Systems and Computers in Japan, Vol. 35, No. 6, 2004
Translated from Denshi Joho Tsushin Gakkai Ronbunshi, Vol. J86-D-II, No. 7, July 2003, pp. 966–973

For Geometric Inference from Images, What Kind of Statistical Model Is Necessary?

Kenichi Kanatani

Department of Information Technology, Okayama University, Okayama, 700-8530 Japan

SUMMARY

In order to promote mutual understanding with researchers in other fields including statistics, this paper investigates the meaning of *statistical methods* for geometric inference based on image feature points. We trace back the origin of feature uncertainty to image processing operations and discuss the meaning of *geometric fitting*, *geometric model selection*, the *geometric AIC*, and the *geometric MDL*. Then, we discuss the implications of asymptotic analysis in reference to *nuisance parameters*, the *Neyman-Scott problem*, and *semiparametric models* and point out that application of statistical methods requires careful considerations about the peculiar nature of geometric inference. © 2004 Wiley Periodicals, Inc. Syst Comp Jpn, 35(6): 1–9, 2004; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.10635

Key words: statistical method; feature point extraction; asymptotic evaluation; Neyman-Scott problem; semiparametric model.

1. Introduction

Statistical inference is one of the key components of computer vision research. Today, sophisticated statistical theories, once regarded as impractical, are applied in practice, taking advantage of the enormous computational power of today's advanced hardware [27]. On the other hand, the author has used statistical methods for inferring inherent relations in geometric data such as points and lines extracted from images, deriving optimization techniques for maximizing the accuracy and proving theoretical accuracy bounds for them [8, 9].

However, the same term *statistical* has different meanings in different contexts. This difference has often been overlooked, frequently causing controversies not only among image processing researchers but also over different research areas including statistics.

Geometric inference from images is one of the few research areas in which Japan has taken the lead. Although it is dominated by overseas research today, there is still a strong interest in this problem in Japan [14, 20]. This paper is not intended to present a new statistical method; rather, we focus on the very question of why we need statistical methods at all, with the motivation of encouraging a wider range of collaborations among researchers of different areas, thereby promoting further theoretical developments of this problem.

2. What Is a Statistical Method?

Most problems in mathematics and physics are deterministic; various properties are deduced from axioms and fundamental equations. A similar approach exists for computer vision research, too; for example, an elaborate theory has been proposed for reconstructing 3D shapes from im-

Contract grant sponsors: Supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C(2) (No. 15500113), the Support Center for Advanced Telecommunications Technology Research, and Kayamori Foundation of Informational Science Advancement.

ages by analyzing the camera imaging geometry [7]. In contrast, the goal of statistical methods is not to describe the properties of observed data themselves but to infer the properties of the *ensemble* by regarding the observed data as *random samples from it*. The ensemble may be a collection of existing entities (e.g., the entire population), but often it is a hypothetical set of conceivable possibilities.

When a statistical method is employed, the underlying ensemble is usually taken for granted. For character recognition, for instance, it is understood that we are thinking of an ensemble of all prints and scripts of individual characters. Since some characters are more likely to appear than others, a probability distribution is naturally defined over the ensemble.

For handwritten character recognition, our attention is restricted to the set of all handwritten characters. The ensemble is further restricted if we want to recognize characters written by a specific writer (e.g., his/her signatures), but these restrictions are too obvious to be mentioned. However, this issue is very crucial for geometric inference from images, yet this fact has not been well recognized in the past. To show this is the main purpose of this paper.

3. What Is Geometric Inference?

3.1. Ensembles for geometric inference

What we call *geometric inference* in this paper deals with a *single* image (or a single *set* of images). For example, we observe an image of a building and extract *feature points* such as the vertices of the building and the corners of windows. We test if they are collinear, and if so, we fit a line to them and evaluate the uncertainty of that fit. This is the simplest example of geometric inference.

The reason why we use a statistical method is that the extracted feature positions have uncertainty. For example, we can fit a line to approximately collinear points in the most reliable manner by considering the uncertainty of individual feature points. We can also infer that those points with large deviations from the fitted line are not collinear with other points in the first place. What is the ensemble that underlies this type of inference?

This question reduces to the question of why the uncertainty of the feature points occurs at all. After all, statistical methods are not necessary if the data are exact. Using a statistical method means regarding the current positions as random samples from its possible positions. But what are the *possible positions*?

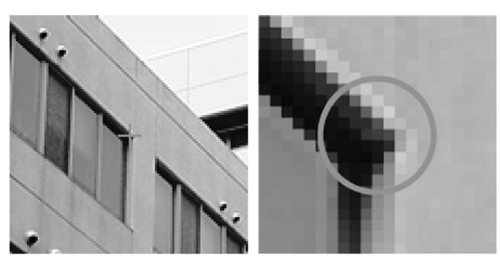
3.2. Uncertainty of feature extraction

Numerous algorithms have been proposed in the past for extracting feature points including the Harris operator [6] and SUSAN [24], and their performance has been extensively compared [3, 21, 23]. For tracking feature points through a video stream, the best known is the Kanade-Lucas-Tomasi algorithm [26]. However, if we use, for example, the Harris operator to extract a particular corner of a particular window from a building image, the output is unique (Fig. 1). No matter how many times we repeat the extraction, we obtain the same point because no external disturbances exist and the internal parameters (e.g., thresholds for judgment) are unchanged. It follows that the current position is the sole possibility. How can we find it elsewhere? In the past, no satisfactory answer seems to have been given to this question.

The reason why we think that *other possibilities should exist* is the fact that *the extracted position is not necessarily correct.* But if it is not correct, why did we extract it? Why didn't we extract the correct position in the first place? The answer is: *we cannot*. Why is this impossible?

3.3. Image processing for computer vision

The reason why there exist so many feature extraction algorithms, none of them being definitive, is that they are aiming at an intrinsically impossible task. If we were to extract a point around which, say, "the intensity varies to the largest degree measured in such and such a criterion," the algorithm would be unique (variations may exist in intermediate steps, but the final output should be the same). For computer vision applications, however, what we want is not *image properties* but *3D properties* such as corners of a building, but the way a 3D property is translated into an image property is intrinsically heuristic. As a result, as



(a) (b)

Fig. 1. (a) A feature point in an image of a building. (b) Its enlargement and the uncertainty of the feature location.

many algorithms can exist as the number of its 2D interpretations. In this sense, feature extraction is essentially heuristic. This fact has not been given much attention in the past, and feature extraction has often been regarded as an objective image processing task.

If we specify a 3D feature that we want to extract, its appearance in the image is not unique. It is affected by various properties of the scene including the details of its 3D shape, the viewing orientation, the illumination condition, and the light reflectance properties of the material. A slight difference of any of them can result in a substantial difference on the image plane. Theoretically, exact feature extraction would be possible if all the properties of the scene were exactly known, but to infer them from images is the very task of computer vision. It follows that we must make a guess in the image processing stage, and there exist as many algorithms as the number of guesses. For the current image, some guesses may be correct, but others may be wrong.

This observation implies that the *possible feature positions* should be associated with *the set of hypothetical algorithms*, but this interpretation has not been considered in the past. According to this interpretation, the current position is regarded as produced by an algorithm sampled from it. This explains why one always obtains the same position no matter how many times one repeats extraction using that algorithm. To obtain a different position, one has to sample another algorithm.

4. Statistical Model of Feature Location

4.1. Covariance matrix of a feature point

For doing statistical analysis based on the above interpretation, we need to hypothesize that the *mean* of the potential positions coincides with the true position. In other words, all hypothetical algorithms as a whole are assumed to be *unbiased*. Without this hypothesis, efforts to devise good algorithms would be meaningless.

The performance of feature point extraction depends on the image properties around that point. If, for example, we want to extract a point in a region with an almost homogeneous intensity, the resulting position may be ambiguous whatever algorithm is used. In other words, the positions that the hypothetical algorithms would extract should have a large spread around the true position. If, on the other hand, the intensity greatly varies around that point, any algorithm could easily locate it accurately, meaning that the positions that the hypothetical algorithms would extract should have a strong peak at the true position. This observation suggests that we may introduce for each feature point its *covariance matrix* that predicts the spread of its potential positions.

Let $V[p_\alpha]$ be the covariance matrix of the α-th feature point p_α. The above argument implies that we can determine the qualitative characteristics of uncertainty in relative terms but not its absolute magnitude. If, for example, the intensity variations around p_α are almost the same in all directions, we can think of the probability distribution as isotropic, a typical equiprobability line, known as the *uncertainty ellipses*, being a circle [Fig. 1(b)]. If, on the other hand, p_α is on an object boundary, distinguishing it from nearby points should be difficult whatever algorithm is used, so its covariance matrix should have an elongated uncertainty ellipse along that boundary.

From these observations, we write the covariance matrix $V[p_\alpha]$ in the form

$$V[p_\alpha] = \epsilon^2 V_0[p_\alpha] \tag{1}$$

where ϵ is an unknown magnitude of uncertainty, which we call the *noise level* [8, 9]. The matrix $V_0[p_\alpha]$, which we call the *normalized covariance matrix*, describes the relative magnitude and the dependence on orientations [8, 9]. However, if we call Eq. (1) the *covariance matrix of feature point* p_α, it is not a property of point p_α but a *property of hypothetical algorithms* applied to the neighborhood of p_α. This fact has not been clearly recognized in the past.

4.2. Characteristics of feature extraction algorithms

Most of the existing feature extraction algorithms are designed to output those points that have large image variations around them [3, 6, 21, 23, 24]. As a result, the covariance matrix of a feature point extracted by such an algorithm can be regarded as nearly isotropic. This has also been confirmed by experiments [13].

The intensity variations around different feature points are usually unrelated, so their uncertainty can be regarded as statistically independent. However, if we track feature points over consecutive video frames, it has been observed that the uncertainty of each point has strong correlations over the frames [25].

Some interactive applications require humans to extract feature points by manipulating a mouse. It has been shown by experiments that humans are likely to choose *easy-to-see* points, such as isolated points and intersections, around which the intensity varies almost in the same degree in all directions [13]. In this sense, the statistical characteristics of human extraction are very similar to machine extraction. This is no surprise if we recall that image processing for computer vision is essentially a heuristic that simulates human perception. It has also been reported that strong microscopic correlations exist when humans manually select corresponding feature points over multiple images [17].

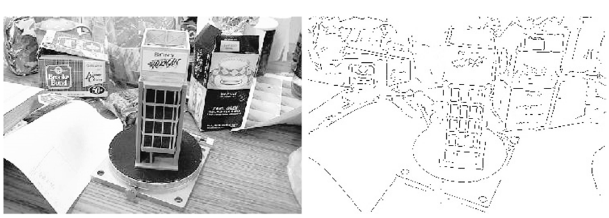
4.3. Image quality and uncertainty

We have observed that the ensemble behind geometric inference is the *set of algorithms* and that statistical assumptions such as normality, independence, unbiasedness, and correlations are *properties of the underlying set of algorithms*. This is the central message of this paper. In the past, however, a lot of confusion occurred because these were taken to be *properties of the image*. The main cause of this confusion may be the tradition that the uncertainty of feature points has simply been referred to as *image noise*. In particular, Eq. (1), which describes the *uncertainty of feature locations*, has been called the *statistical model of image noise*.

Of course, we may obtain better results if we use higher-quality images whatever algorithm is used. The performance of any algorithm naturally depends on the image quality. However, the task of computer vision is not to analyze image properties but to study the 3D properties of the objects that we are viewing.

This observation also applies to *edge detection*, whose goal is to find the boundaries of 3D we are viewing. In reality, all existing algorithms seek *edges*, that is, lines and curves across which the intensity changes discontinuously (Fig. 2). So, it can be argued that edge detection is also a heuristic and hence no definitive algorithm will ever be found.

Thus, we conclude that as long as a hiatus exists between *what we want to extract* and *what we are computing*, any process of computer vision accompanies uncertainty independent of the image quality, and the result must be interpreted *statistically*.The underlying ensemble is the set of hypothetical (inherently imperfect) image processing algorithms, which should be distinguished from image noise caused by random intensity fluctuations of individual pixels. Yet, it has been customary to evaluate the performance of image processing algorithms for extracting 3D properties by adding independent Gaussian noise to individual pixels.



(a) (b)

Fig. 2. (a) An indoor scene. (b) Detected edges.

5. What Is Asymptotic Analysis?

As stated earlier, *statistical estimation* refers to estimating the properties of an ensemble from a finite number of samples chosen from it, assuming some knowledge, or a *model*, about the ensemble. If the uncertainty originates from external conditions, the estimation accuracy can be increased by controlling the measurement environments. For internal uncertainty, on the other hand, there is no way of increasing the accuracy but by repeating the experiment and doing statistical inference. However, doing experiments usually entails costs, and in practice the number of experiments is often limited.

Taking account of such practical considerations, statisticians usually evaluate the performance of estimation *asymptotically*, analyzing the growth in accuracy as the number n of experiments increases. This is justified because a method whose accuracy increases more rapidly as $n \to \infty$ than others can reach admissible accuracy with a fewer number of experiments [Fig. 3(a)].

In contrast, the ensemble for geometric inference based on feature points is, as we have seen, the set of potential feature positions that could be located if other (hypothetical) algorithms were used. The goal is to estimate

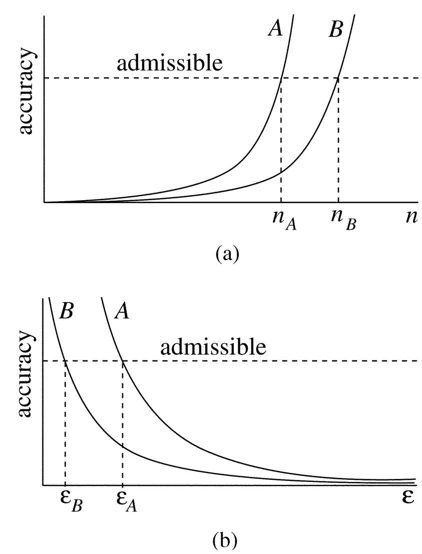


Fig. 3. (a) For the standard statistical estimation problem, it is desired that the accuracy increases rapidly as the number of experiments $n \to \infty$, because admissible accuracy can be reached with a smaller number of experiments. (b) For geometric inference, it is desired that the accuracy increases rapidly as the noise level $\varepsilon \to 0$, because larger data uncertainty can be tolerated for admissible accuracy.

geometric quantities as closely as possible to their expectations over that ensemble, which we assume are their true values. In other words, we want to minimize the discrepancy between obtained estimates and their true values on average over all hypothetical algorithms.

However, the crucial fact is, as stated earlier, we can choose only one sample ($n = 1$) from the ensemble as long as we use a particular image processing algorithm. In other words, the number n of experiments is 1. Then, how can we evaluate the performance of statistical estimation?

Evidently, we want a method whose accuracy is sufficiently high even for large data uncertainty. This implies that we need to analyze the growth in accuracy as the noise level ε decreases, because a method whose accuracy increases more rapidly as $\varepsilon \to 0$ than others can tolerate larger data uncertainty for admissible accuracy [Fig. 3(b)].

6. Asymptotic Analysis for Geometric Inference

We now illustrate our assertion in more specific terms.

6.1. Geometric fitting

Let $p_1, \ldots, p_N$ be extracted feature points. We assume that their true positions $\bar{p}_1, \ldots, \bar{p}_N$ satisfy a constraint

$$\boldsymbol{F}(\bar{p}_\alpha, \boldsymbol{u}) = \boldsymbol{0}, \quad \alpha = 1, ..., N \tag{2}$$

parametrized by a p-dimensional vector $\boldsymbol{u}$. Our task, which we call *geometric fitting*, is to estimate the parameter $\boldsymbol{u}$ from the observed positions $p_1, \ldots, p_N$. Equation (2) is called the (*geometric*) *model* [9].

A typical problem is to fit a line or a curve (e.g., a circle or an ellipse) to given N points in the image, but this can be straightforwardly extended to multiple images. Namely, if a point (x_α, y_α) in one image corresponds to a point (x_α', y_α') in another, we can regard them as a single point p_α in a 4D joint space with coordinates $(x_\alpha, y_\alpha, x_\alpha', y_\alpha')$ (Fig. 4). If the camera imaging geometry is modeled as perspective projection, the constraint (2) corresponds to the *epipolar equation*; the parameter $\boldsymbol{u}$ is the *fundamental matrix* [16]. If the scene is a planar surface or located very far away, Eq. (2) can be regarded as imposing a (two-dimensional) *homography* (or *projective transformation*) on the two images; the parameter $\boldsymbol{u}$ is the *homography matrix* [12].

If we write the covariance matrix of p_α in the form of Eq. (1) and regard the distribution of uncertainty as Gaussian, *maximum likelihood estimation* (*MLE*) over the potential positions of the N feature points is to minimize the squared *Mahalanobis distance* with respect to the normalized covariance matrices $V_0[p_\alpha]$:

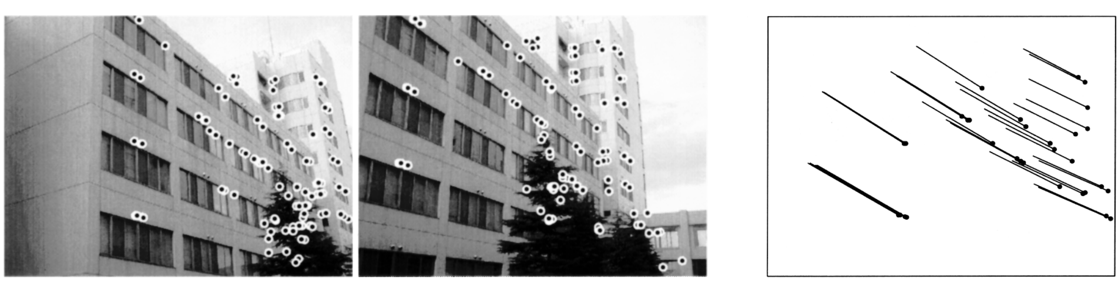
$$J = \sum_{\alpha=1}^{N} (p_\alpha - \bar{p}_\alpha, V_0[p_\alpha]^{-1}(p_\alpha - \bar{p}_\alpha)) \tag{3}$$

Here, p_α and p_α' are identified as two-dimensional vectors and $(\cdot, \cdot)$ designates the inner product of vectors. Equation (3) is minimized with respect to $\{\bar{p}_\alpha\}$, $\alpha = 1, \ldots, N$, and $\boldsymbol{u}$ subject to the constraint (2).

Assuming that the noise level ε is small and using Taylor expansion with respect to ε, one can show that the covariance matrix $V[\hat{\boldsymbol{u}}]$ of the MLE solution $\hat{\boldsymbol{u}}$ converges to $\boldsymbol{O}$ as $\varepsilon \to 0$ (*consistency*) and that $V[\hat{\boldsymbol{u}}]$ coincides with a theoretical accuracy bound if terms of $O(\varepsilon^4)$ are ignored (*asymptotic efficiency*) [9]. Thus, MLE achieves admissible accuracy in the presence of larger uncertainty than other methods.

6.2. Geometric model selection

Geometric fitting is to estimate the parameter $\boldsymbol{u}$ of a given model in the form of Eq. (2). If we have multiple candidate models $\boldsymbol{F}_1(\bar{p}_\alpha, \boldsymbol{u}_1) = \boldsymbol{0}$, $\boldsymbol{F}_2(\bar{p}_\alpha, \boldsymbol{u}_2) = \boldsymbol{0}, \ldots$, from



(a) (b)

Fig. 4. (a) Two images of a building and extracted feature points. (b) *Optical flow* consisting of segments connecting corresponding feature points (black dots correspond to the positions in the left image). The two endpoints can be identified with a point in a four-dimensional space.

which we are to select an appropriate one, the problem is (*geometric*) *model selection* [9].

A naive idea is to first estimate the parameter $\boldsymbol{u}$ by MLE and compute the *residual* (*sum of squares*), that is, the minimum value $\hat{J}$ of J in Eq. (3), for each model and then select the one that has the smallest $\hat{J}$. However, this is not a fair comparison of different constraints, because the MLE solution $\hat{\boldsymbol{u}}$ is determined for each constraint by minimizing the residual $\hat{J}$ of that constraint. This observation leads to the idea of compensating for the bias caused by substituting the MLE solution. This is the principle of Akaike's *AIC* (*Akaike Information Criterion*) [1], whose theoretical basis is the *Kullback-Leibler information* (or *divergence*). Doing a similar analysis to Akaike's and examining the asymptotic behavior as the noise level ε goes to zero, one can obtain the following *geometric AIC* [10]:

$$\text{G-AIC} = \hat{J} + 2(Nd + p)\epsilon^2 + O(\epsilon^4) \tag{4}$$

Here, d is the dimension of the manifold defined by the constraint (2). Its existence in the right-hand side is the major difference, in appearance, from Akaike's AIC, reflecting the uncertainty of N feature positions.

Another well-known criterion is Rissanen's *MDL* (*Minimum Description Length*) [22], which measures the goodness of a model by the minimum information theoretic code length of the data and the model. If, following Rissanen, one quantizes the real-valued parameters, determines the quantization width in such a way that the total code length becomes smallest, and analyzes its asymptotic behavior as the noise level ε goes to zero, one obtains the following *geometric MDL* [10]:

$$\text{G-MDL} = \hat{J} - (Nd + p)\epsilon^2 \log\left(\frac{\epsilon}{L}\right)^2 + O(\epsilon^2) \tag{5}$$

Here, L is a reference length chosen so that its ratio to the magnitude of data is $O(1)$ (e.g., L can be taken to be the image size for feature point data). Its exact determination requires an a priori distribution that specifies where the data are likely to appear, but it has been observed that the model selection is not very much affected by L as long as it has the same order of magnitude [10].

6.3. Equivalent statistical interpretation

Although the term *asymptotic* has opposite meanings, the results in the preceding two sections bear a strong resemblance to the standard statistical analysis. It is known that the covariance matrix of an MLE estimator for the standard statistical problem converges, under a standard condition, to $\boldsymbol{O}$ as the number n of experiments goes to infinity (*consistency*) and that it agrees with the *Cramer-Rao lower bound* for $O(1/n^2)$ (*asymptotic efficiency*). It follows that $1/\sqrt{n}$ plays the same role as ε for geometric inference.

The same correspondence exists for model selection, too. Note that the unknowns are the p parameters of the constraint plus the N true positions specified by the d coordinates of the d-dimensional manifold defined by the constraint. If Eq. (4) is divided by ε^2, we have $\hat{J}/\varepsilon^2 + 2(Nd + p) + O(\varepsilon^2)$, which is (–2 times the logarithmic likelihood) + 2(the number of unknowns), the same form as Akaike's AIC. The same hold for Eq. (5), which reduces to Rissanen's MDL if ε is replaced by $1/\sqrt{n}$.

This correspondence can be interpreted as follows. Since the underlying ensemble is hypothetical, we can actually observe only one sample from it as long as a particular algorithm is used. Suppose we hypothetically apply n different algorithms to find n different positions. The optimal estimate of the true position under the Gaussian model is their sample mean. The covariance matrix of the sample mean is $1/n$ times that of the individual samples. Hence, this hypothetical estimation is equivalent to dividing the noise level ε in Eq. (1) by $\sqrt{n}$.

In fact, there were attempts to generate a hypothetical ensemble of algorithms by randomly varying the internal parameters (e.g., the thresholds for judgments) and sample different points [4], not adding random noise to the image. Then, one can compute their means and covariance matrix. Such a process as a whole can be regarded as one operation that effectively achieves higher accuracy.

Thus, our conclusion is: the asymptotic analysis for $\varepsilon \to 0$ is equivalent to the asymptotic analysis for $n \to \infty$, where n is the number of hypothetical observations. As a result, the expression $\cdots + O(1/\sqrt{n^k})$ in the standard statistical estimation problem turns into $\cdots + O(\varepsilon^k)$ in geometric inference. This type of analysis has already been done by the author for a long time, but he has not made clear what the underlying ensemble was.

7. Nuisance Parameters and Semiparametric Model

7.1. Asymptotic parameters

The number n that appears in the standard statistical analysis is the *number of experiments*. It is also called the *number of trials*, the *number of observations*, and the *number of samples*. Evidently, the properties of the ensemble are revealed more precisely as more elements are sampled from it.

However, the number n is often called the *number of data*, which has caused considerable confusion. For example, if we observe 100-dimensional vector data in one experiment, one may think that the number of data is 100, but this is wrong: the number n of experiments is 1. We are observing one sample from an ensemble of 100-dimensional vectors.

For character recognition, the underlying ensemble is the set of possible character images, and the learning process concerns the number n of training steps necessary to establish satisfactory responses. This is independent of the dimension N of the vector that represents each character. The learning performance is evaluated asymptotically as $n \to \infty$, not $N \to \infty$.

The situation is the same for geometric inference, too. If we extract, for example, 50 feature points, they constitute a 100-dimensional vector consisting of their x and y coordinates. If no other information, such as the image intensity, is used, the image is completely characterized by that vector. Applying a statistical method means regarding it as a sample from a hypothetical ensemble of 100-dimensional vectors.

7.2. Neyman-Scott problem

In the past, many computer vision researchers have analyzed the asymptotic behavior as $N \to \infty$ without explicitly mentioning what the underlying ensemble is. This is perhaps motivated by a similar formulation in the statistical literature. Suppose, for example, a rod-like structure lies on the ground in the distance. We emit a laser beam toward it and estimate its position and orientation by observing the reflection of the beam, which is contaminated by noise. We assume that the laser beam can be emitted in any orientation any number of times but the emission orientation is measured with noise. The task is to estimate the position and orientation of the structure as accurately as possible by emitting as small a number of beams as possible. Naturally, the estimation performance should be evaluated in the asymptotic limit $n \to \infty$ with respect to the number n of emissions.

The underlying ensemble is the set of all response times for all possible directions of emission. Usually, we are interested in the position and orientation of the structure but not the exact orientation of each emission, so the variables for the former are called the *structural parameters*, which are fixed in number, while the latter are called the *nuisance parameters*, which increase indefinitely as the number n of experiments increases [18]. Such a formulation is called the *Neyman-Scott problem* [19]. Since the constraint is an implicit function in the form of Eq. (2), we are considering an *errors-in-variables model* [5]. If we linearize the constraint by changing variables, the noise characteristics differ for each data component, so the problem is *heteroscedastic* [15].

To solve this problem, one can introduce a parametric model for the distribution of possible laser emission orientations, regarding the actual emissions as random samples from it. This formulation is called a *semiparametric model* [2]. An optimal solution can be obtained by finding a good *estimating function* [2, 20].

7.3. Semiparametric model for geometric inference

Since the semiparametric model has something different from the geometric inference problem described so far, a detailed analysis is required for examining if application of a semiparametric model to geometric inference will yield a desirable result [14, 20]. In any event, one should explicitly state what kind of ensemble (or ensemble of ensembles) is assumed before doing statistical analysis. This is the main message of this paper.

This is not merely a conceptual issue. It also affects the performance evaluation of simulation experiments. In doing a simulation, one can freely change the number N of feature points and the noise level ε. If the accuracy of Method A is higher than Method B for particular values of N and ε, one cannot conclude that Method A is superior to Method B, because opposite results may come out for other values of N and ε. Here, we have two alternatives for performance evaluation: fixing ε and varying N to see if admissible accuracy is attained for a smaller number of feature points [18]; fixing N and varying ε to see if larger data uncertainty can be tolerated for admissible accuracy [11]. These two types of evaluation have different meanings. Our conclusion is that the results of one type of evaluation cannot directly be compared with the results of the other.

8. Conclusions

Since image processing for computer vision is based on many peculiar assumptions that are not found in other research areas, the meaning of the term *statistical* is not necessarily the same as in other areas, often causing misunderstandings and controversies. The purpose of this paper is to clarify the background of such assumptions, with the motivation of encouraging a wider range of collaborations among researchers of different areas including statistics, thereby promoting further theoretical developments of computer vision research.

Tracing back the origin of feature uncertainty to image processing operations, we described the meaning of *geometric fitting*, *geometric model selection*, *geometric AIC*, and *geometric MDL*. We have also discussed the implications of asymptotic analysis for performance evaluation in reference to *nuisance parameters*, the *Neyman-*

Scott problem, and *semiparametric models*. We have pointed out that application of statistical methods requires careful considerations about the peculiar nature of the geometric inference problem.

Acknowledgments. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, under a Grant in Aid for Scientific Research C(2) (No. 15500113), the Support Center for Advanced Telecommunications Technology Research, and Kayamori Foundation of Informational Science Advancement.

REFERENCES

1. Akaike H. What is the information criterion AIC?—Its meaning and future prospects. Suri Kagaku 1976;153:5–11.
2. Amari S, Kawanabe M. Estimation of linear relations—Is the least squares optimal? Oyou Suri 1996;6;96–109. (in Japanese)
3. Chabat F, Yang GZ, Hansell DM. A corner orientation detector. Image Vision Comput 1999;10:761–769.
4. Cho K, Meer P, Cabrera J. Performance assessment through bootstrap. IEEE Trans Pattern Anal Mach Intell 1997;19:1185–1198.
5. Fuller WA. Measurement error models. Wiley; 1987.
6. Harris C, Stephens M. A combined corner and edge detector. Proc 4th Alvey Vision Conf, 1988, Manchester, UK, p 147–151.
7. Hartley R, Zisserman A. Multiple view geometry in computer vision. Cambridge University Press; 2000.
8. Kanatani K. Kukan Deta no Suri. Asakura Publishing; 1995. (in Japanese)
9. Kanatani K. Statistical optimization for geometric computation: Theory and practice. Elsevier Science; 1996.
10. Kanatani K. Model selection for geometric fitting. Trans IEICE 2001;J84-A:1385–1393. (in Japanese)
11. Kanatani K, Matsunaga C. Comparing degeneracy detection performance of the geometric AIC and the geometric MDL. Trans IEICE 2002;J85-D-II;1497–1499. (in Japanese)
12. Kanatani K, Ohta N, Kanazawa Y. Optimal homography computation with a reliability measure. Trans IEICE 2000;E83-D:1369–1374.
13. Kanazawa Y, Kanatani K. Do we really have to consider covariance measures for image feature points? Trans IEICE 2002;J85-A:231–239. (in Japanese) Electron Commun Japan Part 3 2003;86:1–10. (English translation)
14. Kurihara Y, Ohta N. Shape reconstruction from optical flow without nuisance parameter estimation. IPSJ SIG Notes, 2002-CVIM-135-14, p 87–94, 2002. (in Japanese)
15. Leedan Y, Meer P. Heteroscedastic regression in computer vision: Problems with bilinear constraint. Int J Comput Vision 2000;37:127–150.
16. Mishima H, Kanatani K. Optimal computation of the fundamental matrix and its reliability evaluation. IPSJ SIG Notes, 99-CVIM-118-10, p 67–74, 1999. (in Japanese)
17. Morris DD, Kanatani K, Kanade T. Gauge fixing for accurate 3D estimation. Proc IEEE Conf Comput Vision Pattern Recog Kauai, Hawaii, 2001, Vol. 2, p 343–350.
18. Nagao J, Han T. Construction of MDL criteria in the presence of nuisance parameters and its application to geometric model selection problems. Trans IEICE 2000;J83-A:83–95. (in Japanese)
19. Neyman J, Scott EL. Consistent estimates based on partially consistent observations. Econometrica 1948;16:1–32.
20. Okatani T, Deguchi K. On bound of estimation accuracy of the structure and motion problem. Proc 6th Meeting Image Recog Understand, 2002, Nagano, p 335–340. (in Japanese)
21. Reisfeld D, Wolfson H, Yeshurun Y. Context-free attentional operators: The generalized symmetry transform. Int J Comput Vision 1995;2:119–130.
22. Rissanen J. Stochastic complexity in statistical inquiry. World Scientific; 1989.
23. Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors. Int J Comput Vision 2000;37:151–172.
24. Smith SM, Brady JM. SUSAN—A new approach to low level image processing. Int J Comput Vision 1997;23:45–78.
25. Sugaya Y, Kanatani K. Outlier removal for motion tracking by subspace separation. IJSJ SIG Notes, 2002-CVIM-133-24, p 177–184. (in Japanese)
26. Tomasi C, Kanade T. Detection and tracking of point features. CMU Tech Rep, CMU-CS-91-132, 1991;http://vision.stanford.edu/~birch/klt
27. Ueda N. Bayesian learning [I]–[IV]. J IEICE 2002;85:265–271, 421–426, 504–509, 633–638.

AUTHOR

Kenichi Kanatani received his M.S. and Ph.D. degrees in applied mathematics from the University of Tokyo in 1974 and 1979. After serving as a professor of computer science at Gunma University, he is currently a professor of information technology at Okayama University. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), *Geometric Computation for Machine Vision* (Oxford, 1993), and *Statistical Optimization for Geometric Computation: Theory and Practice* (Elsevier, 1996). He is an IEEE Fellow.