PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Statistical optimization and geometric inference in computer vision

The Royal Society

Phil. Trans. R. Soc. Lond. A 1998 **356**, 1303-1320
doi: 10.1098/rsta.1998.0223

The Royal Society

Statistical optimization and geometric inference in computer vision

By Kenichi Kanatani

Department of Computer Science, Gunma University, Kiryu, Gunma 376-9515, Japan

This paper gives a mathematical formulation to the computer vision task of inferring three-dimensional structures of a scene based on image data and geometric constraints. Introducing a statistical model of image noise, I define a geometric model as a manifold determined by the constraints and view the problem as model fitting. I then present a general mathematical framework for proving optimality of estimation, deriving optimal schemes, and selecting appropriate models. Finally, I illustrate the theory by applying it to curve fitting and structure from motion.

Keywords: accuracy bound; AIC; Cramer–Rao lower bound; curve fitting; model selection; structure from motion

1. Introduction

The goal of computer vision is to infer three-dimensional structures of the scene from image data. The key that makes this possible is our prior knowledge about the environment. This knowledge takes the form of *constraints*: if something is to be seen in the image, the image data should satisfy a certain relationship; if another thing is to be seen, another relationship should hold. Let us call a particular constraint a (*geometric*) *model*. The inference takes the following two stages:

1. *Model selection.* We decide which model is to be adopted from among all possibilities.
2. *Model fitting.* We obtain a detailed description of the selected model by optimally fitting it to the data.

In this paper, I give a mathematical formulation to these tasks in very general terms and apply it to curve fitting and structure from motion as typical computer vision problems.

2. Model fitting

Suppose we observe m-dimensional vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ constrained to be in an m'-dimensional manifold $\mathcal{A} \in \mathcal{R}^m$, which we call the *data space*. We write

$$\boldsymbol{a}_\alpha = \bar{\boldsymbol{a}}_\alpha + \Delta\boldsymbol{a}_\alpha, \tag{2.1}$$

where $\bar{\boldsymbol{a}}_\alpha$ is the position supposedly observed in the absence of noise. We regard the noise term $\Delta\boldsymbol{a}_\alpha$ as a Gaussian random variable of mean $\boldsymbol{0}$ and covariance matrix

$V[\boldsymbol{a}_\alpha]$, independent for each α. Since each $\boldsymbol{a}_\alpha$ is constrained to be in $\mathcal{A}$, its covariance matrix

$$V[\boldsymbol{a}_\alpha] = E[\Delta\boldsymbol{a}_\alpha \Delta\boldsymbol{a}_\alpha^{\mathrm{T}}] \tag{2.2}$$

is singular ($E[\,\cdot\,]$ denotes expectation, and 'T' denotes transpose). We assume that it is a positive semi-definite symmetric matrix of rank m' whose range coincides with the tangent space $T_{\bar{\boldsymbol{a}}_\alpha}(\mathcal{A})$ to the data space $\mathcal{A}$ at $\bar{\boldsymbol{a}}_\alpha$. We also assume that $V[\boldsymbol{a}_\alpha]$ is known *only up to scale*, i.e. we decompose it into the *noise level* ϵ and the *normalized covariance matrix* $V_0[\boldsymbol{a}_\alpha]$ in the form

$$V[\boldsymbol{a}_\alpha] = \epsilon^2 V_0[\boldsymbol{a}_\alpha], \tag{2.3}$$

and assume that $V_0[\boldsymbol{a}_\alpha]$ (which may be different for each datum) is known, but ϵ is unknown.

Suppose the true values $\bar{\boldsymbol{a}}_\alpha$, $\alpha = 1, \ldots, N$, are known to satisfy a set of equations parametrized by an n-dimensional vector $\boldsymbol{u}$. We assume that the domain of the vector $\boldsymbol{u}$ is an n'-dimensional manifold $\mathcal{U} \subset \mathcal{R}^n$, which we call the *parameter space*. Let $F^{(k)}(\boldsymbol{a}, \boldsymbol{u})$: $\mathcal{R}^m \times \mathcal{R}^n \to \mathcal{R}$, $k = 1, \ldots, L$, be smooth functions of arguments $\boldsymbol{a} \in \mathcal{R}^m$ and $\boldsymbol{u} \in \mathcal{R}^n$, and consider the following problem:

Problem 1. *Estimate the value $\boldsymbol{u} \in \mathcal{U}$ that satisfies*

$$F^{(k)}(\bar{\boldsymbol{a}}_\alpha, \boldsymbol{u}) = 0, \qquad k = 1, \ldots, L, \tag{2.4}$$

from the noisy data $\boldsymbol{a}_\alpha \in \mathcal{A}$, $\alpha = 1, \ldots, N$.

The L equations $F^{(k)}(\boldsymbol{a}, \boldsymbol{u}) = 0$, $k = 1, \ldots, L$, need not be algebraically independent with respect to the argument $\boldsymbol{a}$; we call the number r of independent equations the *rank* of the constraint. We assume that these L equations define a manifold $\mathcal{S} \subset \mathcal{A}$ of *codimension* r parametrized by $\boldsymbol{u} \in \mathcal{U}$; we call $\mathcal{S}$ the *geometric model* of equation (2.4). Then, problem 1 can be rephrased as the following *geometric model fitting*:

Problem 2. *Given noisy data $\{\boldsymbol{a}_\alpha\} \in \mathcal{A}$, let the model $\mathcal{S} \subset \mathcal{A}$ pass through their true positions $\{\bar{\boldsymbol{a}}_\alpha\} \in \mathcal{A}$ exactly by adjusting the parameter $\boldsymbol{u}$.*

3. Theoretical accuracy bound

Let $\hat{\boldsymbol{u}}$ be an arbitrary *unbiased* estimator of $\boldsymbol{u}$. The unbiasedness is usually defined by $E[\hat{\boldsymbol{u}}] = \boldsymbol{u}$. However, the parameter space $\mathcal{U}$ is generally 'curved'. Hence, although $\hat{\boldsymbol{u}} \in \mathcal{U}$, we have $E[\hat{\boldsymbol{u}}] \notin \mathcal{U}$ in general. Here, we define the unbiasedness by

$$E[\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}}(\hat{\boldsymbol{u}} - \boldsymbol{u})] = \boldsymbol{0}, \tag{3.1}$$

where $\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}}$ is the projection matrix onto the tangent space $T_{\boldsymbol{u}}(\mathcal{U})$ to the parameter space $\mathcal{U}$ at $\boldsymbol{u}$. Define the *a posteriori covariance matrix* of the estimator $\hat{\boldsymbol{u}}$ by

$$V[\hat{\boldsymbol{u}}] = \boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} E[(\hat{\boldsymbol{u}} - \boldsymbol{u})(\hat{\boldsymbol{u}} - \boldsymbol{u})^{\mathrm{T}}] \boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}}, \tag{3.2}$$

which defines a positive semi-definite symmetric matrix whose range is restricted to be in $T_{\boldsymbol{u}}(\mathcal{U})$. The estimator $\hat{\boldsymbol{u}}$ is assumed to be such that $\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}}(\hat{\boldsymbol{u}} - \boldsymbol{u})$ distributes in all directions in $T_{\boldsymbol{u}}(\mathcal{U})$ so that the range of $V[\hat{\boldsymbol{u}}]$ coincides with $T_{\boldsymbol{u}}(\mathcal{U})$, having rank n' (= the dimension of $\mathcal{U}$).

Let $\nabla_{\boldsymbol{u}}(\cdot)$ and $\nabla_{\boldsymbol{a}}(\cdot)$ denote the n-dimensional vector $(\partial(\cdot)/\partial u_1, \dots, \partial(\cdot)/\partial u_n)^{\mathrm{T}}$ and the m-dimensional vector $(\partial(\cdot)/\partial a_1, \dots, \partial(\cdot)/\partial a_m)^{\mathrm{T}}$, respectively. The following inequality can be proved (Kanatani 1996*a*):

$$V[\hat{\boldsymbol{u}}] \succ \epsilon^2 \left(\sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} \bar{W}_{\alpha}^{(kl)} (\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} \nabla_{\boldsymbol{u}} \bar{F}_{\alpha}^{(k)}) (\boldsymbol{P}_{\boldsymbol{u}}^{\mathcal{U}} \nabla_{\boldsymbol{u}} \bar{F}_{\alpha}^{(l)})^{\mathrm{T}} \right)^{-}, \tag{3.3}$$

$$(\bar{W}_{\alpha}^{(kl)}) = ((\nabla_{\boldsymbol{a}} \bar{F}_{\alpha}^{(k)}, V_0[\boldsymbol{a}_{\alpha}] \nabla_{\boldsymbol{a}} \bar{F}_{\alpha}^{(l)}))^{-}. \tag{3.4}$$

Here, $(\,\cdot\,)^{-}$ denotes the (Moore–Penrose) generalized inverse; $\nabla_{\boldsymbol{u}} \bar{F}_{\alpha}^{(k)}$ and $\nabla_{\boldsymbol{a}} \bar{F}_{\alpha}^{(k)}$ are the abbreviations of $\nabla_{\boldsymbol{u}} F^{(k)}(\bar{\boldsymbol{a}}_{\alpha}, \boldsymbol{u})$ and $\nabla_{\boldsymbol{a}} F^{(k)}(\bar{\boldsymbol{a}}_{\alpha}, \boldsymbol{u})$, respectively. The relation $\boldsymbol{A} \succ \boldsymbol{B}$ for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is a positive semi-definite matrix. Equation (3.4) means that $\bar{W}_{\alpha}^{(kl)}$ is the (kl) element of the generalized inverse of the matrix whose (kl) element is defined by the expression inside $(\,\cdot\,)^{-}$. It can be proved that $(\bar{W}_{\alpha}^{(kl)})$ is a positive semi-definite symmetric matrix of rank r (Kanatani 1996*a*).

Kanatani (1996*a*) called equation (3.3) the *Cramer–Rao inequality* and the right-hand side the *Cramer–Rao lower bound* because equation (3.3) is derived solely from the unbiasedness condition (3.1) and the definition (3.2) of $V[\hat{\boldsymbol{u}}]$ just as the usual Cramer–Rao lower bound. Equation (3.3) is a special case of a more general form for a non-Gaussian noise distribution, for which the *Fisher information matrix* plays the role of the covariance matrix (Kanatani 1996*a*).

Remark 1. In statistics, a (*statistical*) *model* is a mechanism that *predicts* future observation: it consists of a deterministic part which specifies the structure of the phenomenon and a random fluctuation part which accounts for all factors not modelled in the deterministic part. Mathematically, the problem is to estimate the parameters involved in a probability density by observing data sampled from it. Problem 1 does not fit in this framework: the structure is given as an implicit relationship in the form of equation (2.4). So, the usual Cramer–Rao lower bound does not apply here.

4. Computation methods

The following three are the most widely used methods for solving problem 1:

1. *Minimal approach.* Equation (2.4) gives r constraints on $\boldsymbol{u}$, which has n' degrees of freedom, so we observe $\lceil n'/r \rceil$ data, substitute them for their true values in equation (2.4), and solve the resulting simultaneous equations:

$$F^{(k)}(\boldsymbol{a}_{\alpha}, \boldsymbol{u}) = 0, \qquad k = 1, \dots, L, \quad \alpha = 1, \dots, \lceil n'/r \rceil. \tag{4.1}$$

2. *Least-squares approach.* We observe N ($\geqslant \lceil n'/r \rceil$) data and compute the value of $\boldsymbol{u}$ that minimizes

$$J_{LS} = \sum_{\alpha=1}^{N} \sum_{k=1}^{L} W_{\alpha}^{(k)} F^{(k)}(\boldsymbol{a}_{\alpha}, \boldsymbol{u})^2 \tag{4.2}$$

for appropriate weights $W_{\alpha}^{(k)}$ ($\geqslant 0$).

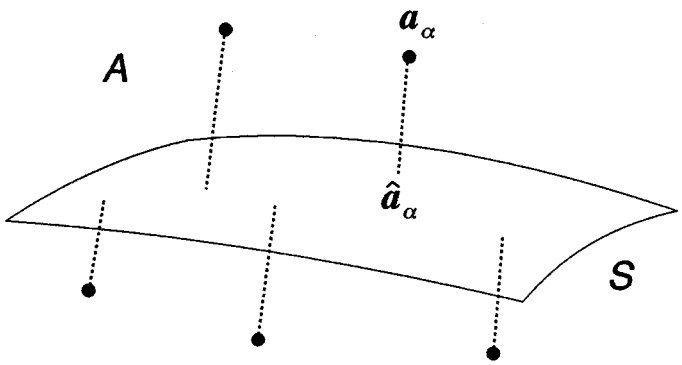

Figure 1. Datum $\boldsymbol{a}_\alpha$ is projected onto manifold $\mathcal{S}$.

3. *Maximum likelihood approach.* We compute the values $\{\bar{\boldsymbol{a}}_\alpha\}$ that minimize

$$J = \sum_{\alpha=1}^{N} (\boldsymbol{a}_\alpha - \bar{\boldsymbol{a}}_\alpha, V_0[\boldsymbol{a}_\alpha]^-(\boldsymbol{a}_\alpha - \bar{\boldsymbol{a}}_\alpha)), \tag{4.3}$$

subject to the constraint (2.4); the value $\boldsymbol{u}$ is chosen so as to minimize J (we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$).

Remark 2. The summand in equation (4.3) is known as the squared *Mahalanobis distance* of $\boldsymbol{a}_\alpha$ from $\bar{\boldsymbol{a}}_\alpha$ with respect to $V_0[\boldsymbol{a}_\alpha]$. Kanatani (1996*a*) called minimization of equation (4.3) *MLE* (*maximum likelihood estimation*) because the likelihood that the data $\{\boldsymbol{a}_\alpha\}$ are observed is const. $\times\, \mathrm{e}^{-J/2\epsilon^2}$ under our assumption; minimizing J is equivalent to maximizing the likelihood.

Remark 3. Many variants exist for the first two methods. For example, we may regard $\boldsymbol{u}$ as an unconstrained variable. This usually simplifies the computation, but the resulting value of $\boldsymbol{u}$ does not necessarily satisfy the constraint $\boldsymbol{u} \in \mathcal{U}$ (the *decomposability condition*). So, we project it onto the manifold $\mathcal{U}$ by some criterion. This approach is called the *linearization technique*.

The MLE approach is evidently the most preferable. The minimum approach is sensitive to noise, though it can be modified into an outlier detection scheme like *RANSAC* (Fischler & Bolles 1981) by incorporating random sampling and voting. The least-squares approach may be robust, but there is no guarantee that the resulting solution is close to the true value. In fact, the solution is usually *statistically biased* whatever weights $W_\alpha^{(kl)}$ are used (Kanatani 1996*a*). In contrast, the MLE solution can be shown to be optimal in the sense we describe shortly.

5. Optimal model fitting

In order to minimize equation (4.3), we first fix the value of $\boldsymbol{u}$ and compute the values $\{\bar{\boldsymbol{a}}_\alpha\}$ that minimize J. This is equivalent to 'projecting' each datum $\boldsymbol{a}_\alpha$ onto a point $\hat{\boldsymbol{a}}_\alpha$ in $\mathcal{S}$ optimally measured in the Mahalanobis distance (figure 1). The first-order solution obtained by ignoring terms of $O(\epsilon^2)$ is given in the form

$$\hat{\boldsymbol{a}}_\alpha = \boldsymbol{a}_\alpha - V_0[\boldsymbol{a}_\alpha] \sum_{k,l=1}^{L} W_\alpha^{(kl)} F_\alpha^{(k)} \nabla_{\boldsymbol{a}} F_\alpha^{(l)}. \tag{5.1}$$

Here, $F_\alpha^{(k)}$ and $\nabla_{\boldsymbol{a}} F_\alpha^{(k)}$ are the abbreviations of $F^{(k)}(\boldsymbol{a}_\alpha, \boldsymbol{u})$ and $\nabla_{\boldsymbol{a}} F^{(k)}(\boldsymbol{a}_\alpha, \boldsymbol{u})$, respectively; $W_\alpha^{(kl)}$ is defined by replacing in equation (3.4) the true value $\bar{\boldsymbol{a}}_\alpha$ by the corresponding data value $\boldsymbol{a}_\alpha$ and the generalized inverse $(\cdots)^-$ by the rank-constrained generalized inverse $(\cdots)_r^-$, by which we mean the operation performed by replacing all the eigenvalues of $\cdots$ other than the r largest ones by zero in the canonical form and computing the (Moore–Penrose) generalized inverse (Kanatani 1996*a*). This operation is necessary for preventing numerical instability, because the operand approaches a singular matrix of rank r in the limit $\boldsymbol{a}_\alpha \to \bar{\boldsymbol{a}}_\alpha$.

Substituting equation (5.1) for $\bar{\boldsymbol{a}}_\alpha$ in equation (4.3), we can express J in terms of $\boldsymbol{u}$ alone in the form

$$J = \sum_{\alpha=1}^{N} \sum_{k,l=1}^{L} W_\alpha^{(kl)} F_\alpha^{(k)} F_\alpha^{(l)}. \tag{5.2}$$

Here, the 'weight' $W_\alpha^{(kl)}$ is not a constant but a function of $\boldsymbol{u}$. We call the value $\hat{\boldsymbol{u}}$ that minimizes equation (5.2) the *MLE estimator*, and the minimum value of J the *residual*, which we denote by $\hat{J}$. 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 (Kanatani 1996*a*). Hence, an unbiased estimator of the squared noise level ϵ^2 is obtained in the form

$$\hat{\epsilon}^2 = \frac{\hat{J}}{rN - n'}. \tag{5.3}$$

It can be shown that MLE in this form is optimal in the sense that the covariance matrix of the resulting solution attains the bound (3.3) in the first order (i.e. if terms of $O(\epsilon^4)$ are ignored) (Kanatani 1996*a*). This corresponds to the well-known fact that MLE in traditional statistics is optimal in that the resulting solution attains the Cramer–Rao lower bound *asymptotically* for repeated observations (i.e. if terms of $O(1/n^2)$ are ignored for n observations).

Remark 4. We should distinguish the 'number of observations' from the 'number of data'. Suppose we make n observations and obtain N data for each observation; all N data are assumed to be samples from a probability density of N variables that involves unknown parameters which we want to estimate. In traditional statistics, MLE is optimal in the limit $n \to \infty$.

Remark 5. In problem 1, the unknown parameters that characterize the probability density of the data are the true values $\{\bar{\boldsymbol{a}}_\alpha\}$ of the data $\{\boldsymbol{a}_\alpha\}$ and the value $\boldsymbol{u}$ that parametrizes the constraint. Hence, as the number of data increases, the number of unknowns increases at the same rate. This means that if we observe an increased number of data, they are still 'one' sample from a new probability density that involves an increased number of unknowns. In other words, the number of observations is always $n = 1$, however large the number of data N is. Thus, asymptotic analysis in the limit $N \to \infty$ does not make sense for problem 1.

Remark 6. Suppose we hypothetically repeat observations of the *same* values $\{\bar{\boldsymbol{a}}_\alpha\}$ n times and obtain n sets of data $\{\boldsymbol{a}_\alpha^{(i)}\}$, $i = 1, \ldots, n$, $\alpha = 1, \ldots, N$. If we take the average

$$\boldsymbol{a}_\alpha = \sum_{i=1}^{n} \boldsymbol{a}_\alpha^{(i)}/n,$$

the resulting values $\{\boldsymbol{a}_\alpha\}$ have errors of $O(1/\sqrt{n})$ times the original errors. In other words, increasing the number of (hypothetical) observations n means *effectively*

reducing the noise level ϵ. Thus, it is natural that the optimality of MLE, which holds asymptotically (i.e. $n \to \infty$) in traditional statistics, should hold for small noise (i.e. $\epsilon \to 0$) for problem 1.

6. Model selection

If there are multiple candidates of geometric models $\mathcal{S}_1, \mathcal{S}_2, \ldots$, we should choose the 'best' one for observed data $\{\boldsymbol{a}_\alpha\}$. But how can we evaluate the 'goodness' of a model $\mathcal{S}$? First of all, a good model should explain the data $\{\boldsymbol{a}_\alpha\}$ well, which implies that the residual $\hat{J}$ should be small. Since $\hat{J}/\epsilon^2$ is subject to a χ^2 distribution with $rN-n'$ degrees of freedom, the residual becomes smaller as n' becomes larger. In particular, the residual is zero in general if the model $\mathcal{S}$ has more than rN free parameters, meaning that we can make $\mathcal{S}$ pass through all the data points by adjusting the parameters. Such an artificial model cannot be regarded as a good model, because it only explains the *current data* $\{\boldsymbol{a}_\alpha\}$ which happen to be observed; there is no guarantee that it could explain the data if the noise occurred differently.

This observation suggests that the 'goodness' of a model can be measured by its 'predicting capability' (Akaike 1974). Let $\{\boldsymbol{a}^*_\alpha\}$ be the *future data* that have the same probability distribution as the current data $\{\boldsymbol{a}_\alpha\}$ but are independent of $\{\boldsymbol{a}_\alpha\}$. Consider the residual J^* for the maximum likelihood estimators $\{\hat{\boldsymbol{a}}_\alpha\} \in \hat{\mathcal{S}}$, which are computed from the current data $\{\boldsymbol{a}_\alpha\}$, *with respect to the future data* $\{\boldsymbol{a}^*_\alpha\}$:

$$J^* = \sum_{\alpha=1}^{N} (\boldsymbol{a}^*_\alpha - \hat{\boldsymbol{a}}_\alpha, V_0[\boldsymbol{a}_\alpha]^-(\boldsymbol{a}^*_\alpha - \hat{\boldsymbol{a}}_\alpha)). \tag{6.1}$$

A model $\mathcal{S}$ is expected to have high prediction capability if J^* is small. Put differently, we are 'validating' the optimal estimate $\hat{\mathcal{S}}$ of $\mathcal{S}$ by measuring its discrepancy from data *yet to be observed*. It can be shown that $\hat{J}$ *is smaller than* J^* *by* $2(pN+n')\epsilon^2$ *in expectation* (Kanatani 1996*a*). So, we define the *geometric information criterion*, or the *geometric Akaike information criterion* (AIC), by

$$\mathrm{AIC}(\mathcal{S}) = \hat{J} + 2(pN + n')\epsilon^2, \tag{6.2}$$

and use it as a measure of the goodness of the model.

Remark 7. The most distinctive characteristic of the geometric AIC is the fact that the number N of data, which does not appear in the usual AIC, explicitly appears in the expression; the traditional AIC contains, other than the residual, only the number of the model parameters. This is because, as we pointed out earlier, the number of data in traditional statistics means the number of 'observations', while the number N of the data $\{\boldsymbol{a}_\alpha\}$ means the number of the parameters $\{\bar{\boldsymbol{a}}_\alpha\}$ (often called *nuisance parameters*) of the model; *the number of observations is always one*. This fact results in the following features of the geometric AIC:

1. 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.

2. The dimension p of the model manifold plays a dominant role in the geometric AIC, while no such geometric concepts are involved in the usual AIC.

7. Noise estimation

In order to compute the geometric AIC, 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. However, estimating the noise level *a priori* is in general very difficult. Here, we resolve this difficulty by focusing on the *inclusion relationship* of the models.

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. Let $\hat{J}_1$ and $\hat{J}_2$ be their respective residuals. 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

$$\mathcal{S}_2 \succ \mathcal{S}_1. \tag{7.1}$$

If model $\mathcal{S}_1$ is correct, the squared noise level ϵ^2 is estimated by equation (5.3). Substituting it into the expression for the geometric AIC, we obtain

$$\mathrm{AIC}(\mathcal{S}_1) = \hat{J}_1 + \frac{2(p_1N + n'_1)}{r_1N - n'_1}\hat{J}_1, \qquad \mathrm{AIC}(\mathcal{S}_2) = \hat{J}_2 + \frac{2(p_2N + n'_2)}{r_1N - n'_1}\hat{J}_1. \tag{7.2}$$

Since the geometric AIC estimates the expected sum of squared Mahalanobis distances, the ratio of the deviations from the two models can be evaluated by

$$K = \sqrt{\frac{\mathrm{AIC}(\mathcal{S}_2)}{\mathrm{AIC}(\mathcal{S}_1)}} = \sqrt{\frac{r_1N - n'_1}{(2p_1 + r_1)N + n'_1}\left(\frac{\hat{J}_2}{\hat{J}_1} + \frac{2(p_2N + n'_2)}{r_1N - n'_1}\right)}. \tag{7.3}$$

This quantity measures *how good* model $\mathcal{S}_2$ is *compared with model* $\mathcal{S}_1$: if $K < 1$, model $\mathcal{S}_2$ is expected to have more predicting capability.

Remark 8. Our approach for noise estimation is very different from that in statistics, where the noise level is estimated *model by model* in such a way that the AIC is minimized, which is equivalent to MLE. If the noise level is so estimated and substituted back into the AIC expression, the residual term is effectively replaced by its logarithm. We cannot do that in our problem for two reasons:

1. In statistics, noise is a characteristic of the model, which specifies how deterministic causes are separated from random effects. Since noise characteristics are model-dependent, it makes sense to estimate them model by model. In our problem, noise is a characteristic of the devices and data-processing operations involved and is *independent of the models we are comparing*. Hence, it should be estimated once for all the models.

2. MLE of the noise level produces a very poor estimator for our problem. In fact, if we estimate ϵ^2 by MLE, its expectation is approximately $r\epsilon^2/m'$, e.g. it is $\epsilon^2/2$ for curve fitting in two dimensions and $\epsilon^2/3$ for surface fitting in three dimensions. This does not occur in traditional statistics, because MLE estimators rapidly approach their true values as the number of observations increases, i.e. as the number of data increases. In our problem, the number of observations is always one; increasing the number of data has no effect in improving the accuracy of estimating ϵ^2. In fact, the optimality of MLE can be established only in the limit $\epsilon^2 \to 0$.

8. Curve fitting

Given N points $(x_1, y_1), \ldots, (x_N, y_N)$ in an image, consider the problem of fitting to them a curve

$$F(x, y, \boldsymbol{u}) = 0 \tag{8.1}$$

parametrized by a vector $\boldsymbol{u}$. Let $(\bar{x}_\alpha, \bar{y}_\alpha)$ be the true position of point (x_α, y_α), i.e. the position that would supposedly be observed if the image were ideal and the detection operation were accurate. We write

$$x_\alpha = \bar{x}_\alpha + \Delta x_\alpha, \qquad y_\alpha = \bar{y}_\alpha + \Delta y_\alpha, \tag{8.2}$$

and regard $(\Delta x_\alpha, \Delta y_\alpha)$ as a two-dimensional random variable subject to a Gaussian distribution, independent for each α, of mean $(0, 0)$ and variance-covariance

$$V[x_\alpha, y_\alpha] = \begin{pmatrix} E[\Delta x_\alpha^2] & E[\Delta x_\alpha \Delta y_\alpha] \\ E[\Delta y_\alpha \Delta x_\alpha] & E[\Delta y_\alpha^2] \end{pmatrix}, \tag{8.3}$$

which we decompose into an unknown noise level ϵ and a known normalized covariance matrix $V_0[x_\alpha, y_\alpha]$ in the form of equation (2.3).

Curve fitting has customarily been defined as finding a curve that passes *as closely to the data points as possible*. In contrast, we state the problem as follows:

Problem 3. *Estimate the value* $\boldsymbol{u}$ *that satisfies*

$$F(\bar{x}_\alpha, \bar{y}_\alpha, \boldsymbol{u}) = 0, \qquad \alpha = 1, \ldots, N, \tag{8.4}$$

from the noisy data $\{(x_\alpha, y_\alpha)\}$.

For this problem, inequality (3.3) takes the following form:

$$V[\hat{\boldsymbol{u}}] \succ \epsilon^2 \left(\sum_{\alpha=1}^{N} \frac{(\nabla_{\boldsymbol{u}} \bar{F}_\alpha)(\nabla_{\boldsymbol{u}} \bar{F}_\alpha)^{\mathrm{T}}}{(\nabla \bar{F}_\alpha, V_0[x_\alpha, y_\alpha] \nabla \bar{F}_\alpha)} \right)^{-1}. \tag{8.5}$$

This bound is attained in the first order by MLE; we minimize

$$J = \sum_{\alpha=1}^{N} \frac{F_\alpha^2}{(\nabla F_\alpha, V_0[x_\alpha, y_\alpha] \nabla F_\alpha)}. \tag{8.6}$$

This criterion is a generalization of the techniques suggested by Bolle & Verumi (1991), Taubin (1991), and Taubin *et al.* (1994), who tacitly assumed homogeneous and isotropic noise.

The covariance matrix $V[\hat{\boldsymbol{u}}]$ of the MLE estimator $\hat{\boldsymbol{u}}$ can be evaluated in the first order by substituting the data $\{(x_\alpha, y_\alpha)\}$ and the estimate $\hat{\boldsymbol{u}}$ for the true values $\{(\bar{x}_\alpha, \bar{y}_\alpha)\}$ and $\boldsymbol{u}$, respectively, in equation (8.5). Since $V[\hat{\boldsymbol{u}}]$ is a positive definite symmetric matrix, it has the spectral decomposition in the form

$$V[\hat{\boldsymbol{u}}] = \sum_{i=1}^{n'} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathrm{T}}, \tag{8.7}$$

where λ_i is the ith largest eigenvalue of $V[\hat{\boldsymbol{u}}]$, and $\boldsymbol{u}_i$ is the corresponding unit eigenvector. The vector $\boldsymbol{u}_1$ indicates the orientation along which deviation is most

likely to occur; $\sqrt{\lambda_1}$ is its standard deviation. Thus, the following two vectors can be regarded as the most typical instances of the parameter $\boldsymbol{u}$:

$$\boldsymbol{u}^+ = \hat{\boldsymbol{u}} + \sqrt{\lambda_1}\boldsymbol{u}_1, \qquad \boldsymbol{u}^- = \hat{\boldsymbol{u}} - \sqrt{\lambda_1}\boldsymbol{u}_1. \tag{8.8}$$

It follows that we can *visualize* the reliability of the optimal fit by displaying two curves $F(x, y, \boldsymbol{u}^+) = 0$ and $F(x, y, \boldsymbol{u}^-) = 0$, which we call the *primary deviation pair* for the optimal fit $F(x, y, \hat{\boldsymbol{u}}) = 0$ (Kanazawa & Kanatani 1995*a*, *b*).

Typical examples of the above procedure include line fitting (Kanazawa & Kanatani 1996*a*), conic fitting (Kanazawa & Kanatani 1996*b*), and fitting a plane to range data (Kanazawa & Kanatani 1995*c*). The bound (8.5) plays an important role in evaluating the accuracy of camera calibration using a planar grid pattern (Kanatani 1992, 1993*a*; Kanatani & Maruyama 1996).

9. Renormalization

Typical curves to fit are a line $Ax + By + C = 0$ and a conic $Ax^2 + 2Bxy + Cy^2 + 2(Dx + Ey) + F = 0$. In most cases, curves to fit have the form

$$(\boldsymbol{a}, \boldsymbol{u}) = 0, \tag{9.1}$$

where $\boldsymbol{u}$ is a parameter vector to estimate and $\boldsymbol{a}$ is a vector whose elements are *polynomial* functions of x and y: for line fitting, we have $\boldsymbol{u} = (A, B, C)^{\mathrm{T}}$ and $\boldsymbol{a} = (x, y, 1)^{\mathrm{T}}$; for conic fitting, we have

$$\boldsymbol{u} = (A, B, C, D, E, F)^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{a} = (x^2, 2xy, y^2, 2x, 2y, 1)^{\mathrm{T}}.$$

From the variances and covariances $E[\Delta x^2]$, $E[\Delta y^2]$, $E[\Delta x \Delta y]$ of the noise terms Δx and Δy, we can compute variances and covariances of any polynomial functions of x and y. In fact, if we let $x_1 = x$ and $x_2 = y$ and put $E[\Delta x_i \Delta x_j] = \Sigma_{ij}$, we can reduce higher-order moments to expressions in Σ_{ij} by using the well-known property for Gaussian random variables:

$$\left.\begin{aligned} E[\Delta x_i \Delta x_j \Delta x_k] &= 0, \\ E[\Delta x_i \Delta x_j \Delta x_k \Delta x_l] &= \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}, \\ E[\Delta x_i \Delta x_j \Delta x_k \Delta x_l \Delta x_m] &= 0, \\ E[\Delta x_i \Delta x_j \Delta x_k \Delta x_l \Delta x_m \Delta x_n] &= \cdots . \end{aligned}\right\} \tag{9.2}$$

Thus, we can obtain the covariance matrix $V[\boldsymbol{a}]$ of the vector $\boldsymbol{a}$ as a function of x and y.

Since the absolute magnitude of the vector $\boldsymbol{u}$ in equation (9.1) is arbitrary, we can normalize it into a unit vector. Hence, if $\boldsymbol{a}$ and $\boldsymbol{u}$ are n-dimensional vectors, the parameter space $\mathcal{U}$ is a unit $(n-1)$-sphere $\|\boldsymbol{u}\| = 1$ in $\mathcal{R}^n$. The data space $\mathcal{A}$ is two-dimensional, since $\boldsymbol{a}$ is parametrized by x and y. It follows the covariance matrix $V[\boldsymbol{a}]$ has dimension n but has rank 2.

In this formalism, an optimal solution can be computed semi-analytically by a procedure called *renormalization* (Kanatani 1994*a*; Kanazawa & Kanatani 1995*a*, *c*). For each datum $\boldsymbol{a}_\alpha$, we decompose $V[\boldsymbol{a}_\alpha]$ into an unknown noise level ϵ and a known normalized covariance matrix $V_0[\boldsymbol{a}_\alpha]$ in the form of equation (2.3). The renormalization procedure goes as follows:

1. Let $c = 0$ and $W_\alpha = 1$, $\alpha = 1, \ldots, N$.

2. Compute the following matrices:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{a}_\alpha \boldsymbol{a}_\alpha^{\mathrm{T}}, \qquad \boldsymbol{N} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{a}_\alpha]. \tag{9.3}$$

3. Compute the smallest eigenvalue λ of the matrix

$$\hat{\boldsymbol{M}} = \boldsymbol{M} - c\boldsymbol{N}, \tag{9.4}$$

and the corresponding unit eigenvector $\boldsymbol{u}$.

4. If $\lambda \approx 0$, return $\boldsymbol{u}$, c, and $\boldsymbol{M}$. Otherwise, update c and W_α as follows:

$$c \leftarrow c + \frac{\lambda}{(\boldsymbol{u}, \boldsymbol{N}\boldsymbol{u})}, \qquad W_\alpha \leftarrow \frac{1}{(\boldsymbol{u}, V_0[\boldsymbol{a}_\alpha]\boldsymbol{u})}. \tag{9.5}$$

5. Go back to Step 2.

The convergence is very fast; usually three or four iterations result in sufficient accuracy. Let $\hat{\boldsymbol{u}}$ be the returned value of $\boldsymbol{u}$. It can be shown that this estimator has the same accuracy as the MLE estimator in the first order (Kanatani 1996*a*). From this fact, we can obtain an unbiased estimator of the squared noise level ϵ^2 in the form

$$\hat{\epsilon}^2 = \frac{c}{1 - 2/N}. \tag{9.6}$$

The covariance matrix of the estimator $\hat{\boldsymbol{n}}$ is estimated by

$$V[\hat{\boldsymbol{u}}] = \frac{\hat{\epsilon}^2}{N}(\hat{\boldsymbol{M}})^{-}_{n-1}. \tag{9.7}$$

The rank-constrained generalized inverse operation $(\cdots)^{-}_{n-1}$ is necessary because the smallest eigenvalue of $\hat{\boldsymbol{M}}$ may not be strictly 0 if the renormalization iterations are prematurely terminated.

10. Three-dimensional motion analysis

Define an XYZ camera coordinate system in such a way that the origin O is at the centre 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. Suppose the camera moves to a position defined by translating the first camera by vector $\boldsymbol{h}$ and rotating it around the centre of the lens by matrix $\boldsymbol{R}$; we call $\{\boldsymbol{h}, \boldsymbol{R}\}$ the *motion parameters* (figure 2).

Let (x_α, y_α), $\alpha = 1, \ldots, N$, be the image coordinates of N feature points before the motion, and (x'_α, y'_α), $\alpha = 1, \ldots, N$, be those after the motion. We use the following three-dimensional vectors to represent them:

$$\boldsymbol{x}_\alpha = (x_\alpha, y_\alpha, 1)^{\mathrm{T}}, \qquad \boldsymbol{x}'_\alpha = (x'_\alpha, y'_\alpha, 1)^{\mathrm{T}}. \tag{10.1}$$

We regard these as Gaussian random variables and denote their covariance matrices by $V[\boldsymbol{x}_\alpha]$ and $V[\boldsymbol{x}'_\alpha]$. Since the third components of $\Delta\boldsymbol{x}_\alpha$ and $\Delta\boldsymbol{x}'_\alpha$ are identically 0, $V[\boldsymbol{x}_\alpha]$ and $V[\boldsymbol{x}'_\alpha]$ are singular matrices of rank 2. Assuming that the covariance matrices are known only up to scale, we decompose them into the noise level ϵ and the normalized covariance matrices $V_0[\boldsymbol{x}_\alpha]$ and $V_0[\boldsymbol{x}'_\alpha]$ in the form of equation (2.3).

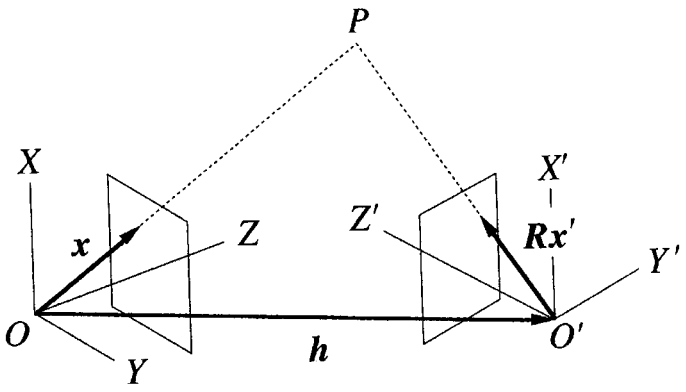

Figure 2. Geometry of camera motion.

(*a*) *General model*

Since vectors $\{\boldsymbol{x}, \boldsymbol{x}'\}$ represent the same point in the scene if and only if three vectors $\boldsymbol{x}$, $\boldsymbol{R}\boldsymbol{x}'$, and $\boldsymbol{h}$ are coplanar, we obtain the *epipolar equation* (Faugeras 1993; Kanatani 1993*a*; Weng *et al.* 1993*a*)

$$|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{R}\boldsymbol{x}'| = 0, \tag{10.2}$$

where $|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}|$ denotes the scalar triple product of vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. This equation defines a model $\mathcal{S}$ of codimension 1 in the four-dimensional data space $\mathcal{X} = \{(x, y, z, x', y', z') | z = 1, z' = 1\} \in \mathcal{R}^6$. The scale of the translation $\boldsymbol{h}$ is indeterminate, so we normalize it into $\|\boldsymbol{h}\| = 1$. As a result, the parameter space $\mathcal{U}$ is a five-dimensional manifold in $\mathcal{R}^{12}$; it is the collection of points

$$(h_1, h_2, h_3, R_{11}, R_{12}, \ldots, R_{33})$$

such that (h_i) is a unit vector and (R_{ij}) is a rotation matrix.

1. *Errors in translation.* Since translations constitute an additive group, we measure the deviation of the computed value $\hat{\boldsymbol{h}}$ from its true value $\boldsymbol{h}$ by the 'difference' $\hat{\boldsymbol{h}} - \boldsymbol{h}$. Both $\hat{\boldsymbol{h}}$ and $\boldsymbol{h}$ are normalized to a unit vector, so they are both on a unit sphere. Assuming that error is small, we identify the domain of the error $\Delta\boldsymbol{h}$ with the tangent plane to the sphere at $\boldsymbol{h}$ and define the error of translation by

$$\Delta\boldsymbol{h} = \boldsymbol{P}_{\boldsymbol{h}}(\hat{\boldsymbol{h}} - \boldsymbol{h}), \tag{10.3}$$

where $\boldsymbol{P}_{\boldsymbol{h}}$ $(= \boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^{\mathrm{T}})$ is the projection matrix onto a subspace perpendicular to $\boldsymbol{h}$ ($\boldsymbol{I}$ denotes the unit matrix).

2. *Errors in rotation.* Since rotations constitute a multiplicative group, we measure the deviation of the computed rotation $\hat{\boldsymbol{R}}$ from the true rotation $\boldsymbol{R}$ by the 'quotient' $\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}$, which is a small rotation. Let $\boldsymbol{l}$ and $\Delta\Omega$ be its axis (unit vector) and angle of rotation, respectively. We define the error of rotation by

$$\Delta\boldsymbol{\Omega} = \Delta\Omega\boldsymbol{l}. \tag{10.4}$$

From these observations, we define the covariance matrices of the motion parameters $\{\hat{\boldsymbol{h}}, \hat{\boldsymbol{R}}\}$ as follows (Kanatani 1993*a*):

$$\left.\begin{aligned} V[\hat{\boldsymbol{h}}] &= E[\Delta\boldsymbol{h}\Delta\boldsymbol{h}^{\mathrm{T}}], & V[\hat{\boldsymbol{h}}, \hat{\boldsymbol{R}}] &= E[\Delta\boldsymbol{h}\Delta\boldsymbol{\Omega}^{\mathrm{T}}], \\ V[\hat{\boldsymbol{R}}, \hat{\boldsymbol{h}}] &= E[\Delta\boldsymbol{\Omega}\Delta\boldsymbol{h}^{\mathrm{T}}], & V[\hat{\boldsymbol{R}}] &= E[\Delta\boldsymbol{\Omega}\Delta\boldsymbol{\Omega}^{\mathrm{T}}]. \end{aligned}\right\} \tag{10.5}$$

For this representation, inequality (3.3) has the following form:

$$\begin{pmatrix} V[\hat{\boldsymbol{h}}] & V[\hat{\boldsymbol{h}}, \hat{\boldsymbol{R}}] \\ V[\hat{\boldsymbol{R}}, \hat{\boldsymbol{h}}] & V[\hat{\boldsymbol{R}}] \end{pmatrix} \succ \epsilon^2 \begin{pmatrix} \sum_{\alpha=1}^{N} \bar{W}_\alpha \bar{\boldsymbol{p}}_\alpha \bar{\boldsymbol{p}}_\alpha^{\mathrm{T}} & \sum_{\alpha=1}^{N} \bar{W}_\alpha \bar{\boldsymbol{p}}_\alpha \bar{\boldsymbol{q}}_\alpha^{\mathrm{T}} \\ \sum_{\alpha=1}^{N} \bar{W}_\alpha \bar{\boldsymbol{q}}_\alpha \bar{\boldsymbol{p}}_\alpha^{\mathrm{T}} & \sum_{\alpha=1}^{N} \bar{W}_\alpha \bar{\boldsymbol{q}}_\alpha \bar{\boldsymbol{q}}_\alpha^{\mathrm{T}} \end{pmatrix}^{-}, \tag{10.6}$$

$$\bar{\boldsymbol{p}}_\alpha = \bar{\boldsymbol{x}}_\alpha \times \boldsymbol{R}\bar{\boldsymbol{x}}'_\alpha, \qquad \bar{\boldsymbol{q}}_\alpha = (\bar{\boldsymbol{x}}_\alpha, \boldsymbol{R}\bar{\boldsymbol{x}}_{\alpha'})\boldsymbol{h} - (\boldsymbol{h}, \boldsymbol{R}\bar{\boldsymbol{x}}'_\alpha)\bar{\boldsymbol{x}}_\alpha, \tag{10.7}$$

$$\bar{W}_\alpha = \frac{1}{(\boldsymbol{h} \times \boldsymbol{R}\bar{\boldsymbol{x}}'_\alpha, V_0[\boldsymbol{x}_\alpha](\boldsymbol{h} \times \boldsymbol{R}\bar{\boldsymbol{x}}'_\alpha)) + (\boldsymbol{h} \times \bar{\boldsymbol{x}}_\alpha, \boldsymbol{R}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{h} \times \bar{\boldsymbol{x}}_\alpha))}. \tag{10.8}$$

This bound is attained in the first order by MLE; we minimize the following function (Kanatani 1994*b*, 1996*a*):

$$J = \sum_{\alpha=1}^{N} \frac{|\boldsymbol{x}_\alpha, \boldsymbol{h}, \boldsymbol{R}\boldsymbol{x}'_\alpha|^2}{(\boldsymbol{h} \times \boldsymbol{R}\bar{\boldsymbol{x}}'_\alpha, V_0[\boldsymbol{x}_\alpha](\boldsymbol{h} \times \boldsymbol{R}\bar{\boldsymbol{x}}'_\alpha)) + (\boldsymbol{h} \times \bar{\boldsymbol{x}}_\alpha, \boldsymbol{R}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{R}^{\mathrm{T}}(\boldsymbol{h} \times \bar{\boldsymbol{x}}_\alpha))}. \tag{10.9}$$

Direct minimization of this function requires numerical search (Kanatani 1993*b*; Weng *et al.* 1993*b*), but the computation can be simplified by combining the *linearization* technique with renormalization and an optimal correction scheme (Kanatani 1994*b*).

(*b*) *Planar surface model*

Consider a planar surface in the scene. Let $\boldsymbol{n}$ be its unit surface normal, and d its distance from the origin O (positive in the direction of $\boldsymbol{n}$); we call $\{\boldsymbol{n}, d\}$ the *surface parameters*. A pair $\{\boldsymbol{x}, \boldsymbol{x}'\}$ is a projection of a feature point on that plane if and only if

$$\boldsymbol{x}' \times \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \tag{10.10}$$

where $\boldsymbol{A}$ is a matrix that determines the projective transformation (or *homography*) of the image and has the following form (Faugeras 1993; Kanatani 1993*a*; Longuet-Higgins 1986; Maybank 1993; Weng *et al.* 1991):

$$\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}(\boldsymbol{h}\boldsymbol{n}^{\mathrm{T}} - d\boldsymbol{I}). \tag{10.11}$$

Equation (10.10) defines a model $\mathcal{S}_{\Pi}$ of codimension 2 in the four-dimensional data space $\mathcal{X} = \{(x, y, z, x', y', z') | z = 1, z' = 1\} \in \mathcal{R}^6$. Since the scale of the homography $\boldsymbol{A}$ is indeterminate, we may normalize it to $\|\boldsymbol{A}\| = 1$. So, the parameter space $\mathcal{U}$ is an eight-dimensional manifold in $\mathcal{R}^9$. An optimal solution of the surface parameters $\{\boldsymbol{n}, d\}$ and the motion parameters $\{\boldsymbol{h}, \boldsymbol{R}\}$ in the sense of MLE is obtained by minimizing the following function (Kanatani 1996*a*; Kanatani & Takeda 1995):

$$J_{\Pi} = \sum_{\alpha=1}^{N} (\boldsymbol{x}'_\alpha \times \boldsymbol{A}\boldsymbol{x}_\alpha, \boldsymbol{W}_\alpha \boldsymbol{x}'_\alpha \times \boldsymbol{A}\boldsymbol{x}_\alpha). \tag{10.12}$$

$$\boldsymbol{W}_\alpha = (\boldsymbol{x}'_\alpha \times \boldsymbol{A}V_0[\boldsymbol{x}_\alpha]\boldsymbol{A}^{\mathrm{T}} \times \boldsymbol{x}'_\alpha + (\boldsymbol{A}\boldsymbol{x}_\alpha) \times V_0[\boldsymbol{x}'_\alpha] \times (\boldsymbol{A}\boldsymbol{x}_\alpha))_2^-. \tag{10.13}$$

The product $\boldsymbol{v} \times \boldsymbol{T}$ of a vector $\boldsymbol{v}$ and a matrix $\boldsymbol{T}$ is defined to be the matrix whose columns are the vector products of $\boldsymbol{v}$ and the three columns of $\boldsymbol{T}$. For a vector $\boldsymbol{v}$ and a matrix $\boldsymbol{T}$, the symbol $\boldsymbol{v} \times \boldsymbol{T} \times \boldsymbol{v}$ is an abbreviation of $\boldsymbol{v} \times \boldsymbol{T}(\boldsymbol{v} \times \boldsymbol{I})^{\mathrm{T}}$. The renormalization procedure can be applied to compute the homography $\boldsymbol{A}$ that minimizes equation (10.12) (Kanatani 1996*a*; Kanatani & Takeda 1995). The surface parameters $\{\boldsymbol{n}, d\}$ and the motion parameters $\{\boldsymbol{h}, \boldsymbol{R}\}$ are analytically computed from the resulting matrix $\boldsymbol{A}$ (Kanatani 1993*a*; Longuet-Higgins 1986; Weng *et al.* 1991).

(*c*) *Pure rotation model*

No three-dimensional information can be obtained if the camera motion is pure rotation around the centre of the lens; all we can estimate is the amount of the camera rotation $\boldsymbol{R}$. A pair $\{\boldsymbol{x}, \boldsymbol{x}'\}$ is a projection of the same feature point if and only if

$$\boldsymbol{x} \times \boldsymbol{R}\boldsymbol{x}' = \boldsymbol{0}. \tag{10.14}$$

This equation defines a model $\mathcal{S}_R$ of codimension 2 in the four-dimensional data space $\mathcal{X} = \{(x, y, z, x', y', z') | z = 1, z' = 1\} \in \mathcal{R}^6$. The parameter space $\mathcal{U}$ is a three-dimensional manifold defined by all rotation matrices in $\mathcal{R}^9$. An optimal solution of $\boldsymbol{R}$ in the sense of MLE is obtained by minimizing the following function (Kanatani 1996*a*):

$$J_R = \sum_{\alpha=1}^{N} (\boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha, \boldsymbol{W}_\alpha \boldsymbol{x}_\alpha \times \boldsymbol{R}\boldsymbol{x}'_\alpha), \tag{10.15}$$

$$\boldsymbol{W}_\alpha = ((\boldsymbol{R}\boldsymbol{x}'_\alpha) \times V_0[\boldsymbol{x}_\alpha] \times (\boldsymbol{R}\boldsymbol{x}'_\alpha) + \boldsymbol{x}_\alpha \times \boldsymbol{R}V_0[\boldsymbol{x}'_\alpha]\boldsymbol{R}^{\mathrm{T}} \times \boldsymbol{x}_\alpha)_2^-. \tag{10.16}$$

(*d*) *Model selection*

Equation (10.10) implies the epipolar equation (10.2). Equation (10.13) is obtained from equation (10.10) by letting $d = 1$ and $\boldsymbol{h} = \boldsymbol{0}$ in equation (10.11). Hence, $\mathcal{S}_R \succ \mathcal{S}_{\Pi} \succ \mathcal{S}$. Thus, the following tests can be done (Kanatani 1996*a*, *b*):

1. *Planarity test.* The object is judged to be planar if

$$K_{\Pi} = \sqrt{\frac{N-5}{7N+5}\left(\frac{\hat{J}_{\Pi}}{\hat{J}} + \frac{4N+16}{N-5}\right)} < 1. \tag{10.17}$$

2. *Rotation test.* The camera motion is judged to be a pure rotation if

$$K_R = \sqrt{\frac{N-5}{7N+5}\left(\frac{\hat{J}_R}{\hat{J}} + \frac{4N+6}{N-5}\right)} < 1. \tag{10.18}$$

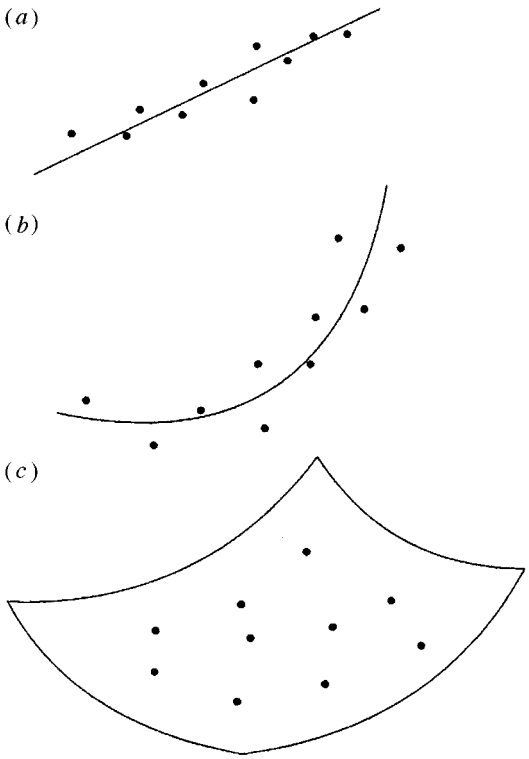

Figure 3. (*a*) A two-dimensional manifold with three degrees of freedom can be fitted. (*b*) A two-dimensional manifold with eight degrees of freedom can be fitted. (*c*) A three-dimensional manifold with five degrees of freedom can be fitted.

(*e*) *Self-evaluation*

In order to do robust three-dimensional reconstruction, the camera must be displaced over a long distance so that the resulting disparity is sufficiently large. As the disparity increases, however, it is increasingly difficult to match feature points over the two images due to possible occlusions and illumination changes. Feature matching becomes easier as the disparity decreases; for each feature point, only a small neighbourhood needs to be searched. It is therefore desirable to keep the camera displacement minimum in such a way that the resulting disparity is sufficient for reliable three-dimensional reconstruction. This can be done by measuring the 'goodness' of the images by the values (K_R, K_Π) defined by equations (10.17) and (10.18) if we note the following (figure 3):

1. *Small motion*. If the camera motion is small, the data points are concentrated near a two-dimensional manifold in the four-dimensional data space $\mathcal{X}$. As a result, we can robustly fit the rotation model $\mathcal{S}_R$, but we cannot robustly fit $\mathcal{S}_\Pi$ or $\mathcal{S}$, so we are unable to perceive any three-dimensional structure of the scene.

2. *Intermediate motion*. As the camera motion increases, the data spread more in $\mathcal{X}$, so we can robustly fit the planar surface model $\mathcal{S}_\Pi$, but we cannot robustly fit $\mathcal{S}$. As a result, the scene we can perceive is merely a planar surface.

3. *Large motion.* If the camera motion is sufficiently large, the distribution of the data is sufficiently three-dimensional in $\mathcal{X}$, so we can robustly fit the general model $\mathcal{S}$ and thereby perceive the full three-dimensional structure of the scene.

11. Concluding remarks

We have given a mathematical formulation to the computer vision task of inferring three-dimensional structures of the scene based on image data and geometric constraints. Introducing a statistical model of image noise, we defined a geometric model as a manifold determined by the constraints and viewed the problem as model fitting. We then presented a general mathematical framework for proving optimality of estimation, deriving optimal schemes, and selecting appropriate models. Finally, we illustrated our theory by applying it to curve fitting and structure from motion. Geometric model selection based on the geometric AIC has many other applications than are described here—inferring true two-dimensional structures from distorted shapes (Kanatani 1997; Triono & Kanatani 1996) and inferring three-dimensional structures by stereo vision (Kanazawa & Kanatani 1997), for instance.

This work was in part supported by the Ministry of Education, Science, Sports and Culture, Japan under a Grant in Aid for Scientific Research C(2) (No. 09680352).

References

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

Bolle, R. M. & Verumi, B. C. 1991 On three-dimensional surface reconstruction methods. *IEEE Trans. Patt. Analysis Mach. Intell.* **13**, 1–13.

Faugeras, O. 1993 *Three-dimensional computer vision: a geometric viewpoint.* Cambridge, MA: MIT Press.

Fischler, M. A. & Bolles, R. C. 1981 Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**, 381–395.

Kanatani, K. 1992 Statistical analysis of focal-length calibration using vanishing points. *IEEE Trans. Patt. Analysis Mach. Intell.* **8**, 767–775.

Kanatani, K. 1993*a* *Geometric computation for machine vision.* Oxford University Press.

Kanatani, K. 1993*b* Unbiased estimation and statistical analysis of 3-D rigid motion from two views. *IEEE Trans. Patt. Analysis Mach. Intell.* **15**, 37–50.

Kanatani, K. 1994*a* Statistical analysis of geometric computation. *CVGIP: Image Understanding* **59**, 286–306.

Kanatani, K. 1994*b* Renormalization for motion analysis: statistically optimal algorithm. *IEICE Trans. Inf. Sys.* E **77-D**, 1233–1239.

Kanatani, K. 1996*a* *Statistical optimization for geometric computation: theory and practice.* Amsterdam: Elsevier Science.

Kanatani, K. 1996*b* Automatic singularity test for motion analysis by an information criterion. In *Proc. 4th European Conf. Computer Vision, Cambridge, UK, 15–19 April 1996*, pp. 697–708.

Kanatani, K. 1997 Comments on 'Symmetry as a continuous feature'. *IEEE Trans. Patt. Analysis Mach. Intell.* **19**, 246–247.

Kanatani, K. & Maruyama, T. 1996 Optimal grid pattern for focal length calibration. *Advanced Robotics* **10**, 81–103.

Kanatani K. & Takeda, S. 1995 3-D motion analysis of a planar surface by renormalization. *IEICE Trans. Inf. Sys.* E **78-D**, 1074–1079.

Kanazawa, Y. & Kanatani, K. 1995*a* Direct reconstruction of planar surfaces by stereo vision. *IEICE Trans. Inf. Sys.* E **78-D**-7, 917–922.

Kanazawa, Y. & Kanatani, K. 1995*b* Reliability of 3-D reconstruction by stereo vision. *IEICE Trans. Inf. Sys.* E **78-D**, 1301–1306.

Kanazawa, Y. & Kanatani, K. 1995*c* Reliability of fitting a plane to range data. *IEICE Trans. Inf. Sys.* E **78-D**, 1630–1635.

Kanazawa, Y. & Kanatani, K. 1996*a* Optimal line fitting and reliability evaluation. *IEICE Trans. Inf. Sys.* E **79-D**, 1317–1322.

Kanazawa, Y. & Kanatani, K. 1996*b* Optimal conic fitting and reliability evaluation. *IEICE Trans. Inf. Sys.* E **79-D**, 1323–1328.

Kanazawa, Y. & Kanatani, K. 1997 Infinity and planarity test for stereo vision. *IEICE Trans. Inf. Sys.* E **80-D**.

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

Maybank, S. 1993 *Theory of reconstruction from image motion*. Berlin: Springer.

Taubin, G. 1991 Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Trans. Patt. Analysis Mach. Intell.* **13**, 1115–1138.

Taubin, G., Cukierman, F., Sullivan, S., Ponce, J. & Kriegman, D. J. 1994 Parametrized families of polynomials for bounded algebraic curve and surface fitting. *IEEE Trans. Patt. Analysis Mach. Intell.* **16**, 287–303.

Triono, I. & Kanatani, K. 1996 Automatic recognition of regular figures by geometric AIC. In *Proc. IAPR Workshop Machine Vision Applications, Tokyo, Japan, 12–14 November 1996*, pp. 393–396.

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

Weng, J., Huang, T. S. & Ahuja, N. 1993*a* *Motion and structure from image sequences*. Berlin: Springer.

Weng, J., Ahuja, N. & Huang, T. S. 1993*b* Optimal motion and structure estimation. *IEEE Trans. Patt. Analysis Mach. Intell.* **15**, 864–884.

Discussion

R. Cipolla (*University of Cambridge, UK*). Can Professor Kanatani's method handle points on the critical surface?

K. Kanatani. Yes, it can. My theory deals with data with noise, and in the presence of noise data points are almost always not on a critical surface. Doing 3D reconstruction from points on a critical surface is like fitting a planar surface to collinear points in three dimensions. In the presence of noise, there always exists a solution, but it is very unreliable. And the degree of unreliability can be evaluated by my geometric AIC, which compares the goodness measures of two models: general model versus degeneracy in the data configuration.

R. I. Hartley (*GE Corporate Research and Development, Niskayuna, NY, USA*). Why choose to use the Cramer–Rao lower bound rather than some other bounds known in statistics?

K. Kanatani. The problem of geometric fitting is in many respects very different from traditional statistical estimation problems, for which various lower bounds are known. For geometric fitting, no such bounds have been known, so I derived one by

modifying the Cramer–Rao lower bound to fit in my geometric fitting framework. This bound is attained by maximum likelihood solution in the first order. Of course, there could be other bounds better in higher orders, but according to my experiments my bound is very tight, so I don't think we should worry about higher-order differences.

W. Triggs (*INRIA, France*). In Professor Kanatani's estimation of the fundamental matrix did he normalize the data first? Also, is a full statistical error the best thing to minimize?

K. Kanatani. First, about data normalization. Since my statistical theory deals with noise in the *physical* space, data can be represented in any way you like; no matter how they are represented, the solution always has the same meaning, as it should. In my treatment, I represent a point in the image by a three-dimensional vector starting from the focal point towards that point in the image plane, and I normalize it so that its z-component is 1. Of course, you can use other normalizations, but you must give its covariance matrix/tensor so that it is compatible with your normalization; if you change the normalization, you also have to change its covariance matrix/tensor accordingly, thereby guaranteeing the solution to be representation-free.

Next, about the minimizing criterion. According to my experiments, minimizing full statistical errors definitely improves accuracy over minimizing algebraic distances. Richard Hartley asserted in his talk that minimizing algebraic distances was advantageous because it was computationally efficient, yet the difference in accuracy from minimizing full statistical errors was very small. But we must be careful when we talk about 'accuracy'. Take the fundamental matrix F for example. If we measure $\|F - \bar{F}\|$ ($\bar{F}$ is the true value of F), the difference between statistical error minimization and algebraic distance minimization is very small indeed. But if we compute the *3D motion* from F, the difference is very large, as I illustrated in my talk. The same thing happens for conic fitting, too: although the difference $\|Q - \bar{Q}\|$ is very small, the curve represented by the matrix Q can deviate significantly from the curve represented by the matrix $\bar{Q}$. So, we cannot be content with the fact that errors *in matrix representation* are very small. Very slight as the improvement may be, we should minimize full statistical errors; the difference in the end result is often very striking.

Richard Hartley (this volume) also asserts that minimizing full statistical error was time-consuming when constraints were involved, as in the case of the fundamental matrix: one has to introduce minimal parametrization so as to satisfy the constraints and then do numerical search in the (minimal) parameter space. This is time-consuming indeed. However, my proposed approach is different. We first do minimization without regard to the constraints, which is usually very efficient, and compute an initial solution *and its covariance matrix/tensor*. Then, we correct the solution optimally *according to that covariance matrix/tensor*. This correction stage can be done by repeated *linear* computation. It can be confirmed by many experiments that the resulting accuracy is comparable to the parametrized minimization, yet the computation is several orders faster. Thus, there is no reason to avoid full statistical error minimization.

A. Fitzgibbon (*Department of Engineering, University of Oxford, UK*). Does Professor Kanatani's method reduce to minimizing the exact geometric distance in

the case of small isotropic noise? Also, is it a valid assumption to assume that noise is small?

K. Kanatani. Yes, what I call 'maximum likelihood estimation' reduces to minimizing the sum of squares of distances over which the data points must be displaced so as to satisfy the constraint, *provided* the probability density of the noise is a function of the square distance, which is the case if the noise is subject to an isotropic and identical Gaussian distribution. And I introduced various first-order approximations for simplifying the equations, assuming that noise is small. According to my experiments, this assumption is very valid as long as we are dealing with image and sensor data, as opposed to data in medicine, biology, agriculture, sociology, economics, etc., which traditional statistics concerns.

R. I. Hartley. I would like to comment on the previous question. From my own experiments I have found that observed residual error seems to be quite close to linear in the injected noise. This suggests that surface representing correct measurements is well approximated locally by its tangent plane. This suggests that the small-noise assumption is justified.

O. Faugeras (*INRIA, France*). There have been some recent advances in estimating quantities such as R under fairly severe noise. Does Professor Kanatani not think we need a theory for describing random processes on manifolds under all noise conditions?

K. Kanatani. I do not see strong evidence yet that we need higher-order theories. Although my algorithms have been derived by introducing first-order approximations, the resulting algorithms can be applied to data with very large noise, and the result is usually fairly good. In order to obtain a better method, we need to exploit the particular mathematical structure and noise characteristics of the problem in question, introducing additional parameters that describe higher-order effects. This means that the analysis is inevitably less general. Such problem-dependent higher-order theories may have advantages over general first-order theories in some situations.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES