Proc. Workshop on Science of Computer Vision, September 2002, Okayama Japan, pp. 22–30.

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

Kenichi KANATANI*
Department of Information Technology, Okayama University
Okayama 700-8530 Japan

In order to facilitate smooth communications with researchers in other fields including statistics, this paper investigates the meaning of "statistical methods" for geometric inference based on image feature points. We point out that statistical analysis does not make sense unless the underlying "statistical ensemble" is clearly defined. We trace back the origin of feature uncertainty to image processing operations for computer vision in general and discuss the implications of asymptotic analysis for performance evaluation in reference to "geometric fitting", "geometric model selection", the "geometric AIC", and the "geometric MDL". Referring to such statistical concepts as "nuisance parameters", the "Neyman-Scott problem", and "semiparametric models", we point out that simulation experiments for performance evaluation will lose meaning without carefully considering the assumptions involved and intended applications.

1. Introduction

Statistical inference from images has been one of the key components of computer vision, and today's advanced computer power makes feasible many statistical methods once regarded as mere theoretical curiosities.

While statistical methods have usually been employed for recognition and classification purposes, the author has introduced statistical methods for *geometric inference* based on geometric primitives such as points and lines extracted by image processing operations and derived theoretical accuracy bounds and optimization techniques that achieve those bounds [7].

However, the term "statistical" has somewhat a different meaning for geometric inference than for recognition and classification purposes. This difference has often been overlooked, causing controversies over the validity of the statistical approach for geometric problems in general.

In this paper, we focus on this difference, starting with the question of why we need a statistical method at all. We point out that statistical analysis does not make sense unless the underlying "ensemble" is clearly defined. We trace back the origin of feature uncertainty to image processing operations for computer vision in general and discuss the implications of asymptotic analysis for performance evaluation. This is illustrated in reference to "geometric fitting" and "geometric model selection". Referring to such statistical concepts as "nuisance parameters", the "Neyman-Scott problem", and "semiparametric models", we point out that simulation experiments for performance evaluation will lose meaning without carefully considering the assumptions involved and intended applications.

*E-mail kanatani@suri.it.okayama-u.ac.jp

2. What Is a Statistical Method?

2.1 Statistical ensembles

Most problems of mathematics and physics are deterministic: various properties and propositions are deduced from a fixed set of axioms and fundamental equations. Such an approach can be found in computer vision research, too, a typical example being the geometric and algebraic theories of 3-D reconstruction from images based on an assumed camera imaging geometry [6].

Statistical methods, on the other hand, are not for studying the properties of observed data themselves but for inferring the properties of the *ensemble* from which we regard the observed data as having been sampled. 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 often taken for granted without being specifically mentioned what it is. For character recognition, for instance, it is understood that we are thinking of an ensemble of all prints and scripts of individual characters printed or written in all circumstances. Since some characters are more likely to appear than others, a *probability distribution* is defined over the ensemble.

The definition of the ensemble depends on the purpose of the analysis. For handwritten character recognition, for example, 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 the difference is too obvious to be mentioned. For geometric inference from images, however, this is a very crucial issue.

2.2 Characteristics of uncertainty

All elements of an ensemble share some specified characteristics, but otherwise their individual properties are different. Referring to this fact, we say that the elements have *uncertainty*. It can be classified into *external uncertainty* and *internal uncertainty*.

A. External uncertainty

This occurs in many experiments in physics. Even if an identical physical phenomenon is measured, uncertainty exists because of the imperfection of the measurement device (due to, e.g., uncontrollable impurities, thermal noise, and the use of approximate values of physical constants) as well as environmental disturbances (e.g., temperature, pressure, wind, radiant heat, light, and small oscillations of the device). As a result, the observed value, which should ideally be the same, fluctuates unpredictably from measurement to measurement. It follows that the underlying ensemble is the set of all values that could be observed in that experiment. This set can be identified with the set of *all possible observation processes*. Theoretically, this ensemble can be reduced by using a higher accuracy device and better controlling the environment; we would reach in the limit a set of a single element (the true value) with the delta function as the probability distribution.

B. Internal uncertainly

This uncertainty occurs because individual elements are inherently different even when the measurement is exact. The term "statistical method" often implies existence of this type of uncertainty. The ensemble for handwritten character recognition, for example, contains different characters because they are written by different writers and in different circumstances, while the character reading device is assumed to be accurate. Medicine and pharmacology deal with ensembles of patients affected with the same disease but otherwise with different characteristics (e.g., age, sex, occupation, medical history, physical strength, and health conditions), but the diagnosis is assumed to be correct. In some experiments in physics, observations by an ideal measuring device can fluctuate if the individual items are in different microscopic states (e.g., atomic decay, thermal fluctuations, and turbulence). The uncertainty of meteorological data, for instance, is intrinsic and independent of the uncertainty of the thermometers and pressure gauges used. As a result, this type of uncertainty cannot be reduced by controlling measurement devices or environments. It can be reduced only *statistically* by repeating measurements, resorting to the *law of large numbers*.

Then, what is the ensemble underlying geometric inference from images, and what kind of uncertainty is involved there?

3. What Is Geometric Inference?

3.1 Ensembles for geometric inference

Although images are used as input, the geometric inference problem studied by the author and others has a different characteristic from *recognition* problems using images. The ensemble for recognizing, say, persons is a set of images of different persons in different poses taken under different illumination conditions.

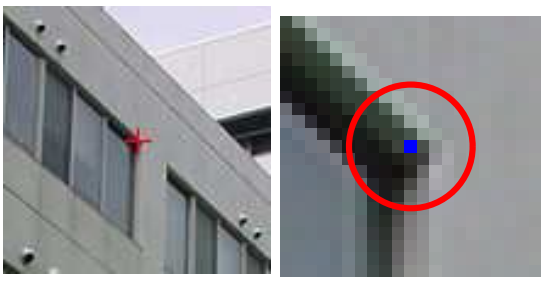
Geometric inference, on the other hand, deals with a *single* image (or a single set of images). For example, we observe an image of a building and extract *feature points* such as isolated points, corners, and intersections of lines. Our task is to test if a particular geometric constraint exists. If so, we estimate the parameters of the constraint and evaluate the degree of uncertainty of that estimation.

The reason why we need a statistical method is that *the extracted feature positions have uncertainty*. For example, we judge the extracted feature points as collinear if they are sufficiently aligned even though they are not strictly collinear. We also evaluate the degree of uncertainty of the fitted line by propagating the uncertainty of the individual points. 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 feature position as randomly sampled from a set of its *possible positions*. But *where else could it be if not in the current position*?

3.2 Uncertainty of feature extraction

Many algorithms have been proposed for extracting feature points including the Harris operator [5]



(a) (b)

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

and SUSAN [20], and their performance has been extensively compared [3, 16, 19]. Feature points can also be extracted and traced over a video sequence, for which the Kanade-Lucas-Tomasi method [22] is best known.

However, if we use, for example, the Harris operator to extract a particular corner of a particular 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 only possible position. How can we find it elsewhere?

If we closely examine the situation, we are compelled to conclude that other possibilities should exist for the feature position *because the extracted position is not necessarily correct*. But if the extracted position 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 to achieve an *essentially 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 the intermediate steps, but the final output should be the same).

However, what we want is not "image properties" but "3-D properties" such as corners of a building, and the way a 3-D property is translated into an image property is essentially a heuristic. Hence, as many algorithms can exist for extracting a 3-D property as the number of heuristics for its 2-D interpretation.

If we specify a 3-D 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 3-D shape, the viewing orientation, the illumination condition, and the light reflectance properties of the material, and a slight difference in the image capturing process may result in a different appearance of the image.

Theoretically, exact feature extraction would be possible if the properties of the scene were exactly known, but *to infer them from images is the very task of computer vision*. Thus, we must necessarily make a *guess* by a heuristic means in the image processing stage. For the current image, some guesses may be correct, but others may be wrong.

It follows that the exact feature position could be located only by a (non-existing) "ideal" algorithm that could guess everything correctly, but in reality a wrong position may be located because we use a non-ideal algorithm based on wrong guesses. This observation allows us to interpret the "possible feature positions" to be the positions that would be located by different (non-ideal) algorithms based on different guesses.

In this sense, the set of hypothetical positions can be associated with the set of hypothetical algorithms. The current position can be regarded as produced by an algorithm sampled from it. That is why one always obtains the same position no matter how many times one repeats extraction using that algorithm. In order to obtain a different position, one has to sample another algorithm from that ensemble.

4. Statistical Model of Feature Location

4.1 Covariance matrix of a feature point

For actual statistical analysis based on the interpretation described above, we need some additional assumptions. First, we must assume that the "mean" of the potential positions coincide with the true position. In other words, we assume that all hypothetical algorithms are *unbiased*.

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 the point that we want to extract, any algorithm will 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 measures the spread of its potential positions.

Let $V[p_\alpha]$ be the covariance matrix of the αth 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, often called 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* [7]. The matrix $V_0[p_\alpha]$, which we call the *normalized covariance matrix* [7], describes the relative magnitude and the dependence on orientations.

We should note, however, that although we call eq. (1) "the covariance matrix of p_α", it is not a property of p_α; it is a property of the set of hypothetical feature extraction algorithms applied to the neighborhood of p_α.

4.2 Characteristics of feature extraction

Most of existing feature extraction algorithms are designed to output those points that have large image variations around them, so points in a region with an almost homogeneous intensity or on object boundaries are rarely chosen as feature points. As a result, the covariance matrix of a feature point extracted by such an algorithm can be regarded as isotropic. This has also been confirmed by experiments [10], justifying the use of the identity as the normalized covariance matrix.

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 [21].

Many interactive applications require humans to extract feature points by manipulating a mouse. Extraction by a human is also an "algorithm", and it has been shown by experiments that humans are likely to choose "easy-to-see" points such as isolated points and intersections, avoiding points in a region with an almost homogeneous intensity or on object boundaries [10]. In this sense, the statistical characteristics of human extraction are very similar to machine extraction. This is no wonder if we recall that image processing for computer vision is essentially a heuristic to simulate human perception. It has also been reported that strong microscopic correlations exist when humans manually select corresponding feature points over multiple images [13].

(a) (b)

Figure 2: (a) An indoor scene. (b) Detected edges.

4.3 Image processing in computer vision

Thus, we have observed that the ensemble behind geometric inference from images is the set of algorithms and that statistical concepts and assumptions such as normality, independence, unbiasedness, and correlations are *properties of the underlying set of algorithms*. In the past, however, a lot of confusion occurred because these were often taken to be properties of the image.

The main cause of this confusion may be the tradition that the uncertainty of feature points is simply referred to as "image noise". In fact, the assumption described by eq. (1) is usually called the "noise model" rather than "the model of uncertainty of feature location". This confusion is compounded by the convention that the constant ϵ in eq. (1) is called the "noise level", giving a misleading impression as if the feature location is fluctuating by a mysterious random force.

Of course, we may obtain better results if we use higher-quality images whatever algorithm is used (including human intervention). Hence, the performance of any algorithm depends on the image quality. As stated earlier, however, the task of computer vision is not to analyze "image properties" but to study the "3-D properties" of the objects that we are viewing. Since the image properties and the 3-D properties do not correspond to each other one to one, any image processing for computer vision inevitably entails some degree of uncertainty, however high the image quality may be.

Historically, the study of image feature extraction has mainly focused on what is known as *edge detection* (Fig. 2). Its goal is to find the boundaries of 3-D objects in the scene, but in reality all existing algorithms seek *edges*, i.e., lines and curves across which the intensity changes discontinuously. Since this is essentially a heuristic, no definitive algorithm has yet been found and perhaps will ever be.

One of recent studies of edge detection is to remove the boundaries of shadows, which qualify as edges in the usual sense, as non-edges. For this purpose, various clues including the spectrum of color values are

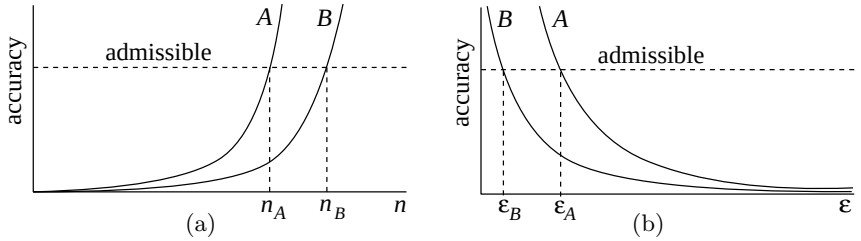


Figure 3: (a) For the standard statistical estimation, it is desired that the accuracy increases rapidly as the number of experiments $n \to \infty$, because admissible accuracy can be reached with a smaller number of experiments. (b) For geometric inference, it is desired that the accuracy increases rapidly as the noise level $\epsilon \to 0$, because admissible accuracy can be reached in the presence of larger uncertainty.

analyzed to judge if both sides of an edge belong to the same object. However, the judgment of whether two parts belong to the same object or not should be different from application to application (e.g., the face and the body are different objects for face recognition while they belong to the same object for human recognition). Thus, the 3-D analysis via image analysis has an inherent limitation.

After all, *any* process of computer vision accompanies uncertainty independent of the image quality, and the result must be interpreted *statistically* in such terms as *likelihood* and *confidence*. The underlying ensemble is the set of *hypothetical (inherently imperfect) algorithms of image processing*, which should be distinguished from "image noise" such as poor resolution and poor illumination. Also, it should be distinguished from the ensemble of input images (for face recognition, for example) of different 3-D objects taken in different conditions.

5. What Is Asymptotic Analysis

5.1 Standard statistical estimation

As stated earlier, *statistical estimation* is to estimate 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, as in experiments in physics, we can increase the accuracy of estimation by controlling the measurement devices and environments. For internal uncertainty, on the other hand, there is no way of increasing the accuracy except by repeating experiments and doing statistical inference based on the assumed model. However, repeating experiments usually entails costs, and often the number of experiments is limited in practice.

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)).

5.2 Geometric inference

As we have seen, we can think of the ensemble for geometric inference based on feature points as the set of potential feature positions that could be located if other (hypothetical) algorithms were used. The goal is to estimate 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* 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, seeking a method whose accuracy rapidly increases as the number n of experiments is meaningless, since we have always $n = 1$. Rather, we want a method whose accuracy is sufficiently high *even for large feature uncertainty*. This observation implies that we need to analyze the growth in accuracy as the noise level ϵ decreases, since a method whose accuracy increases more rapidly as $\epsilon \to 0$ than others can reach admissible accuracy with larger uncertainty of feature extraction (Fig. 3(b)).

6. Asymptotic Analysis for Geometric Inference

We now illustrate our strategy described in the preceding section in more specific terms.

6.1 Geometric fitting

Let p_1, ..., p_N be the extracted feature points. Their true positions $\bar{p}_1$, ..., $\bar{p}_N$ are assumed to satisfy the constraint

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

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

(a) (b)

Figure 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.

A typical problem to is to fit a line or a curve (e.g., a circle or an ellipse) to given N points in the image. For example, assuming that the true positions of the N points are on a parameterized line or curve, we estimate the parameters of the line or curve. The same formulation also applies to constraints on multiple images [7]. For example, if a point (x_α, y_α) in the first image corresponds to a point (x'_α, y'_α) in the second image, we can regard these two points as a single point p_α in a 4-dimensional 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 what is known as the *epipolar equation*, and the parameter $\boldsymbol{u}$ corresponds to the *fundamental matrix*, which encodes the relative positions of the two cameras that took these images [6]. If the scene is a planar surface or located very far away, eq. (2) can be regarded as imposing a (2-dimensional) *homography* (or *projective transformation*) on the two images, where the parameter $\boldsymbol{u}$ is the *homography matrix* [9].

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* over the potential positions of the N feature points is to minimize the squared *Mahalanobis distance* with respect to the normalized covariance matrices $V_0[p_\alpha]$:

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

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

Assuming that the noise level ϵ is small and using Taylor expansion with respect to ϵ, we can show that the covariance matrix $V[\hat{\boldsymbol{u}}]$ of the maximum likelihood solution $\hat{\boldsymbol{u}}$ converges to $\boldsymbol{O}$ as $\epsilon \to 0$ (*consistency*) and that $V[\hat{\boldsymbol{u}}]$ coincides with a theoretical accuracy bound if terms of $O(\epsilon^4)$ are ignored (*asymptotic efficiency*) [7]. Thus, maximum likelihood estimation 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 $\boldsymbol{F}_1(\bar{p}_\alpha, \boldsymbol{u}_1) = \boldsymbol{0}$, $\boldsymbol{F}_2(\bar{p}_\alpha, \boldsymbol{u}_2) = \boldsymbol{0}$, ..., from which we are required to select an appropriate one, the problem is called (*geometric*) *model selection* [7].

A naive idea is to first estimate the parameter $\boldsymbol{u}$ by maximum likelihood estimation and compute the *residual* (*sum of squares*), i.e., the minimum value $\hat{J}$ of J given by (3), for each model and then select the one that has the smallest residual. This does not work, however, because the maximum likelihood solution $\hat{\boldsymbol{u}}$ is determined so as to minimize the residual $\hat{J}$. As a result, the residual $\hat{J}$ can be made smaller if the model has more parameters to adjust.

This observation leads to the idea of compensating for the bias caused by substituting the maximum likely solution. This is the principle of Akaike's *AIC* (*Akaike Information Criterion*) [1], whose theoretical basis is the *Kulback-Leibler information* (or *divergence*) and its asymptotic behavior as the number n of experiments goes to infinity. If we do a similar analysis to Akaike's and examine the asymptotic behavior as the noise level ϵ goes to zero, we obtain the following *geometric AIC* [8]:

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

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

Another well known criterion is Rissanen's *MDL* (*Minimum Description Length*) [17, 18], which measures the goodness of a model by the minimum information theoretic code length of the data and the model. The standard form of the MDL is derived by asymptotic analysis as the number n of experiments

goes to infinity. If, following Rissanen, we quantize the real-valued parameters, determine the quantization width in such a way that the total code length becomes smallest, and analyze its asymptotic behavior as the noise level ϵ goes to zero, we obtain the following *geometric MDL* [8]:

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

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 the model selection is not very much affected by L as long as it has the same order of magnitude [8].

6.3 Equivalent statistical interpretation

Although the above asymptotic analysis is in the "opposite" direction to that of the standard statistical estimation, the final results are similar to the corresponding standard statistical estimation in many respects.

It is known that the covariance matrix of a maximum likelihood estimator for a standard statistical problem converges, under a mild condition, to $\boldsymbol{O}$ as the number n of experiments goes to infinity (*consistency*) and that it agrees with the *Cramer-Rao lower bound* expect for $O(1/n^2)$ (*asymptotic efficiency*). If follows that $1/\sqrt{n}$ plays the same role as ϵ for geometric inference.

The same correspondence exist for model selection, too. Akaike's AIC is derived not from eq. (3) but from its division by ϵ^2. In other words, the normalized covariance matrix $V_0[p_\alpha]$ in eq. (3) is replaced by the covariance matrix $V[p_\alpha]$ given by eq. (1), so that J can be identified with -2 times the logarithmic likelihood. Then, the right-hand side of eq. (4) becomes $\hat{J}/\epsilon^2 + 2(Nd+p) + O(\epsilon^2)$, which is ($-2$ times the logarithmic likelihood)+2(the number of unknowns). This is the same form as Akaike's AIC, since the unknowns are the p parameters of the constraint plus the N true positions, each specified by d coordinates of the d-dimensional manifold defined by the constraint. 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. However, suppose we can repeatedly sample other possibilities. If we observe n samples, an optimal estimate of the true position is the sample mean under the assumption of Gaussian noise. 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 multiple points by a single feature detector by randomly varying the internal parameters (e.g., the thresholds for judgments) [12]. One can then compute the means of the resulting positions and evaluate their covariance matrix. Such a process as a whole can be regarded as one operation that effectively achieves higher accuracy.

In short, the asymptotic analysis for $\epsilon \to 0$ is equivalent to the asymptotic analysis for $n \to \infty$, where n is the number of hypothetical observations. Naturally, the asymptotic behavior of the standard statistical estimation as $n \to \infty$ appears in the asymptotic analysis of geometric inference for $\epsilon \to 0$. As a result, the expression $\cdots + O(1/\sqrt{n^k})$ in the standard statistical estimation turns into $\cdots + O(\epsilon^k)$ in geometric inference.

7. Nuisance Parameters and Semiparametric Model

7.1 Asymptotic parameters

The number n that appears in the asymptotic analysis of the standard statistical estimation 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 an ensemble are revealed more precisely as we sample more elements from it.

However, the number n is often called the *number of data*, which has caused considerable confusion. For example, if we observe a 100-dimensional vector data in one experiment, one may think that the "number of data" is 100, but of course 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 a set of 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 $N \to \infty$.

For geometric inference, however, many researchers have taken the dimension of the data as the "number of data" perhaps because the ensemble is hypothetical and hence one cannot sample more than one datum from it. 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 intensity value, 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 and inferring its properties based

on an assumed model.

7.2 Neyman-Scott problem

Many studies of geometric inference for computer vision in the past have analyzed the asymptotic behavior of estimation as $N \to \infty$ with respect to the number N of extracted feature points without explicitly mentioning what the underlying ensemble is. However, increasing N means considering a *different set of feature points*. This means we are considering *nested ensembles*: an ensemble of "images" having different number of feature points, each image being equipped with an ensemble that describes the feature position uncertainty.

A similar formulation exists in the statistical literature. Suppose, for example, a long 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, which may depend on that orientation. 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.

Evidently, this problem has nested ensembles behind. Since we are interested in the position and orientation of the structure but not the exact orientation of each emission, the variables for the former are called the *structural parameters* while the latter the *nuisance parameters* [2]. This type of formulation is called the *Neyman-Scott problem* [14]. A similar mathematical structure is also found in what is known as the *errors-in-variables model* [4]. In the above example, an optimal solution can be obtained by introducing a parametric model for the laser emission orientations and regarding the actual emissions as randomly sampled from it. This type of formulation is called a *semiparametric model* [2]. However, if each laser emission is regarded as a randomly chosen independent experiment, rather than a single prefixed set of experiments, the noise characteristics change from experiment to experiment. Such a pathological situation is said to be *heteroscedastic* [11].

7.3 Semiparametric model for geometric inference

Since the semiparametric model described above has something different from the geometric inference problem described in Sec. 6, a detailed analysis is required for examining if application of a semiparametric model to geometric inference will yield a desirable result [15]. In any event, one should explicitly state what kind of ensemble (or ensemble of ensembles) is assumed before doing statistical analysis.

This is not merely a conceptual issue. It also affects the performance evaluation of simulation experiments using artificial noise. 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, since opposite results may be obtained for other values of N and ϵ. Here, we have two approaches for the comparison: fixing ϵ and varying N to see if admissible accuracy is attained for a smaller number of feature point; fixing N and varying ϵ to see if admissible accuracy is attained for less certain feature extraction. Since these two approaches have different meanings, the results of one approach cannot directly be compared with the results of the other.

8. Conclusions

In this paper, we have investigated the meaning of "statistical methods" for geometric inference based on image feature points. We traced back the origin of feature uncertainty to image processing operations for computer vision in general and discussed the implications of asymptotic analysis for performance evaluation. This was illustrated in reference to "geometric fitting", "geometric model selection", "nuisance parameters", the "Neyman-Scott problem", and "semiparametric models". The main conclusions of this paper are as follows:

- A *statistical method* is not to study the properties of observed data but to infer the properties of the *ensemble* from which we regard the observed data as having been sampled, assuming some knowledge (or *model*) of the ensemble.
- Statistical analysis does not make sense unless the underlying ensemble is clearly defined.
- The uncertainty of feature location reflects the imperfection of image processing operations. This imperfection is unavoidable, because it is inherent to all computer vision problems.
- The ensemble that reflects the uncertainty of feature location is the set of potential feature positions that could be located by other hypothetical image processing operations.
- If we extract N feature points from an image, we are considering an ensemble of $2N$-dimensional vectors consisting of their x and y coordinates. This ensemble is hypothetical, and only one sample can be observed.

- In such a case, the performance of estimation can be evaluated by an asymptotic analysis for $\epsilon \to 0$ with respect to the noise level ϵ.

- If we could repeat sampling from the hypothetical ensemble, the asymptotic analysis for $\epsilon \to 0$ is equivalent to the standard asymptotic analysis for $n \to \infty$ with respect to the number n of hypothetical observations. Hence, all properties of the standard statistical estimation in the asymptotic limit $n \to \infty$ appear as asymptotic properties of geometric inference for $\epsilon \to 0$.

- The asymptotic analysis for $N \to \infty$ with respect to the number N of feature points is a non-standard mathematical process based on a *semi-parametric model* with nested ensembles. This type of analysis requires careful considerations about the assumptions involved and intended applications, without which performance evaluation by simulation experiments will lose meaning.

Acknowledgements: The author thank Shun-ichi Amari of Riken Brain Science Institute, Japan, David Suter of Monash University, Australia, Pierre-Louis Bazin of Brown University, U.S.A, Peter Meer of Rutgers University, U.S.A., Te-Sun Han of the University of Electro-Communications, Japan, Gérald Battail in France, Naoya Ohta of Gunma University, Japan, Takayuki Okatani of Tohoku University, Japan, and Keisuke Kinoshita of ATR Human Information Sciences Laboratories, Japan for helpful discussions, through which the view presented in this paper has evolved. 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. 13680432), the Support Center for Advanced Telecommunications Technology Research, and Kayamori Foundation of Informational Science Advancement.

References

[1] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control, **16**-6 (1977-12), 716–723.

[2] S. Amari and M. Kawanabe, Information geometry of estimating functions in semiparametric statistical models, *Bernoulli*, **3** (1997), 29–54.

[3] F. Chabat, G. Z. Yang and D. M. Hansell, A corner orientation detector, *Image Vision Comput.*, **17** (1999), 761–769.

[4] W. A. Fuller, *Measurement Error Models*, Wiley, New York, 1987.

[5] C. Harris and M. Stephens, A combined corner and edge detector, *Proc. 4th Alvey Vision Conf.*, Aug. 1988, Manchester, U.K., pp. 147–151.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.

[7] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, the Netherlands, 1996.

[8] K. Kanatani, Model selection for geometric inference, plenary talk, *Proc. 5th Asian Conf. Comput. Vision*, January 2002, Melbourne, Australia, Vol. 1, pp. xxi–xxxii.

[9] K. Kanatani and N. Ohta, Accuracy bounds and optimal computation of homography for image mosaicing applications, *Proc. 7th Int. Conf. Comput. Vision*, September 1999, Kerkyra, Greece, Vol. 1, pp. 73–78.

[10] Y. Kanazawa and K. Kanatani, Do we really have to consider covariance matrices for image features? *Proc. 8th Int. Conf. Comput. Vision*, July 2001, Vancouver, Canada, Vol. 2, pp. 586–591.

[11] Y. Leedan and P. Meer, Heteroscedastic regression in computer vision: Problems with bilinear constraint, *Int. J. Comput. Vision.*, **37**-2 (2000), 127–150.

[12] P. Meer, Personal communications, 2002.

[13] D. D. Morris, K. Kanatani and T. Kanade, Gauge fixing for accurate 3D estimation, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, December 2001, Kauai, Hawaii, U.S.A., Vol. 2, pp. 343–350.

[14] J. Neyman and E. L. Scott, Consistent estimates based on partially consistent observations, *Econometrica*, **16**-1 (1948-1), 1–32.

[15] T. Okatani and K. Deguchi, Is there room for improving estimation accuracy of the structure and motion problem? *Proc. Statistical Methods in Video Processing Workshop*, June, 2002, Copenhagen Denmark, pp. 25–30.

[16] D. Reisfeld, H. Wolfson and Y. Yeshurun, Context-free attentional operators: The generalized symmetry transform, *Int. J. Comput. Vision*, **14** (1995), 119–130.

[17] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.

[18] J. Rissanen, Fisher information and stochastic complexity, *IEEE Trans. Inf. Theory*, **42**-1 (1996-1), 40–47.

[19] C. Schmid, R. Mohr and C. Bauckhage, Evaluation of interest point detectors, *Int J. Comput. Vision*, **37**-2 (2000), 151–172.

[20] S. M. Smith and J. M. Brady, SUSAN—A new approach to low level image processing, *Int. J. Comput. Vision*, **23**-1 (1997-5), 45–78.

[21] Y. Sugaya and K. Kanatani, Outlier removal for feature tracking by subspace separation, *Proc. 8th Symp. Sensing via Imaging Information*, July 2002, Yokohama, Japan, pp. 603–608.

[22] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," CMU Tech. Rep. CMU-CS-91-132, April 1991; `http://vision.stanford.edu/~birch/klt/`.