distribution. We call $I(\mathcal{S})$ the *expected residual* of model $\mathcal{S}$ and regard $\mathcal{S}$ as good if $I(\mathcal{S})$ is small.

This idea was first introduced by Akaike (1969) in relation to an autoregressive model and then extended to general statistical inference for estimating the parameters of a probability distribution from which the data are sampled (Akaike 1974). Akaike considered the expectation of the logarithmic likelihood with respect to the future data and derived a criterion called the *AIC* (*Akaike information criterion*), which has been widely used in many applications of statistical inference. It can be shown that the AIC can be viewed as an approximation to the entropy crierion (the *Kullback-Leibler information*) (Kullback 1959). Following a reasoning similar to that of Akaike, we now prove the following theorem.

THEOREM 1

$$I(\mathcal{S}) = E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2] + 2(dN + n'). \tag{7}$$

Theorem 1 implies that *the future residual (5) is larger than the current residual (3) by* $2(dN + n')$ *in expectation*. Since the residual (3) measures the goodness of the fit, the additional term $2(dN + n')$ can be interpreted to be the *penalty* for the demension of the model and its degree of freedom. In other words, when we minimize the residual (3), we have to pay a penalty for using a model with a high dimension and a large degree of freedom. This observation leads to the following definition of the *geometric information criterion*, or *geometric AIC* for short:

$$AIC(\mathcal{S}) = \sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2 + 2(dN + n'). \tag{8}$$

This is an unbiased estimator of the expected residual $I(\mathcal{S})$, and we use this as a measure for the goodness of the model: if $AIC(\mathcal{S}_1) < AIC(\mathcal{S}_2)$ for models $\mathcal{S}_1$ and $\mathcal{S}_2$, we prefer model $\mathcal{S}_1$ because $\mathcal{S}_1$ is expected to have more predicting capacity than model $\mathcal{S}_2$. We give some mathematical preliminaries in Section 3 and prove Theorem 1 in Section 4.

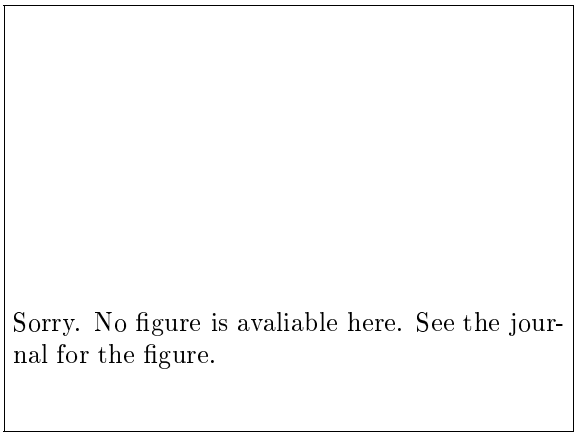

Fig. 5. The relation between datum $\mathbf{a}_\alpha$, its true value $\bar{\mathbf{a}}_\alpha \in \mathcal{S}$ and the Mahalanobis projection $\hat{\mathbf{a}}_\alpha$ of $\mathbf{a}_\alpha$ onto $\mathcal{S}$.

3. Mahalanobis Geometry

3.1. Mahalanobis Projection

First, we consider the case in which the model $\mathcal{S}$ has no free parameters ($n' = 0$). Since the noise is assumed to be Gaussian, the maximum likelihood estimator $\hat{\mathbf{a}}_\alpha$ of $\bar{\mathbf{a}}_\alpha$ is the *tangent point* of the *equilikelihood surface* $(\mathbf{a}-\mathbf{a}_\alpha, V[\mathbf{a}_\alpha]^-(\mathbf{a}-\mathbf{a}_\alpha))$ = constant, $\mathbf{a} \in \mathcal{A}$, to the manifold $\mathcal{S}$ (Fig. 4). Let us call $\hat{\mathbf{a}}_\alpha$ the *Mahalanobis projection* of $\mathbf{a}_\alpha$ onto $\mathcal{S}$ with respect to $V[\mathbf{a}_\alpha]$. The proviso "with respect to $V[\mathbf{a}_\alpha]$" is omitted when understood from the context.

We assume that the *noise is sufficiently small*, by which we mean that $V[\mathbf{a}_\alpha] = O(\epsilon^2)$, $\alpha = 1, ..., N$, for an appropriately defined noise level ϵ, as compared with which the data space $\mathcal{A}$ and the manifold $\mathcal{S}$ are assumed to be sufficiently smooth. Intuitively, this means that ϵ is small as compared with a quantity that measures the "radius of curvature" for the manifolds $\mathcal{A}$ and $\mathcal{S}$. In the following, we understand that we ignore high order effects due to the "curvatures" of $\mathcal{A}$ and $\mathcal{S}$ and omit the proviso "in the first order approximation".

The following proposition is easily obtained (Kanatani 1996a):

PROPOSITION 1

$$\|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2 = \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2 + \|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2. \tag{9}$$

This can be interpreted as the *Pythagoras theorem*, stating that three points $\mathbf{a}_\alpha$, $\hat{\mathbf{a}}_\alpha$, and $\bar{\mathbf{a}}_\alpha$ define a "right-angled triangle," where the length is measured in the Mahalanobis norm and the orthogonality of vectors $\mathbf{u}$ and $\mathbf{v}$ is defined by $(\mathbf{u}, V[\mathbf{a}_\alpha]^-\mathbf{v}) = 0$ (Fig. 5). In this sense, $\hat{\mathbf{a}}_\alpha$ is the "foot" of the "perpendicular line" drawn from $\mathbf{a}_\alpha$ to $\mathcal{S}$.

PROPOSITION 2

$$\begin{aligned} E[\|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] &= m', \\ E[\|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2] &= r, \\ E[\|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] &= d. \end{aligned} \tag{10}$$

Proof: By definition, $\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha$ is a Gaussian random variable of mean $\mathbf{0}$ and covariance matrix $V[\mathbf{a}_\alpha]$ of rank m'. Hence, $\|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2$ is subject to a χ^2 distribution with m' degrees of freedom. On the other hand, $\|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2$ is subject to a χ^2 distribution with r degrees of freedom (Kanatani 1996a). Since $\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha$ is obtained from $\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha$ by a linear mapping, $\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha = (\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha) - (\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha)$ is subject to a χ^2 distribution with $m' - r$ $(= d)$ degrees of freedom (Kanatani 1996a). Since the expectation of a χ^2 distribution is equal to its degree of freedom, we obtain eqs. (10). □

3.2. Residual of Model Fitting

Now, we consider the case where the model $\mathcal{S}$ has n' free parameters. Let $\hat{\mathcal{S}}$ be its maximum likelihood estimator. We assume that this model contains the true manifold $\bar{\mathcal{S}}$. Let $\tilde{\mathbf{a}}_\alpha$ and $\hat{\mathbf{a}}_\alpha$ be the Mahalanobis projections of $\mathbf{a}_\alpha$ onto $\bar{\mathcal{S}}$ and $\hat{\mathcal{S}}$, respectively (Fig. 6). Since each datum $\mathbf{a}_\alpha$ is independent, $\sum_{\alpha=1}^N \|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2$ and $\sum_{\alpha=1}^N \|\mathbf{a}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2$ are subject to χ^2 distributions with $m'N$ and rN degrees of freedom, respectively. In other words, the degrees of freedom of the residual decrease from $m'N$ to rN if the true value $\bar{\mathbf{a}}_\alpha$ is replaced by the Mahalanobis projection $\tilde{\mathbf{a}}_\alpha$. However, if $\mathbf{a}_\alpha$ is projected onto the the maximum likelihood estimator $\hat{\mathcal{S}}$, the degrees of freedom of the residual further decrease by n' (Kanatani 1996a) (Fig. 6). Namely, the residual $\sum_{\alpha=1}^N \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2$ is subject to a χ^2 distribution with $rN - n'$ degrees of freedom. Thus, we obtain

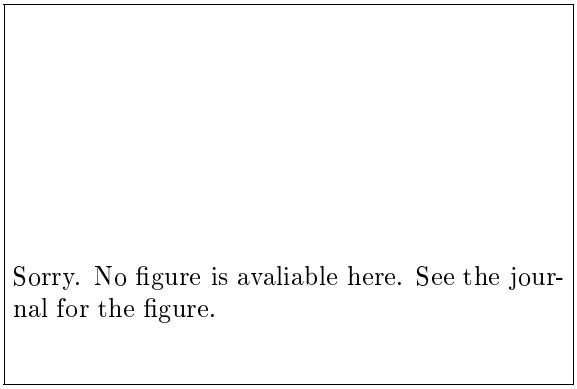

Fig. 6. The Mahalanobis projections $\tilde{\mathbf{a}}_\alpha$ and $\hat{\mathbf{a}}_\alpha$ of datum $\mathbf{a}_\alpha$ onto the true manifold $\bar{\mathcal{S}}$ and the the optimally fitted manifold $\hat{\mathcal{S}}$, respectively.

PROPOSITION 3

$$E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] = m'N,$$
$$E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2] = rN,$$
$$E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2] = rN - n'. \tag{11}$$

4. Expected Residual

4.1. Evaluation of the Expected Residual

We now evaluate the expected residual $I(\mathcal{S})$ defined by eq. (6). Let $\tilde{\mathbf{a}}_\alpha$ and $\hat{\mathbf{a}}_\alpha$ be the Mahalanobis projections of $\mathbf{a}_\alpha$ onto $\bar{\mathcal{S}}$ and $\hat{\mathcal{S}}$, respectively. Let $\mathbf{a}_\alpha^*$ be the future datum corresponding to $\mathbf{a}_\alpha$ (Fig. 7).

LEMMA 1

$$I(\mathcal{S}) = E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2] + E[\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2] + 2dN + n'. \tag{12}$$

Proof: Since the maximum likelihood estimators $\{\hat{\mathbf{a}}_\alpha\}$ are determined from the current data $\{\mathbf{a}_\alpha\}$, they are independent of the future data $\{\mathbf{a}_\alpha^*\}$.

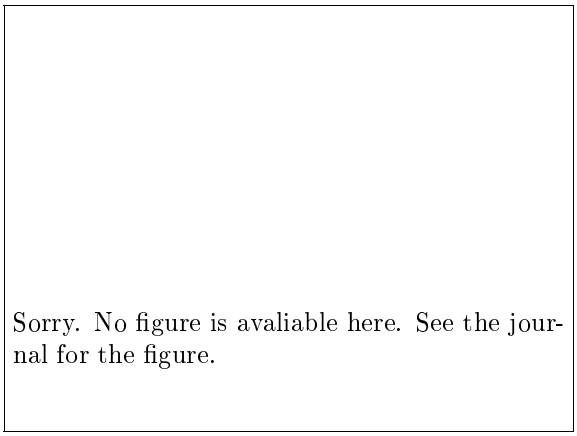

Fig. 7. Evaluation of the expected residual.

Hence, eq. (6) reduces to

$$\begin{aligned} I(\mathcal{S}) &= E^*[E[\sum_{\alpha=1}^{N} \|(\mathbf{a}_\alpha^* - \bar{\mathbf{a}}_\alpha) \\ &\quad -(\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha)\|_\alpha^2]] \\ &= E^*[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha^* - \bar{\mathbf{a}}_\alpha\|_\alpha^2] \\ &\quad +E[\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2]. \end{aligned} \tag{13}$$

Since $\{\mathbf{a}_\alpha^*\}$ and $\{\mathbf{a}_\alpha\}$ have the same distribution, we have $E^*[\|\mathbf{a}_\alpha^* - \bar{\mathbf{a}}_\alpha\|_\alpha^2] = E[\|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2]$. So, eq. (13) can be written as

$$\begin{aligned} I(\mathcal{S}) &= E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] \\ &\quad +E[\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2]. \end{aligned} \tag{14}$$

Consider the first term on the right-hand side. Applying Proposition 1, we can write

$$\begin{aligned} &E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] \\ &= E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2] \\ &\quad +E[\sum_{\alpha=1}^{N} \|\tilde{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2]. \end{aligned} \tag{15}$$

From Proposition 3, we have

$$E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2] = E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2] + n'. \quad (16)$$

We also have

$$E[\sum_{\alpha=1}^{N} \|\tilde{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] = dN. \quad (17)$$

Hence, eq. (15) reduces to

$$E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2] = E[\sum_{\alpha=1}^{N} \|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2] + dN + n'. \quad (18)$$

Next, consider the second term on the right-hand side of eq. (14). If the noise is small, the Mahalanobis projection of $\hat{\mathbf{a}}_\alpha \in \hat{\mathcal{S}}$ onto the true manifold $\bar{\mathcal{S}}$ coincides to a first approximation with the Mahalanobis projection $\tilde{\mathbf{a}}_\alpha$ of the datum $\mathbf{a}_\alpha$ onto $\bar{\mathcal{S}}$ (Fig. 7). Hence, we obtain from Proposition 1 and eq. (17)

$$\begin{aligned}\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2 &= \sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2 + \sum_{\alpha=1}^{N} \|\tilde{\mathbf{a}}_\alpha - \bar{\mathbf{a}}_\alpha\|_\alpha^2 \\ &= \sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2 + dN. \end{aligned} \quad (19)$$

Substituting eqs. (18) and (19) into eq. (14), we obtain eq. (12). □

4.2. Accuracy of Parametric Fitting

In order to evaluate the second term on the right-hand side of eq. (12), we need an explicit representation of the model $\mathcal{S}$. Suppose the model $\mathcal{S}$ is represented by L equations in the form

$$F^{(k)}(\mathbf{a}, \mathbf{u}) = 0, \qquad k = 1, ..., L, \quad (20)$$

parameterized by an n-dimensional vector $\mathbf{u}$ constrained to be in an n'-dimensional manifold $\mathcal{U} \subset \mathcal{R}^n$, which we call the *parameter space*. The L equations need not be algebraically independent; we call the number r of independent equations the *rank* of eq. (20). We assume that eq. (20) is *nonsingular* in the sense that each of the L equations defines a manifold of codimension 1 in the m'-dimensional data space $\mathcal{A}$ such that they intersect with each other *transversally* (Kanatani 1996a) (Fig. 8). This assumption ensures that eq. (20) is a manifold of codimension r.

Let $\bar{\mathbf{u}}$ be the true value of $\mathbf{u}$, i.e., the value that realizes the true manifold $\bar{\mathcal{S}}$. The *moment matrix* is defined as follows (Kanatani 1996a):

$$\bar{\mathbf{M}} = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} \bar{W}_\alpha^{(kl)} (\mathbf{P}_{\bar{\mathbf{u}}}^{\mathcal{U}} \nabla_{\mathbf{a}} \bar{F}_\alpha^{(k)}) (\mathbf{P}_{\bar{\mathbf{u}}}^{\mathcal{U}} \nabla_{\mathbf{a}} \bar{F}_\alpha^{(l)})^\top, \quad (21)$$

$$(\bar{W}_\alpha^{(kl)}) = \Big((\nabla_{\mathbf{a}} \bar{F}_\alpha^{(k)}, V[\mathbf{a}_\alpha] \nabla_{\mathbf{a}} \bar{F}_\alpha^{(l)}) \Big)^- . \quad (22)$$

Here, $\mathbf{P}_{\bar{\mathbf{u}}}^{\mathcal{U}}$ is the n-dimensional projection matrix onto the tangent space $T_{\bar{\mathbf{u}}}(\mathcal{U})$ to the manifold $\mathcal{U}$ at $\bar{\mathbf{u}}$. The symbol $\nabla_{\mathbf{a}}(\,\cdot\,)$ denotes the vector whose ith component is $\partial(\,\cdot\,)/\partial a_i$, where a_i is the ith component of $\mathbf{a}$, and $\nabla_{\mathbf{a}} \bar{F}_\alpha^{(k)}$ is the abbreviation of $\nabla_{\mathbf{a}} F^{(k)}(\bar{\mathbf{a}}_\alpha, \bar{\mathbf{u}})$. In the following, we use an abbreviated notation to define a matrix: $(\bar{W}_\alpha^{(kl)})$ denotes an L-dimensional matrix whose (kl) element is $\bar{W}_\alpha^{(kl)}$; the right-hand side of eq. (22) denotes the generalized inverse of the L-dimensional matrix whose (kl) element is $(\nabla_{\mathbf{a}} \bar{F}_\alpha^{(k)}, V[\mathbf{a}_\alpha] \nabla_{\mathbf{a}} \bar{F}_\alpha^{(l)})$.

Remark 2. Let $\mathcal{V}$ be a q-dimensional linear subspace of $\mathcal{R}^p$ ($q \leq q$). The (*orthogonal*) *projection* $\mathbf{P}^{\mathcal{V}}$ of $\mathcal{R}^p$ onto $\mathcal{V}$ is a p-dimensional matrix such that $\mathbf{P}^{\mathcal{V}}\mathbf{r} \in \mathcal{V}$ and $\mathbf{r} - \mathbf{P}^{\mathcal{V}}\mathbf{r} \in \mathcal{V}^\perp$ for any $\mathbf{r} \in \mathcal{R}^p$. It is symmetric ($(\mathbf{P}^{\mathcal{V}})^\top = \mathbf{P}^{\mathcal{V}}$), idempotent ($(\mathbf{P}^{\mathcal{V}})^2 = \mathbf{P}^{\mathcal{V}}$), and positive semi-definite with rank q. If $\{\mathbf{v}_i\}$, $i = 1, ..., q$, is an orthonormal basis of $\mathcal{V}$, the projection $\mathbf{P}^{\mathcal{V}}$ has the expression $\mathbf{P}^{\mathcal{V}} = \sum_{i=1}^{q} \mathbf{v}_i \mathbf{v}_i^\top$.

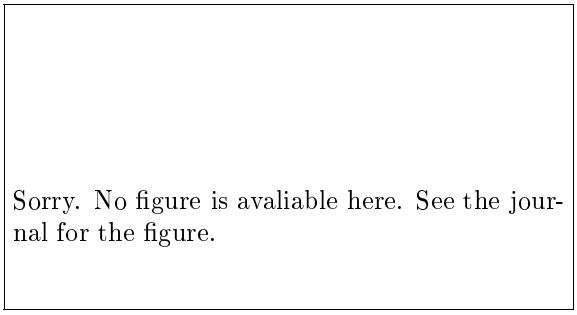

Fig. 8. (a) Two surfaces intersecting transversally. (b) Two surfaces meeting non-transversally.

Let $\hat{\mathbf{u}}$ be the maximum likelihood estimator of $\mathbf{u}$, i.e., the value of $\mathbf{u}$ that realizes the maximum likelihood estimator $\hat{\mathcal{S}}$ of $\mathcal{S}$.

LEMMA 2

$$\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2 = (\hat{\mathbf{u}} - \bar{\mathbf{u}}, \bar{\mathbf{M}}(\hat{\mathbf{u}} - \bar{\mathbf{u}})). \tag{23}$$

Proof: The Mahalanobis projection $\tilde{\mathbf{a}}_\alpha$ of $\hat{\mathbf{a}}_\alpha \in \hat{\mathcal{S}}$ onto $\bar{\mathcal{S}}$ is given as follows (Kanatani 1996a):

$$\tilde{\mathbf{a}}_\alpha = \hat{\mathbf{a}}_\alpha - V[\mathbf{a}_\alpha] \sum_{k,l=1}^{L} \tilde{W}_\alpha^{(kl)} F^{(k)}(\hat{\mathbf{a}}_\alpha, \bar{\mathbf{u}}) \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(l)}, \tag{24}$$

$$(\tilde{W}_\alpha^{(kl)}) = \left((\nabla_{\mathbf{a}} \tilde{F}_\alpha^{(k)}, V[\mathbf{a}_\alpha] \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(l)}) \right)^{-}. \tag{25}$$

Here, $\nabla_{\mathbf{a}} \tilde{F}_\alpha^{(k)}$ is the abbreviation of $\nabla_{\mathbf{a}} F^{(k)}(\tilde{\mathbf{a}}_\alpha, \bar{\mathbf{u}})$. Since $\hat{\mathbf{a}}_\alpha \in \hat{\mathcal{S}}$, we have

$$F^{(k)}(\hat{\mathbf{a}}_\alpha, \hat{\mathbf{u}}) = 0, \quad k = 1, ..., L. \tag{26}$$

Letting $\Delta\hat{\mathbf{u}} = \hat{\mathbf{u}} - \bar{\mathbf{u}}$ and taking a linear approximation, we obtain

$$\begin{aligned} F^{(k)}(\hat{\mathbf{a}}_\alpha, \bar{\mathbf{u}}) &= F^{(k)}(\hat{\mathbf{a}}_\alpha, \hat{\mathbf{u}} - \Delta\hat{\mathbf{u}}) \\ &= -(\nabla_{\mathbf{u}} \hat{F}_\alpha^{(k)}, \Delta\hat{\mathbf{u}}), \end{aligned} \tag{27}$$

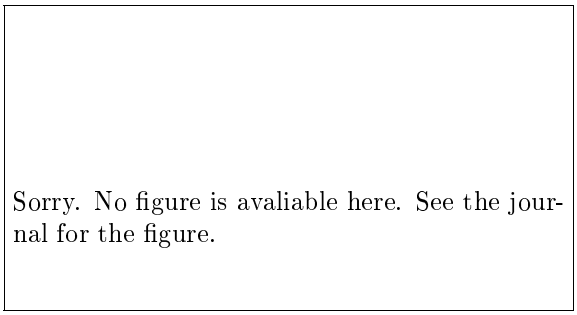

Fig. 9. Accuracy of the optimal fitting

where $\nabla_{\mathbf{u}}$ denotes differentiation with respect to $\mathbf{u}$, and $\nabla_{\mathbf{u}} \hat{F}_\alpha^{(k)}$ is the abbreviation of $\nabla_{\mathbf{u}} F^{(k)}(\hat{\mathbf{a}}_\alpha, \hat{\mathbf{u}})$. Since eq. (27) is a linear approximation in $\Delta\hat{\mathbf{u}}$, we can replace $\nabla_{\mathbf{u}} \hat{F}_\alpha^{(k)}$ by $\nabla_{\mathbf{u}} \bar{F}_\alpha^{(k)}$ (= the abbreviation of $\nabla_{\mathbf{u}} F^{(k)}(\bar{\mathbf{a}}_\alpha, \bar{\mathbf{u}})$) with errors of $O(\Delta\hat{\mathbf{u}})^2$ (Fig. 9). Consequently, eq. (24) is written to a first approximation as

$$\hat{\mathbf{a}}_\alpha - \tilde{\mathbf{a}}_\alpha = V[\mathbf{a}_\alpha] \sum_{k,l=1}^{L} \tilde{W}_\alpha^{(kl)} \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(k)} \nabla_{\mathbf{u}} \bar{F}_\alpha^{(l)\top} \Delta\hat{\mathbf{u}}. \tag{28}$$

From this, we see that

$$\begin{aligned} &\sum_{\alpha=1}^{N} \|\hat{\mathbf{a}}_\alpha - \tilde{\mathbf{a}}_\alpha\|_\alpha^2 \\ &= \sum_{\alpha=1}^{N} (V[\mathbf{a}_\alpha] \sum_{k,l=1}^{L} \tilde{W}_\alpha^{(kl)} \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(k)} \nabla_{\mathbf{u}} \bar{F}_\alpha^{(l)\top} \\ &\quad \Delta\hat{\mathbf{u}}, V[\mathbf{a}_\alpha]^{-} V[\mathbf{a}_\alpha] \\ &\quad \sum_{m,n=1}^{L} \tilde{W}_\alpha^{(mn)} \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(m)} \nabla_{\mathbf{u}} \bar{F}_\alpha^{(n)\top} \Delta\hat{\mathbf{u}}) \\ &= (\Delta\hat{\mathbf{u}}, \left(\sum_{\alpha=1}^{N} \sum_{k,l,m,n=1}^{L} \tilde{W}_\alpha^{(kl)} \tilde{W}_\alpha^{(mn)} \right. \\ &\quad \nabla_{\mathbf{u}} \bar{F}_\alpha^{(l)} \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(k)\top} V[\mathbf{a}_\alpha] V[\mathbf{a}_\alpha]^{-} V[\mathbf{a}_\alpha] \\ &\quad \left. \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(m)} \nabla_{\mathbf{u}} \bar{F}_\alpha^{(n)\top} \right) \Delta\hat{\mathbf{u}}) \\ &= (\Delta\hat{\mathbf{u}}, \left(\sum_{\alpha=1}^{N} \sum_{l,n=1}^{L} \left(\sum_{k,m=1}^{L} \tilde{W}_\alpha^{(lk)} \right. \right. \\ &\quad \left. \left. (\nabla_{\mathbf{a}} \tilde{F}_\alpha^{(k)}, V[\mathbf{a}_\alpha] \nabla_{\mathbf{a}} \tilde{F}_\alpha^{(m)}) \tilde{W}_\alpha^{(mn)} \right) \right) \end{aligned}$$

$$\nabla_{\mathbf{u}}\bar{F}_\alpha^{(l)}\nabla_{\mathbf{u}}\bar{F}_\alpha^{(n)\top}\Big)\Delta\hat{\mathbf{u}})$$
$$= (\Delta\hat{\mathbf{u}}, \Big(\sum_{\alpha=1}^{N}\sum_{l,n=1}^{L}\tilde{W}_\alpha^{(ln)}$$
$$\nabla_{\mathbf{u}}\bar{F}_\alpha^{(l)}\nabla_{\mathbf{u}}\bar{F}_\alpha^{(n)\top}\Big)\Delta\hat{\mathbf{u}}), \tag{29}$$

where we have used the defining equation (25) of the L-dimensional matrix $\tilde{\mathbf{W}}_\alpha = (\tilde{W}_\alpha^{(kl)})$ and the identities $V[\mathbf{a}_\alpha]V[\mathbf{a}_\alpha]^-V[\mathbf{a}_\alpha] = V[\mathbf{a}_\alpha]$ and $\tilde{\mathbf{W}}_\alpha\tilde{\mathbf{W}}_\alpha^-\tilde{\mathbf{W}}_\alpha = \tilde{\mathbf{W}}_\alpha$. To a first approximation, $\tilde{W}_\alpha^{(kl)}$ can be replaced by $\bar{W}_\alpha^{(kl)}$ defined by eq. (22). Since we have $\Delta\hat{\mathbf{u}} \in T_{\bar{\mathbf{u}}}(\mathcal{U})$ to a first approximation and hence $\mathbf{P}_{\bar{\mathbf{u}}}^{\mathcal{U}}\Delta\hat{\mathbf{u}} = \Delta\hat{\mathbf{u}}$, we can write eq. (29) as $(\Delta\hat{\mathbf{u}}, \bar{\mathbf{M}}\Delta\hat{\mathbf{u}})$. □

Proof of Theorem 1: The covariance matrix of the maximum likelihood estimator $\hat{\mathbf{u}}$ is given to a first approximation by the *Cramer-Rao lower bound* defined as the generalized inverse of $\bar{\mathbf{M}}$ (Kanatani 1996a):

$$\bar{V}[\hat{\mathbf{u}}] = \bar{\mathbf{M}}^-. \tag{30}$$

Since this matrix generally has rank n', the quadratic form

$$(\hat{\mathbf{u}} - \bar{\mathbf{u}}, \bar{\mathbf{M}}(\hat{\mathbf{u}} - \bar{\mathbf{u}}))$$
$$= (\hat{\mathbf{u}} - \bar{\mathbf{u}}, \bar{V}[\hat{\mathbf{u}}]^-(\hat{\mathbf{u}} - \bar{\mathbf{u}})) \tag{31}$$

is subject to a χ^2 distribution with n' degrees of freedom (Kanatani 1996a). Hence, its expectation is n, and Lemmas 1 and 2 imply eq. (7). □

5. Geometric Information Criterion

5.1. Model Selection by AIC

Eq. (8) is not a convenient form for actual applications, because computing the residual $\sum_{\alpha=1}^{N}\|\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha\|_\alpha^2$ requires knowledge of the covariance matrices $V[\mathbf{a}_\alpha]$. It is generally very difficult to estimate the noise characteristics exactly. However, it is relatively easy to predict qualitative characteristics such as homogeneity/inhomogeneity, isotropy/anisotropy, and their relative degrees from the characteristics of the imaging device and the image processing operation. So, we assume that the covariance matrices $V[\mathbf{a}_\alpha]$ are known only *up to scale* and write

$$V[\mathbf{a}_\alpha] = \epsilon^2 V_0[\mathbf{a}_\alpha], \tag{32}$$

where ϵ is an appropriately defined constant that measures the average magnitude of the noise. We call ϵ the *noise level* and $V_0[\mathbf{a}_\alpha]$ the *normalized covariance matrix*. We assume that $V_0[\mathbf{a}_\alpha]$ is known but ϵ is unknown.

Define the *normalized residual* by

$$J_0[\hat{\mathcal{S}}] = \sum_{\alpha=1}^{N}(\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha, V_0[\mathbf{a}_\alpha]^-(\mathbf{a}_\alpha - \hat{\mathbf{a}}_\alpha)). \tag{33}$$

Multiplying eq. (8) by ϵ^2, we define the *normalized geometric AIC* of model $\mathcal{S}$ by

$$AIC_0(\mathcal{S}) = J_0[\hat{\mathcal{S}}] + 2(dN + n')\epsilon^2. \tag{34}$$

In the following, we call $AIC_0(\mathcal{S})$ and $J_0[\hat{\mathcal{S}}]$ simply the *AIC* and the *residual*, respectively. Given a set of N data $\{\mathbf{a}_\alpha\}$ and two models $\mathcal{S}_1$ and $\mathcal{S}_2$, we regard model $\mathcal{S}_1$ better than $\mathcal{S}_2$ if $AIC_0(\mathcal{S}_1) < AIC_0(\mathcal{S}_2)$.

Remark 3. The usual AIC is obtained in the *asymptotic limit* $N \to \infty$ by applying the *law of large numbers* and the *central limit theorem* (Akaike 1974). However, eq. (8) is obtained in the limit of *small noise* for a *fixed* number N of the data. This leads to the following distinctive features of the geometric AIC as compared with the usual AIC:

- The degree of freedom n' of the model has no significant effect for the geometric AIC if the number N of data is large, whereas it plays a dominant role in the usual AIC.
- The dimension d of the model manifold plays a dominant role in the geometric AIC, while no such geometric concepts are involved in the usual AIC.
- The number N of data explicitly appears in the expression for the geometric AIC, but it does not in the usual AIC. This is because the true

positions of the data play the role of what is called *nuisance parameters* in statistics.

5.2. Model Comparison by AIC

In order to apply the AIC criterion, we need to estimate the noise level ϵ appropriately. This is obvious; distinguishing one model from another is meaningless if the noise level is high, while a small difference between the residuals gives a strong clue if the noise level is low. Note that noise is a characteristic of the devices and data processing operations involved and is *independent of the models we are comparing*, as we mentioned in Section 1.2. However, estimating the noise level a priori is in general very difficult.

If model $\mathcal{S}$ is correct, $J[\hat{\mathcal{S}}] = J_0[\hat{\mathcal{S}}]/\epsilon^2$ is subject to a χ^2 distribution with $rN - n'$ degrees of freedom. Hence, an unbiased estimator of ϵ^2 is obtained in the form

$$\hat{\epsilon}^2 = \frac{J_0[\hat{\mathcal{S}}]}{rN - n'}, \tag{35}$$

as long as $rN - n' \neq 0$. However, we need the true noise level ϵ to judge if the model $\mathcal{S}$ is correct. This difficulty can be avoided if we focus on comparing two models such that one *implies* the other.

Let $\mathcal{S}_1$ be a model of dimension d_1 and codimension r_1 with n_1' degrees of freedom, and $\mathcal{S}_2$ a model of dimension d_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

$$\mathcal{S}_2 \succ \mathcal{S}_1. \tag{36}$$

In this case, $J_0[\hat{\mathcal{S}}_2] \geq J_0[\hat{\mathcal{S}}_1]$ for whatever data $\{\mathbf{a}_\alpha\}$.

Suppose $\mathcal{S}_1$ is a general model which is assumed to be correct. Then, the squared noise level ϵ^2 is estimated by eq. (35) as log as $rN - n' \neq 0$. Substituting it to ϵ^2 in the expression for the geometric AIC, we see that the condition for $AIC_0(\mathcal{S}_2) < AIC_0(\mathcal{S}_1)$ is

$$\frac{J_0[\hat{\mathcal{S}}_2]}{J_0[\hat{\mathcal{S}}_1]} < 1 + \frac{2(d_1 - d_2)N + 2(n_1' - n_2')}{r_1 N - n_1'}. \tag{37}$$

If this condition is satisfied, the predicting capability is expected to increase by replacing the general model $\mathcal{S}_1$ by the strong model $\mathcal{S}_2$.

Example 7. Consider Example 3. We see that

$$P' \succ P, \quad L' \succ L, \quad \Pi' \succ \Pi, \quad P \succ L \succ \mathcal{R}^3. \tag{38}$$

- If the true positions are known to be coplanar, we can infer that the plane on which the true positions lie passes through the coordinate origin when

$$\frac{J_0[\hat{\Pi}']}{J_0[\hat{\Pi}]} < 1 + \frac{2}{N-3}, \tag{39}$$

  and infer that the true positions are collinear when

$$\frac{J_0[\hat{L}]}{J_0[\hat{\Pi}]} < 3 + \frac{4}{N-3}. \tag{40}$$

- If the true positions are known to be collinear, we can infer that the line on which the true positions lie passes through the coordinate origin when

$$\frac{J_0[\hat{L}']}{J_0[\hat{L}]} < 1 + \frac{2}{N-2}, \tag{41}$$

  and infer that the true positions are identical when

$$\frac{J_0[\hat{P}]}{J_0[\hat{P}]} < 2 + \frac{3}{N-2}. \tag{42}$$

- If the true positions are known to be identical, we can infer that the true position is at the coordinate origin when

$$\frac{J_0[\hat{P}']}{J_0[\hat{P}]} < 1 + \frac{2}{N-1}. \tag{43}$$

5.3. Model Selection vs. Testing of Hypotheses

The model selection criterion given by eq. (37) has a positive implication in contrast to the negative meaning of the statistical *testing of hypotheses*, according to which the procedure is given as follows. Since $J_0[\hat{\mathcal{S}}_2]/\epsilon^2$ is subject to a χ^2 distribution with $r_2N - n_2'$ degrees of freedom, the *hypothesis* that "model $\mathcal{S}_2$ is correct" is rejected if

$$\frac{J_0[\hat{\mathcal{S}}_2]}{\epsilon^2} > \chi^2_{r_2N-n_2',a} \tag{44}$$

with *significance level* $a\%$, where $\chi^2_{q,a}$ is the *upper* $a\%$ *point* of the χ^2 distribution with q degrees of freedom. If the square noise level ϵ^2 is approximated by the estimator $\hat{\epsilon}^2$ given by eq. (35), we can rewrite (44) as

$$\frac{J_0[\hat{\mathcal{S}}_2]}{J_0[\hat{\mathcal{S}}_1]} > \frac{\chi^2_{r_2N-n_2',a}}{r_1N - n_1'}. \tag{45}$$

The left-hand side equals the ratio of the logarithmic likelihoods of the two models, so this test belongs to a class called the *logarithmic likelihood ratio test*.

The interpretation of this test is that if eq. (45) holds, the hypothesis that the model is $\mathcal{S}_2$ is very questionable with *confidence level* $(100 - a)\%$ because if the hypothesis is true, we are observing a very rare event that occurs only with a probability $a\%$. Hence, we decide that *there exists no reason to favor model* $\mathcal{S}_2$ *over* $\mathcal{S}_1$. In other words, a statistical test can only *reject* a hypothesis when the data do not support it within a specified allowance threshold. Its ultimate purpose is to *negate* a hypothesis (hence called the *null hypothesis*) in favor of a default hypothesis (called the *alternative hypothesis*).

After all, *any* hypothesis is rejected in the presence of noise if the significance level is lowered (or the confidence level is raised); the judgment is not definitive in this sense, and it does not address the issue of choosing one model in favor of another. In contrast, the criterion given by eq. (37) gives a positive and definitive assertion that model $\mathcal{S}_2$ is *preferable* to $\mathcal{S}_1$ with regard to the predicting capability; it requires no knowledge about the noise magnitude and no arbitrarily set thresholds.

5.4. Other criteria

In traditional statistics, the MDL and the AIC are not the only criteria for the goodness of a model. Existing criteria can be classified roughly into *Bayesian*, *non-non-Bayesian*, and *empirical* types. Bayesian types include the MDL and the *BIC* (*Bayesian information criterion*) of Schwarz (1978) (see Clarke (1994), Clarke and Barron (1990) and Matsushima et al. (1991) for other criteria). Non-Bayesian types include the AIC and the *Cp* of Mallows (1973). A typical empirical method is *cross validation*: dividing the data into two parts, the *estimation* (or *learning*) *part* and the *validation part*, we fit a model to the estimation part and evaluate the residual for the validation part. Many variations are conceivable for this process. For example, we may fit a model to the data after removing one datum, validate the resulting fit by that datum, and repeat this for each datum in turn. This technique is called *jackknife*. Or we may generate "simulation data" by a computer many times in such a way that they as a whole have the same statistical properties as the original data and do fitting and validation by using them. Such a simulation-oriented method is called *bootstrap* (Efron and Tibshirani 1993).

The AIC can be viewed as "hypothetical cross validation"; we validate our fit by imagining the future data and compute the expectation. It appears at first sight that one can invent arbitrarily many criteria by juggling with equations. However, a lot of comparative studies of different criteria in the past indicate that all behave more or less similarly—particularly so asymptotically (e.g., see Hannan and Quinn (1979), Shibata (1981), Stone (1977)). All the criteria defined for traditional statistical inference could be modified to fit to our geometric inference problem described in Section 2.1, just as we modified the AIC into the geometric AIC, and it is very likely that we would end up with more or less similar results. With no good "criterion for choosing a criterion" at hand, we do not have any evidence that other criteria would be superior or inferior to the geometric AIC.

6. 3-D Motion Analysis

7. Structural Singularity

The *structure-from-motion* algorithm for reconstructing the 3-D object shape and the 3-D camera motion from two views has been studied by many researchers (Faugeras and Maybank 1990; Kanatani 1994b; Longuet-Higgins 1981; Tsai and Huang 1984; Weng, Ahuja and Huang 1993). However, the algorithm fails if all the feature points are coplanar in the scene. This is because a planar surface is a degenerate *critical surface* that gives rise to ambiguity of 3-D interpretation (Horn 1990; Maybank 1993; Negahdaripour 1990a,b). Hence, a different algorithm is necessary for a planar surface. The planar surface algorithm has also been studied by many researchers (Kanatani and Takeda 1995; Longuet-Higgins 1986, Weng et al. 1991). 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, the incoming rays of light are the same before and after the motion, so no 3-D information can be obtained. If follows that the structure-from-motion analysis must take the following steps:

1. *Rotation test*: We test if the translation is 0. If so, output a warning message and stop.
2. *Planarity test*: We test if the object is a planar surface. 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 arbitrarily set threshold. For example, based on the fact that the smallest eigenvalue 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 judgment?*

- 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 is different from image to image, so 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. 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. We now show that this problem can be systematically solved by the use of the geometric AIC criterion without knowing the magnitude of the image noise and without introducing any arbitrarily set thresholds.

7.1. General Model

Define an XYZ camera coordinate system such 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 $Z = 1$ such 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 $\mathbf{x} = (x, y, 1)^\top$.

Suppose the camera is rotated around the center of the lens by $\mathbf{R}$, and finally translated by $\mathbf{h}$ (Fig. 10). We call $\{\mathbf{h}, \mathbf{R}\}$ 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 after the motion, a point (x', y'), which can be represented by vector $\mathbf{x}' = (x', y', 1)$ with respect to the $X'Y'Z'$ coordinate system, can be represented by vector $\mathbf{R}\mathbf{x}'$ with respect to the XYZ coordinate system. If follows that vectors $\mathbf{x}$ and $\mathbf{x}'$ can be images of the same point in the scene if and only if the following *epipolar equation* holds (Faugeras 1993; Kanatani 1993a; Maybank 1993;

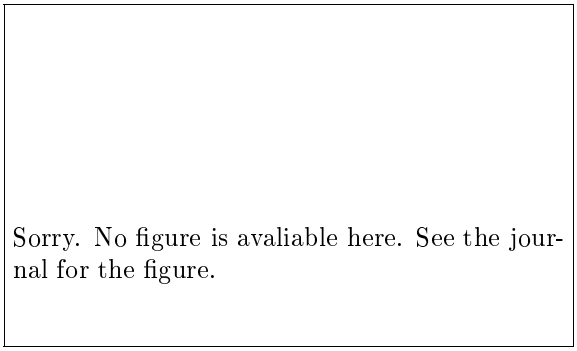

Fig. 10. Epipolar geometry.

Weng, Huang and Ahuja 1993), where $|\cdot, \cdot, \cdot|$ denotes the scalar triple product of vectors:

$$|\mathbf{x}, \mathbf{h}, \mathbf{R}\mathbf{x}'| = 0. \tag{46}$$

Suppose we have N pairs of corresponding points $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$, $\alpha = 1, ..., N$. We decompose the vectors $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$ into the form

$$\mathbf{x}_\alpha = \bar{\mathbf{x}}_\alpha + \Delta\mathbf{x}_\alpha, \quad \mathbf{x}'_\alpha = \bar{\mathbf{x}}'_\alpha + \Delta\mathbf{x}'_\alpha, \tag{47}$$

where $\bar{\mathbf{x}}_\alpha$ and $\bar{\mathbf{x}}'_\alpha$ represent the positions supposedly observed if noise did not exist. We regard the noise terms $\Delta\mathbf{x}_\alpha$ and $\Delta\mathbf{x}'_\alpha$ as independent Gaussian random variables of mean $\mathbf{0}$ and respective covariance matrices $V[\mathbf{x}_\alpha]$ and $V[\mathbf{x}'_\alpha]$. 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.

The $2N$ vectors $\mathbf{x}_\alpha$ and $\mathbf{x}'_\alpha$, $\alpha = 1, ..., N$, can be viewed as *six*-dimensional data $\mathbf{a}_\alpha = \mathbf{x}_\alpha \oplus \mathbf{x}'_\alpha \in \mathcal{R}^6$ sampled from the *four*-dimensional data space

$$\mathcal{A} = \{(x, y, 1, x', y', 1)^\top | x, y, x', y' \in \mathcal{R}\} \subset \mathcal{R}^6. \tag{48}$$

Since the epipolar equation (46) defines a *three*-dimensional manifold $\mathcal{S}$ in the four-dimensional data space $\mathcal{A}$.

From the epipolar equation (46), it is immediately seen that the translation $\mathbf{h}$ can be determined only up to scale. Hence, the unknown motion parameters $\{\mathbf{h}, \mathbf{R}\}$ have five degrees of freedom. Consequently, the manifold $\mathcal{S}$ is a model of dimension 3, codimension 1, and *five* degrees of freedom. We decompose the covariance matrices $V[\mathbf{x}_\alpha]$ and $V[\mathbf{x}'_\alpha]$ into the noise level ϵ and the normalized covariance matrices $V_0[\mathbf{x}_\alpha]$ and $V_0[\mathbf{x}'_\alpha]$ in the form

$$\begin{aligned} V[\mathbf{x}_\alpha] &= \epsilon^2 V_0[\mathbf{x}_\alpha], \\ V[\mathbf{x}'_\alpha] &= \epsilon^2 V_0[\mathbf{x}'_\alpha]. \end{aligned} \tag{49}$$

The AIC of this model is

$$AIC_0(\mathcal{S}) = J_0[\hat{\mathcal{S}}] + 2(3N + 5)\epsilon^2. \tag{50}$$

Let $\{\hat{\mathbf{h}}, \hat{\mathbf{R}}\}$ be the maximum likelihood estimators of the motion parameters $\{\mathbf{h}, \mathbf{R}\}$. They can be computed accurately and effectively by a numerical scheme called *renormalization*, and the residual $J_0[\hat{\mathcal{S}}]$ is computed as follows (Kanatani 1994b):

$$\begin{aligned} J_0[\hat{\mathcal{S}}] = \sum_{\alpha=1}^{N} & |\mathbf{x}_\alpha, \hat{\mathbf{h}}, \hat{\mathbf{R}}\mathbf{x}'_\alpha|^2 \\ & \Big/ \Big((\hat{\mathbf{h}} \times \hat{\mathbf{R}}\mathbf{x}'_\alpha, V_0[\mathbf{x}_\alpha](\hat{\mathbf{h}} \times \hat{\mathbf{R}}\mathbf{x}'_\alpha)) \\ & + (\hat{\mathbf{h}} \times \mathbf{x}, \hat{\mathbf{R}} V_0[\mathbf{x}'_\alpha] \hat{\mathbf{R}}^\top (\hat{\mathbf{h}} \times \mathbf{x})) \Big). \end{aligned} \tag{51}$$

If this model is correct, $J_0[\hat{\mathcal{S}}]/\epsilon^2$ is subject to a χ^2 distribution with $N-5$ degrees of freedom. Hence, the squared noise level ϵ^2 can be estimated by

$$\hat{\epsilon}^2 = \frac{J_0[\hat{\mathcal{S}}]}{N - 5}. \tag{52}$$

Remark 4. As is well known, the solution $\{\hat{\mathbf{h}}, \hat{\mathbf{R}}\}$ of the motion parameters was not unique in general (Faugeras 1993; Kanatani 1993a; Maybank 1993; Weng, Huang and Ahuja 1993). The existence of multiple solutions is due to the fact that different motion parameters can define the same manifold $\mathcal{S}$. However, *the maximum likelihood estimator $\hat{\mathcal{S}}$ of the manifold $\mathcal{S}$ is generally unique.*

7.2. Planar Surface Model

Suppose the object is a planar surface. Let $\mathbf{n}$ be its unit surface normal with respect to the XYZ coordinate system, and d its distance (positive in the direction $\mathbf{n}$) from the origin O (Fig. 11); we

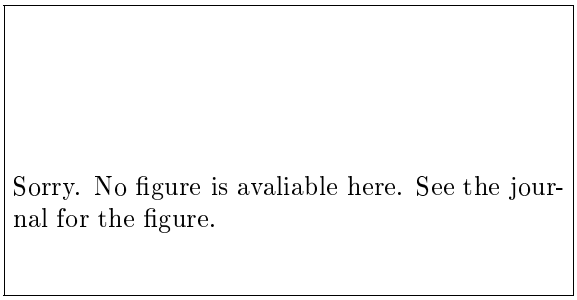

Fig. 11. Geometry of camera motion relative to a planar surface.

call $\{\mathbf{n}, d\}$ the *surface parameters*. It can be easily shown that $\mathbf{x}$ and $\mathbf{x}'$ are images of the same point on that plane if and only if the following equation holds (Kanatani 1993a; Maybank 1993; Weng, Huang and Ahuja 1993):

$$\mathbf{x}' \times \mathbf{A}\mathbf{x} = \mathbf{0}. \tag{53}$$

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

$$\mathbf{A} = \mathbf{R}^\top(\mathbf{h}\mathbf{n}^\top - d\mathbf{I}), \tag{54}$$

which determines an image transformation called *homography*. Only two component equations of eq. (53) are independent. Hence, it defines a *two*-dimensional manifold $\mathcal{S}_\Pi$ in the four-dimensional data space $\mathcal{A}$. The unknown surface parameters $\{\mathbf{n}, d\}$ have three degrees of freedom, and the unknown motion parameters $\{\mathbf{h}, \mathbf{R}\}$ have five degrees of freedom. Hence, the manifold $\mathcal{S}_\Pi$ is a model of dimension 2, codimension 2, and *eight* degrees of freedom. Let $\hat{\mathcal{S}}_\Pi$ be the maximum likelihood estimator of $\mathcal{S}_\Pi$. The AIC of this model is

$$AIC_0(\mathcal{S}_\Pi) = J_0[\hat{\mathcal{S}}_\Pi] + 2(2N+8)\epsilon^2. \tag{55}$$

Let $\{\hat{\mathbf{n}}, \hat{d}\}$ and $\{\hat{\mathbf{h}}, \hat{\mathbf{R}}\}$ be the maximum likelihood estimators of the surface and motion parameters $\{\mathbf{n}, d\}$ and $\{\mathbf{h}, \mathbf{R}\}$, respectively. They can be computed accurately and effectively by a numerical scheme called *renormalization*, and the residual $J_0[\hat{\mathcal{S}}_\Pi]$ is computed as follows (Kanatani and Takeda 1995):

$$J_0[\hat{\mathcal{S}}_\Pi] = \sum_{\alpha=1}^{N} (\mathbf{x}'_\alpha \times \hat{\mathbf{A}}\mathbf{x}_\alpha, \hat{\mathbf{W}}_0(\mathbf{x}'_\alpha \times \hat{\mathbf{A}}\mathbf{x}_\alpha)), \tag{56}$$

$$\hat{\mathbf{W}}_0 = \Big(\mathbf{x}'_\alpha \times \hat{\mathbf{A}} V_0[\mathbf{x}_\alpha]\hat{\mathbf{A}}^\top \times \mathbf{x}'_\alpha + (\hat{\mathbf{A}}\mathbf{x}_\alpha) \times V_0[\mathbf{x}'_\alpha] \times (\hat{\mathbf{A}}\mathbf{x}_\alpha)\Big)_2^-. \tag{57}$$

Here, $\hat{\mathbf{A}}$ is the estimate of matrix $\mathbf{A}$ obtained by replacing $\{\mathbf{n}, d\}$ and $\{\mathbf{h}, \mathbf{R}\}$ by their maximum likelihood estimators $\{\hat{\mathbf{n}}, \hat{d}\}$ and $\{\hat{\mathbf{h}}, \hat{\mathbf{R}}\}$, respectively, in eq. (54). The product $\mathbf{a} \times \mathbf{U} \times \mathbf{a}$ of a vector $\mathbf{a} = (a_i)$ and a matrix $\mathbf{U} = (U_{ij})$ is defined to be a matrix whose (ij) element is $\sum_{k,l=1}^{3} \varepsilon_{ikl}\varepsilon_{jmn}a_k a_m U_{ln}$, where 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. The symbol $(\cdot)_r^-$ denotes the *rank-constrained generalized inverse*.

Remark 5. Let $\mathbf{A} = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$ be the spectral decomposition of a p-dimensional positive semidefinite symmetric matrix $\mathbf{A}$, where $\{\mathbf{v}_i\}$ is an orthonormal system of the eigenvectors of $\mathbf{A}$ for eigenvalues $\{\lambda_i\}$ (see Remark 1). If $\lambda_1 \geq \cdots \geq \lambda_m > \lambda_{m+1} = \cdots = \lambda_p = 0$, the *rank-constrained* (*Moore-Penrose*) *generalized inverse* of $\mathbf{A}$ for rank r $(\leq m)$ is given by $(\mathbf{A})_r^- = \sum_{i=1}^{r} \mathbf{v}_i \mathbf{v}_i^\top / \lambda_i$.

Remark 6. As is well known, the solution $\{\hat{\mathbf{n}}, \hat{d}\}$ and $\{\hat{\mathbf{h}}, \hat{\mathbf{R}}\}$ of the surface and motion parameters was not unique in general (Faugeras 1993; Kanatani 1993a; Maybank 1993; Weng, Huang and Ahuja 1993). This ambiguity is due to the fact that different surface and motion parameters can define the same manifold. However, *the maximum likelihood estimator $\hat{\mathcal{S}}_\Pi$ of the manifold $\mathcal{S}_\Pi$ is generally unique.*

7.3. Rotation Model

If the camera motion is a pure rotation around the center of the lens, the incoming rays of light are the same before and after the camera motion, and hence no 3-D information can be obtained. The necessary and sufficient condition for $\mathbf{h} = \mathbf{0}$ is that all corresponding image points $\mathbf{x}$ and $\mathbf{x}'$ satisfy the following equation (Fig. 12):

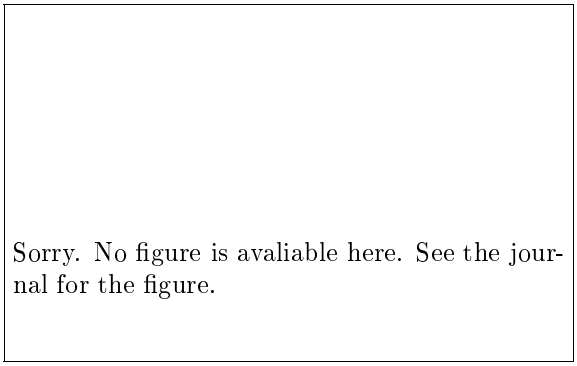

Fig. 12. Pure rotation of the camera.

$$\mathbf{x} \times \mathbf{R}\mathbf{x}' = \mathbf{0}. \tag{58}$$

This equation has only two independent component equations. Hence, it defines a two-dimensional manifold $\mathcal{S}_R$ in the four-dimensional data space $\mathcal{A}$. Since the unknown rotation $\mathbf{R}$ has three degrees of freedom, the manifold $\mathcal{S}_R$ is a model of dimension 2, codimension 2, and *three* degrees of freedom. Let $\hat{\mathcal{S}}_R$ be the maximum likelihood estimator of $\mathcal{S}_R$. The AIC of this model is

$$AIC_0(\mathcal{S}_R) = J_0[\hat{\mathcal{S}}_R] + 2(2N+3)\epsilon^2. \tag{59}$$

Let $\hat{\mathbf{R}}$ be the maximum likelihood estimator of the rotation $\mathbf{R}$. The residual $J_0[\hat{\mathcal{S}}_R]$ is computed as follows (Kanatani 1996a):

$$J_0[\hat{\mathcal{S}}_R] = \sum_{\alpha=1}^{N} (\mathbf{x}_\alpha \times \hat{\mathbf{R}}\mathbf{x}'_\alpha, \hat{\mathbf{W}}_0(\mathbf{x}_\alpha \times \hat{\mathbf{R}}\mathbf{x}'_\alpha)), \tag{60}$$

$$\hat{\mathbf{W}}_0 = \Big(\mathbf{x}_\alpha \times \hat{\mathbf{R}}V_0[\mathbf{x}'_\alpha]\hat{\mathbf{R}}^\top \times \mathbf{x}_\alpha + (\hat{\mathbf{R}}\mathbf{x}'_\alpha) \times V_0[\mathbf{x}_\alpha] \times (\hat{\mathbf{R}}\mathbf{x}'_\alpha)\Big)_2^-. \tag{61}$$

Remark 7. The maximum likelihood estimator $\hat{\mathbf{R}}$ of the camera rotation $\mathbf{R}$ can be computed by a numerical search algorithm (e.g., steepest descent), minimizing eq. (11) viewed as a function of $\hat{\mathbf{R}}$. A good approximation of $\hat{\mathbf{R}}$ is analytically obtained by applying the singular value decomposition (Kanatani 1994a). Note that what we actually need is not the estimate $\hat{\mathbf{R}}$ itself but the residual $J_0[\hat{\mathcal{S}}_R]$. Since $\hat{\mathbf{R}}$ is the value that minimizes $J_0[\hat{\mathcal{S}}_R]$, the minimum value can be accurately computed even from an approximate value of $\hat{\mathbf{R}}$.

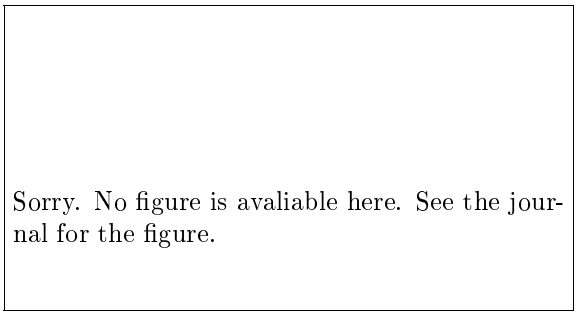

Fig. 13. Two planar grids are hinged together in the scene.

7.4. Planarity Test and Rotation Test

It is easy to see that eqs. (53) and (58) both imply the epipolar equation (46). It can also be seen that eq. (53) reduces to eq. (58) if the matrix $\mathbf{A}$ in eq. (53) is constrained to be a rotation matrix (i.e., an orthogonal matrix of determinant 1). Hence, the following relation holds:

$$\mathcal{S}_R \succ \mathcal{S}_\Pi \succ \mathcal{S}. \tag{62}$$

It follows that we always have $J_0[\hat{\mathcal{S}}_R] \geq J_0[\hat{\mathcal{S}}_\Pi] \geq J_0[\hat{\mathcal{S}}]$. Applying the comparison criterion (37), we obtain the following procedures:

1. *Planarity test*: The object surface can be inferred to be planar if
$$\frac{J_0[\hat{\mathcal{S}}_\Pi]}{J_0[\hat{\mathcal{S}}]} < 3 + \frac{4}{N-5}. \tag{63}$$
2. *Rotation test*: The camera motion can be inferred to be a pure rotation if
$$\frac{J_0[\hat{\mathcal{S}}_R]}{J_0[\hat{\mathcal{S}}]} < 3 + \frac{14}{N-5}. \tag{64}$$

7.5. Experiments

We defined two planar grids hinged together with angle $\pi - \theta$ in the scene and generated images that

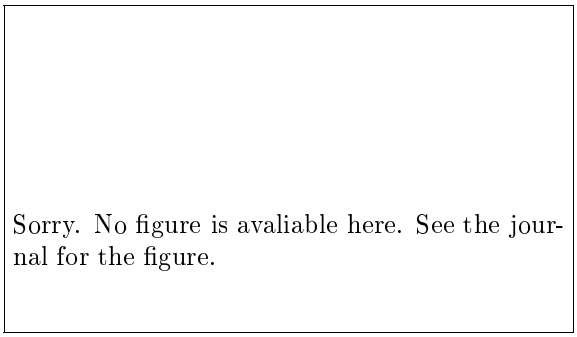

Fig. 14. Percentage of the instances judged to be planar.

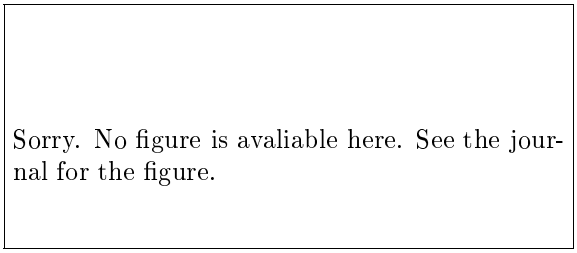

Fig. 15. An instance for which the object is judged to be planar ($\theta = 22^\circ$ and $\sigma = 1$).

simulate two views from different camera positions (Fig. 13). The image size and the focal length were assumed to be 512×512 (pixels) and 600 (pixels), respectively. 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 different noise. For various values of θ and for $\sigma = 0.5 \sim 3.0$, we computed the percentage of the instances for which the object is judged to be planar (Fig. 14). We see that the threshold for the test is automatically adjusted to the noise. Note that the purpose of the test is not to know the *true* shape; the purpose is to test if the object shape *can* be regarded as planar.

If $\sigma = 1.0$, the percentage is approximately 50% for $\theta = 22^\circ$. Fig. 15 shows one instance for which the object is judged to be planar; Fig. 16 shows one instance for which the object is judged to be non-planar. Fig. 17 shows the 3-D shapes reconstructed from the general and planar surface models. The true shape is superimposed in dashed lines. 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 to be planar. In contrast, the non-planar shape can be reconstructed fairly well if the object is judged to be non-planar.

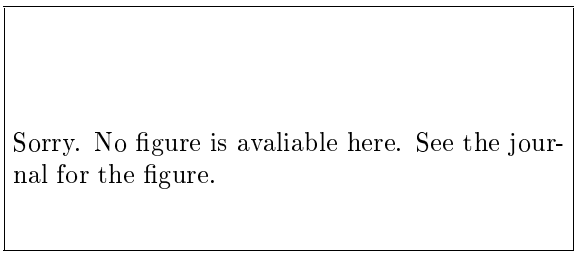

Fig. 16. An instance for which the object is judged to be non-planar ($\theta = 22^\circ$ and $\sigma = 1$).

8. Further Applications

The geometric AIC can be applied to many other problems in computer vision and robotics problems where geometric inference is involved. Here, we merely list such possibilities without going into the details.

8.1. 3-D Reconstruction by Stereo Vision

Stereo vision can be regarded as structure from motion with known motion parameters $\{\mathbf{h}, \mathbf{R}\}$. Then, the epipolar equation (46) defines a model $\mathcal{S}$ of dimension 3, codimension 1, and zero degrees of freedom; the maximum likelihood estimator $\hat{\mathcal{S}}$ of $\mathcal{S}$ is $\mathcal{S}$ itself. Hence, its AIC is

$$AIC_0(\mathcal{S}) = J_0[\hat{\mathcal{S}}] + 6N\epsilon^2. \quad (65)$$

If the object is a planar surface, eq. (53) defines a model $\mathcal{S}_\Pi$ of dimension 2, codimension 2, and three degrees of freedom for known motion parameters $\{\mathbf{h}, \mathbf{R}\}$. Hence, its AIC is

$$AIC_0(\mathcal{S}_\Pi) = J_0[\hat{\mathcal{S}}_\Pi] + 2(2N + 3)\epsilon^2. \quad (66)$$

If the base line between the two cameras is very short, it is difficult to distinguish objects at a finite distance from objects at infinity in the presence of image noise. The condition that corresponding pair of image points $\{\mathbf{x}, \mathbf{x}'\}$ is a projection of a point at infinity is given by eq. (58), which defines a model $\mathcal{S}_\infty$ of dimension 2, codimension 2, and zero degrees of freedom. Its AIC is

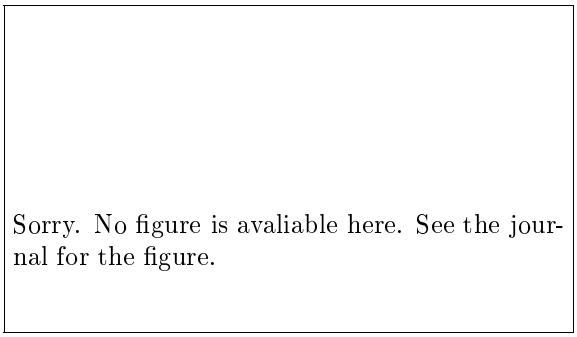

Fig. 17. (a) 3-D reconstruction from the images in Fig. 15. (b) 3-D reconstruction from the images in Fig. 16.

$$AIC_0(\mathcal{S}_\infty) = J_0[\hat{\mathcal{S}}_R] + 4N\epsilon^2. \tag{67}$$

Since $\mathcal{S}_\infty \succ \mathcal{S}_\Pi \succ \mathcal{S}$, the object surface can be inferred to be planar if

$$\frac{J_0[\hat{\mathcal{S}}_\Pi]}{J_0[\hat{\mathcal{S}}]} < 3 - \frac{6}{N}, \tag{68}$$

and the object can be inferred to be at infinity if

$$\frac{J_0[\hat{\mathcal{S}}_\infty]}{J_0[\hat{\mathcal{S}}]} < 3. \tag{69}$$

Thus, we can build an *intelligent* stereo sytem (Kanazawa and Kanatani 1996): if the object is judged to be a planar surface, the system outputs an optimally fitted planar surface instead of a point-wise reconstructed shape (see Kanazawa and Kanatani (1995) for this optimal fitting scheme). Such ability would be very useful in indoor environments where many objects have planar surfaces. On the other hand, if the object is judged to be infinitely far away, the system outputs a warning message, telling us that the distparity is too small to do reliable 3-D reconstruction. Such ability would be very useful for "mircostereo system" embedded in a robot manipulator with a very short baseline.

8.2. 3-D Interpretation of Optical Flow

Optical flow is a flow field in an image caused by an instantaneous motion of an object or a camera in the scene. Hence, its 3-D analysis is done in almost the same way as in the case of finite motion; all we need to do is take the limit of infinitesimal motion. The structure-from-motion algorithm for optical flow has been studied by many researchers (Adiv 1985; Heeger and Jepson 1992; Kanatani 1993b; Ohta and Kanatani 1995; Yasumoto and Medioni 1986). As in the case of finite motion, the algorithm fails if all the object is a planar surface, because a planar surface is also a degenerate *critical surface* for optical flow (Horn 1987; Maybank 1985; Negahdaripour 1989). The planar surface algorithm has also been studied by many researchers (Kanatani 1987; Longuet-Higgins 1984; Subbarao and Waxman 1986). 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. Applying the same procedure as in the case of finite motion except taking the limit of infinitesimal motion, we can formulate the rotation test and the planarity test for optical flow (we omit the details; see Kanatani 1996a).

Although the finite motion analysis and optical flow analysis are completely parallel, one notable difference exists: optical flow analysis can also be viewed as traditional statistical estimation. In fact, the detected optical flow $\mathbf{u}$ can be expressed in the form

$$\mathbf{u} = F(\mathbf{x}, Z(\mathbf{x}), \mathbf{v}, \omega) + \mathbf{e}. \tag{70}$$

Here, $Z(\mathbf{x})$ is the depth of the object surface at image point $\mathbf{x}$; $\{\mathbf{v}, \omega\}$ are instantaneous translation velocity and angular velocity of the camera motion, respectively; $\mathbf{e}$ denotes the error in optical flow detection (see Barron et al. (1996) and Mitiche and Bouthemy (1996) for optical flow detection techniques). Estimation of the depth $Z(\mathbf{x})$ and the motion parameters $\{\mathbf{v}, \omega\}$ from observed flow $\mathbf{u}$ belongs to a standard problem of traditional estimation called *nonlinear regression* and can also be treated in that framework (Endoh et al. 1994; Tagawa et al. 1993, 1994; Young and Chellappa 1992).

8.3. Curve Representation

In order to recognize objects in an image, the first task is to detect boundaries that describe individual objects. In such a processing, edges obtained by an edge filter are often represented as

a collection of simple primitives such as straight line segments, circular segments, conic segments, and high order polynomial curves (Rosin and West 1995).

However, such primitives have inclusion relations among themselves, so we cannot simply base the judgement on the degree of discrepancy between the model and the data (Kanatani 1997b); a high order polynomial curve is always selected whatever distance measure is used. Although the usual AIC can be used if the problem is in the form of non-linear regression (Boyer et al. 1994), the geometric AIC introduced here can be applied to any type of parametric fitting of curves.

Using the geometric AIC, we can not only select the best primitive for each segment but also automatically stop the segmentation process *without an artificial threshod*; we recursively subdivide a curve until the geometric AIC no longer decreases by a further subdivision.

8.4. Symmetry Detection

Finding if a figure found in an image has a certain symmetry in the presence of noise is an interesting problem, and Zabrodsky et al. (1995) introduced as a measure of symmetry the sum of squared distances over which individual vertices must be moved to enforce the assumed symmetry.

This idea can be also used in computer aided graphics (CAD) tools. For example, a user wants draws a symmetric figure of some kind, such as a parallelogram, a rectangle, and a square, by manipulating a mouse, but manually input figures do not generally have the required symmetry. It would be nice if the computer can detect the intended symmetry in the input figure and automatically modify it to enforce the required symmetry.

However, symmetry classes have inclusion relations among themselves; e.g., the set of squares is a subset of the set of rectangles. This means that we cannot simply base the judgement on the degree of discrepancy from the required symmetry; the weakest symmetry always has the smallest discrepancy whatever distance measure is used (Kanatani 1997a). The geometric AIC introduced here can be applied to distinguish classes with inclusion relations without introducing any empirically adjustable thresholds (Triono and Kanatani 1996).

8.5. Continuous Measure for Active Vision

So far, we have focused only on comparing models based on the geometric AIC. However, we can also use the geometric AIC to define a *continuous measure* that indicates how preferable a particular model is (Kanatani 1996b). Recall that the geometric AIC is an estimate of the expected residual $E^*[E[\|\mathbf{a}_\alpha^* - \hat{\mathbf{a}}_\alpha\|_\alpha^2]]$. Hence, given two models $\mathcal{S}_1$ and $\mathcal{S}_2$ such that $\mathcal{S}_2 \succ \mathcal{S}_1$, we can view the ratio

$$K = \sqrt{\frac{AIC_0(\mathcal{S}_2)}{AIC_0(\mathcal{S}_1)}} \tag{71}$$

as a continuous measure of how preferable the model $\mathcal{S}_2$ is to the model $\mathcal{S}_1$. The model $\mathcal{S}_2$ is preferred if $K < 1$, but the magnitude of K itself can be viewed as indicating the confidence of this judgement.

In structure from motion and stereo, This measure is useful not only for detcting degeneracy but also measuing the *information content* of the images. As (i) image noise increases, (ii) the disparity between the two images decreases, and (iii) the depth variance of the feature points decreases, this measure decreases and we observe "virtual genencracy". Hence, it serves as a criterion for *active vision* (Aloimonos 1993); a robot actively controls the camera in such a way that the information content of the images measured by them increases.

9. Concluding Remarks

9.1. Summary

In this paper, we have presented a statistical framework for detecting degeneracies of a geometric model by evaluating its predictive capability in terms of the *expected residual* and derived the *geometric AIC*. We have applied it to structure-from-motion analysis and shown that we can judge whether or not the object can be regarded as a planar surface and whether or not the camera motion can be regarded as pure rotation. We have illustrated our approach by simulation examples. We have also suggested potential applications of our

theory to 3-D reconstruction by stereo vision, 3-D interpretation of optical flow, curve representation, symmetry detection, and active vision criteria. Our theory is expected to play a crucial role in model fitting for building a 3-D model of the environment from image and sensor data.

9.2. Remaining Issues

Throughout this paper, we have assumed Gaussian noise because Gaussian noise is most fundamental and practically important. Theoretically, however, our theory can be extended to non-Gaussian noise. For a general noise model, we need to consider the logarithmic likelihood instead of the residual, and the generalized inverse of the Fisher information matrix plays the role the covariance matrix of Gaussian noise (Kanatani 1996a). However, such an extension does not seem to have much practical significance because of the difficulty of estimating the parameters of a non-Gaussian noise distribution

A more important issue in real applications is detecting *outliers* that have very different characteristics from other data. Research in this direction is also in progress (Meer et al. 1991; Stewart 1995; Torr and Murray 1993), and we do not go into the details in this paper, but the problem of model selection still remains after removing outliers.

References

Adiv, G. 1985, Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Trans. Patt. Anal. Mach. Intell.* 7(4): 384–401.

Akaike, H. 1969, Fitting autoregressive model for prediction. *Ann. Inst. Statist. Math.* 21: 234–247.

Akaike, H. 1974, A new look at the statistical model identification. *IEEE Trans. Automation Control* 19(6): 176–723.

Aloimonos, Y. (Ed.) 1993, *Active Perception*, Lawrence Erlbaum Associates; Hillsdale, NJ.

Barron, A. R. and Cover, M. 1991, Minimum complexity density estimation, *IEEE Trans. Inform Theory* 37(4): 1034–1054.

Barron, J. L.,Fleet, D. J. and Beauchemin S. S. 1994, Performance of optical flow techniques. *Int. J. Comput. Vision* 12(1): 43–77.

Boyer, K. L., Mirza, M. J. and Ganguly, G. 1994, The robust sequential estimator: A general approach and its application to surface organization in range data, *IEEE Trans. Patt. Anal. Mach. Intell.* 16(10): 987–1001.

Clarke, B. S. 1994, Jeffery's prior is asymptotically least favorable under entropy risk. *J. Statist. Planning Inference* 41: 37–60.

Clarke, B. S. and Barron, A. R. 1990, Information-theoretic asymptotics of Bayes methods. *IEEE Trans. Inf. Theory* 36(3): 453–471.

Efron, B. and Tibshirani, R. J. 1993, *An Introduction to Bootstrap*, Chapman-Hall: New York.

Endoh, T., Toriu, T. and Tagawa, N. 1994, A superior estimator to the maximum likelihood estimator on 3-D motion estimation from noisy optical flow. *IEICE Trans. Inf. & Syst.* E77-D(1): 1240–1246.

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

Faugeras O. D. and Maybank, S. 1990: Motion from point matches: Multiplicity of solutions. *Int. J. Comput. Vision* 4(3): 225–246.

Gu, H., Shirai, Y. and Asada, M. 1996, MDL-based segmentation and motion modeling in a long sequence of scene with multiple independently moving objects. *IEEE Trans. Patt. Anal. Mach. Intell.* 18(1): 58–64.

Hannan, E. J. and Quinn, B. G. 1979, The determination of the order of an autoregression. *J. Roy. Statist. Soc.* B-41: 190–195.

Heeger, D. J. and Jepson, A. D. 1992, Subspace methods for recovering rigid motion I: Algorithm and implementation. *Int. J. Comput. Vision* 7(2): 95–117.

Horn, B. K. P. 1987, Motion fields are hardly ever ambiguous. *Int. J. Comput. Vision* 1(3): 259–274.

Horn, B. K. P. 1990, Relative orientation. *Int. J. Comput. Vision* 4(1): 59–78.

Kanatani, K. 1987, Structure and motion from optical flow under perspective projection. *Comput. Vision Graphics Image Process.* 38: 122–146.

Kanatani, K. 1993a, *Geometric Computation for Machine Vision*. Oxford University Press: Oxford.

Kanatani, K. 1993b, 3-D interpretation of optical flow by renormalization. *Int. J. Comput. Vision* 11(3): 267–282.

Kanatani, K. 1994a, Analysis of 3-D rotation fitting. *IEEE Trans. Patt. Anal. Mach. Intell.* 16(5): 543–549.

Kanatani, K. 1994b, Renormalization for motion analysis: Statistically optimal algorithm. *IEICE Trans. Inf. Sys.* E77-D(1): 1233–1239.

Kanatani, K. 1996a, *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science: Amsterdam.

Kanatani, K. 1996b, Automatic singularity test for motion analysis by an information criterion, *Proc. 4th European Conf. Comput. Vision*, April 1996, Cambridge, Vol. 1, pp. 697–708.

Kanatani, K. 1997a, Comments on "Symmetry as a continuous feature". *IEEE Trans. Patt. Anal. Mach. Intell.* to appear.

Kanatani, K. 1997b, Comments on "Nonparametric segmentation of curves into various representation". *IEEE Trans. Patt. Anal. Mach. Intell.* to appear.

Kanatani K. and Takeda, S. 1995, 3-D motion analysis of a planar surface by renormalization. *IEICE Trans. Inf. & Syst.* E78-D(8): 1074–1079.

Kanazawa, Y. and Kanatani, K. 1995, Direct reconstruction of planar surfaces by stereo vision. *IEICE Trans. Inf. & Syst.* E78-D(7): 917–922.

Kanazawa, Y. and Kanatani, K. 1996, Statistical inference by stereo vision. *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, November 1996, Osaka, Japan, pp. 1272–1279.

Kullback, S. 1959, *Information Theory and Statistics*, Wiley: New York.

Leclerc, Y. G. 1989. Constructing simple stable descriptions for image partitioning. *Int. J. Comput. Vision* 3(1): 73–102.

Longuet-Higgins, H. C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature* 293(10): 133–135.

Longuet-Higgins, H. C. 1984, The visual ambiguity of a moving plane. *Proc. Royal Soc. Lond.* B-223: 165–175.

Longuet-Higgins, H. C. 1986. The reconstruction of a plane surface from two perspective projections. *Proc. Roy. Soc. Lond.* B-227: 399–410.

Mallows, C. L. 1973, Some comments on C_p, *Technometrics* 15(4): 661–675.

Matsushima, T., Inazumi, H. and Hirasawa, S. 1991, A class of distortionless codes designed by Bayes decision theory. *IEEE Trans. Inf. Theory* 37(5): 1288–1293.

Maybank, S. 1985, The angular velocity associated with the optical flowfield arising from motion through rigid environment. *Proc. Roy. Soc. Lond.* A-401: 317–326.

Maybank, S. 1993, *Theory of Reconstruction from Image Motion*. Springer: Berlin.

Meer, P., Mintz, D., Rosenfeld, A. and Kim, D. Y. 1991, Robust regression methods for computer vision: A review. *Int. J. Comput. Vision* 6(1): 59–70.

Mitiche, A. and Bouthemy, P. 1996, Computation and analysis of image motion: A synopsis of current problems and methods. *Int. J. Comput. Vision* 19(1): 29–55.

Negahdaripour, S. 1989, Critical surface pairs and triplets. *Int. J. Comput. Vision* 3(4): 293–312.

Negahdaripour, S. 1990a, Multiple interpretations of the shape and motion of objects from two perspective images. *IEEE Trans. Patt. Anal. Mach. Intell.* 12(11): 1025–1030.

Negahdaripour, S. 1990b, Closed-form relationship between the two interpretations of a moving plane. *J. Opt. Soc. Amer.* 7(2): 279–285.

Ohta, N. and Kanatani, K. 1995, Optimal structure from motion algorithm for optical flow. *IEICE Trans. Inf. & Syst.* E78-D(12): 1559–1566.

Rissanen, J. 1984, Universal coding, information, prediction and estimation. *IEEE Trans. Inf. Theory* 30(4): 629–636.

Rissanen, J. 1987, *Stochastic Complexity in Statistical Inquiry*. World Scientific: Singapore.

Rosin, P. L., and West, G. A. W. 1995, Nonparametric segmentation of curves into various representation. *IEEE Trans. Patt. Anal. Mach. Intell.* 17(12): 1140–1153.

Schwarz, G. 1978, Estimating the dimension of a model, *Ann. Statist.* 6(2): 461–464.

Shibata, S. 1981, An optimal selection of regression variables. *Biometrika* 68: 45–54.

Stewart, C. V. 1995, MINPRAN: A new robust estimator for computer vision. *IEEE Trans. Patt. Anal. Mach. Intell.* 17(10): 925–938.

Stone, M. 1977, An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *J. Roy. Statist. Soc.* B39: 44–47.

Subbarao, M. and Waxman, A. M. 1986, Closed form solution to image flow equations for a planar surface in motion. *Comput. Vision Graphics Image Process.* 36: 208–228.

Tagawa, N., Toriu, T. and Endoh, T. 1993, Un-biased linear algorithm for recovering three-dimensional motion from optical flow. *IEICE Trans. Inf.& Syst.* E76-D(10): 1263–1275.

Tagawa, N., Toriu, T. and Endoh, T. 1994, Estimation of 3-D motion from optical flow with unbiased objective function. *IEICE Trans. Inf. & Syst.* E77-D(10) 1148–1161.

Torr, P. H. and Murray, D. W. 1993, Statistical detection of independent movement from a moving camera. *Image Vision Comput.* 11(4): 180–187.

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

Tsai, R. Y. and Huang, H. S. 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. Patt. Anal. Mach. Intell.* 6(1): 13–27.

Weng, J., Ahuja, N. and Huang, T. S. 1991. Motion and structure from point correspondences with error estimation: Planar surfaces. *IEEE Trans. Sig. Proc.* 39(12): 2691–2717.

Weng, J., Ahuja, N., and Huang, T. S. 1993. Optimal motion and structure estimation. *IEEE Trans. Patt. Anal. Mach. Intell.* 15(9): 864–884.

Weng, J., Huang, T. S. and Ahuja, N. 1993, *Motion and Structure from Image Sequences*. Springer: Berlin.

Yasumoto, Y. and Medioni, G. 1986, Robust estimation of three-dimensional motion parameters from a sequence of image frames using regularization. *IEEE Trans. Patt. Anal. Mach. Intell.* 8(4): 464–471.

Young, B.-S. J. and Chellappa, R. 1992, Statistical analysis of inherent ambiguities in receiving 3-D motion from a noisy flow field. *IEEE Trans. Patt. Anal. Mach. Intell.* 14(10): 995–1013.

R. Zabrodsky, R., Peleg, S., and Avnir, D. 1995, Symmetry as a continuous feature. *IEEE Trans. Patt. Anal. Mach. Intell.* 17(12): 1154–1166.

Zhang, P. 1992, On the distributional properties of model selection criteria. *J. Amer. Statist. Assoc.* 87: 732–737.