Model Selection Criteria in Computer Vision: Are They Different?

Niloofar Gheissari and Alireza Bab-Hadiashar

School of Engineering and Science, Swinburne University of Technology,
Hawthorn, Victoria, 3122, Australia.
Email: ngheissari@swin.edu.au, abab-hadiashar@swin.edu.au

Abstract This paper presents an extensive survey of model selection techniques for computer vision applications. A large number of existing model selection criteria and a new model selection criterion (SSC) are introduced and their performance for two important computer vision tasks: motion estimation and range segmentation are evaluated and compared. Various factors affecting the performance of different criteria are introduced and their effects are compared by virtue of conducting controlled experiments using synthetic and real data. Our results show that the performance of different model selection criteria are affected by the size of data and the amount (and distribution) of noise as well as of complexity of models used in an application.

1 Introduction

Many computer vision algorithms rely on using a parametric model, which are usually determined broadly by examining the underlying physical phenomenon. Such physical constraints are often represented by a family of parametric models that can be applicable to various situations of a given task ([2,8,11,13,18,20-22]). Hence, a complete solution to most vision tasks is likely to depend on how well the true underlying model can be chosen.

On the other hand, the model selection problem, which refers to choosing the most appropriate and concise model to express given data in an abstract fashion, has attracted the attention of many statisticians for several decades. Since the introduction of Akaikeís An Information Criterion (AIC) [1], which had a fundamental effect on model selection research, many model selection criteria have been introduced [6,14-16,22]. It comes as no surprise that many of those model selection techniques have been employed in many computer vision algorithms for various applications ([2,8,11,13,18,20-23]).

Although the general form of a model that underpins a specific task is most likely to be mandated by its physical characteristics, there are situations where simpler models (than the general form) can be truly applicable to those particular situations. For example, in 2D motion recovery, while a third order polynomial represents the rigid motion of a generic surface, the motion of a planar surface can be represented by a subset of that polynomial (a partial quadratic model). If there is no a priori information about the shape of surfaces in the scene, a method to choose the correct model is an important ingredient.
To determine the correct underlying model of a data set, one may simply suggest the most appropriate model is the one, which best fits to the data. This idea, however, does not work because it will always favour the most complex model of a model library. The reason is that the most complex model has more degrees of freedom and can therefore fit to the data better than any other model in that library. Thus, to choose the correct model, one needs to establish a trade off between fidelity (how well a model fits the data, which is often measured by the sum of squared residuals) and the complexity of that model. In practice, higher order models have to be penalized so that the selected model would be chosen based on its suitability rather than its fidelity to data. In fact, the salient difference between all the

existing model selection criteria is in the way by which they penalize the higher order models.

In this paper, we explain, evaluate and compare an extensive set of existing model selection criteria for two important computer vision applications namely 3D range segmentation and 2D motion segmentation. Although we think that there are many factors affecting the suitability of a criterion for employment in a computer vision application, the most important ones are:

- *Application:* The nature of physical constraints (which inturn depends on the application) can affect the performance of a model selection criterion. For example, the performance of a model selection criterion for range segmentation differs from its performance for motion segmentation application, as the types of models used in these applications are different.

- *Data Size:* As we will show, the size of data (size of image, region etc) can significantly change the performance of a model selection criterion.

- *Noise:* The scale and distribution of the noise, which are different in synthetic and real data, can affect the performance of a model selection criterion. Most model selection criteria are derived based on a priori assumption about the distribution of the noise.

- *Model Library*: The model library, from which the most appropriate model is chosen, may include very similar models. In this case, there can be a reduction in the performance of a model selection criterion. For example, the distinction between closely nested models, which have very similar terms, can be challenging to any criterion.

2 Statistical Model Selection Criteria

Here, we briefly explain (in chronological order) a number of popular and effective model selection criteria. These techniques appear to perform well in computer vision applications and have been commonly used in various applications. In the following, P refers to the number of parameters of a model, r_i, denotes the residual for the i^{th} data point (Σr_i^2 is therefore the sum of squared residuals). We show the scale of noise by δ and the number of data points by N. The dimension of the surface that fits to the data is denoted by d.

Akaikeís An Information Criterion (AIC)

AIC [1] was one of the first model selection criteria introduced in statistics literature. AIC is based on the idea that a chosen model is correct if it can sufficiently describe any future data with the same distribution and therefore AIC can be regarded as a *hypothetical cross validation* method [12]. In other words, AIC selects a model that minimises the expected error of the new observation with the same distribution as the data used for fitting (which is the current observation). AIC uses Maximum Likelihood Estimation (MLE) technique for computing the residuals and has the following form:

$$AIC = \sum_{i=1}^{N} r_i^2 + 2P\delta^2 .$$

Since 1973, AIC has been modified in many ways. For example, many model selection criteria including CAIC [6], CAICF [6], GAC [22], GAIC [14] and MAIC [5] are derived from AIC.

CP

Around the same time as the introduction of AIC, Mallow pioneered another model selection criterion called CP [15]. CP selects the model that minimises the mathematical expectation of *ì scaled sum of squared errorî* , where the error is defined as the algebraic distance between the predicted and observed data. It means that the error is calculated by a linear regression method instead of MLE. Therefore, unlike most of the statistical model selection criteria, CP does not rely on MLE technique for evaluating the residuals (errors). CP has the following form:

$$CP = \sum_{i=1}^{N} r_i^2 + (-N + 2P)\delta^2$$

In the recent years, Kanatani [12] reported that Geometric CP is equivalent to GAIC (explained below).

Minimum Description Length (MDL)

Later in 1978, Rissanen introduced MDL [16,17]. The underlying logic of MDL is that the simplest model that sufficiently describes the data is the best model. For example, if one aims at encoding and transmitting a given data set, the best encoding model is the model that generates the least total size of transmitted data. MDL has the following form:

$$MDL = \sum_{i=1}^{N} r_i^2 + (P/2)\log(N)\delta^2 \cdot$$

Shortest Data Description (SSD)

Similarly in 1978, Rissanen proposed another model selection criterion termed SSD [17]. SSD selects the model that minimises the bit representation of the data and therefore has a very similar underlying logic with MDLs and has shown (Figures 1, 2 and 3) to be efficient where the noise distribution is Gaussian. SSD has the following form:

$$SSD = \sum_{i=1}^{N} r_i^2 + (P\log((N+2)/24) + 2\log(P+1))\delta^2 \cdot$$

Bayesian Information Criterion (BIC)

BIC chooses the model that maximises the conditional probability of describing a data set by a model constrained by some priori information. Thus, BIC can take various forms based on the nature of the assumed priori information. For example, an instance of BIC was introduced by Schwarz [19] in 1978 and has the following form:

$$BAYES = (2\pi)^{P/2} Log(\theta_m)\sqrt{1 \Big/ \sum_{i=1}^{N} r_i^2}$$

where θ_m denotes the estimated parameters of each model.

Consistent AIC (CAIC)

CAIC [6], proposed by Bozdogan in 1987, is an attempt to overcome the tendency of the AIC to overestimate the complexity of the underlying model. In formulating CAIC, a correction factor based on the sample size (N) is employed to compensate for the overestimating nature of AIC. Logarithm of *N* has been suggested as an appropriate instance of such factor. CAIC can be written as:

$$CAIC = \sum_{i=1}^{N} r_i^2 + P(\log N + 1)\delta^2 .$$

In computer vision context, Bubna and Stewart [7] reported that CAIC has a satisfactory performance in surface merging for reconstruction algorithms.

Geometric Bayesian Information Criterion (GBIC)

Almost twenty years after the introduction of BIC, Heckerman and Chickering proposed GBIC [9]. This criterion, which is based on BIC, appears to be more effective than BIC perhaps, due to the fact that it somehow reduces the tendency of BIC to over-predict the complexity of the model. GBIC can be written as:

$$GBIC = \sum_{i=1}^{N} r_i^2 + (Nd \log(4) + P\log(4N))\delta^2 \cdot$$

Geometric Akaikeís Information Criterion (GAIC)

GAIC is based on AIC. The main difference between Geometric AIC (GAIC), introduced by Kanatani [14], and AIC is in the way that the expectation functions are evaluated. GAIC is specifically derived for geometric fitting. As described by Kanatani, the objective of geometric fitting is to estimate the model parameters from observed data. In geometric fitting, the properties of the noise are assumed to be known a priori. In contrast, in the statistical inference procedures, one aims at estimating both the model parameters and the properties (mean and variance) of the noise. In other words, in geometric fitting the main objective is to study the observe data themself whilst in statistical methods one aims at studying the ensemble from which the observed data are sampled. According to Kanatani, in geometric fitting the ensemble is the set of all algorithms that can be applied to solve a specific problem. Therefore, the estimation accuracy only improves if noise is decreased. This is in sharp contrast to the fact that in statistical inference, the estimation accuracy is improved by increasing the number of sample points.

Kanatani derives a first order approximation of GAIC which can be written as:

$$GAIC = \sum_{i=1}^{N} r_i^2 + 2(dN + P)\delta^2 .$$

GIC

In 1998, Torr [22] proposed GIC that is a modified version of AIC. In AIC, if the number of data points (*N*) is much larger than the number of parameters of a model (P) then the influence of *P* on the AICís decision is greatly diminished. In practice, this is a crucial problem since the main task of a model selection criterion is to determine the correct number of parameters of the appropriate model. To avoid this problem, Torr proposed using adjustable coefficients for *N* and *P* and rewrote the GAIC to be:

$$GIC = \sum_{i=1}^{N} r_i^2 + \lambda_1 dN\delta^2 + \lambda_2 P\delta^2 .$$

Torr suggested that $1>\lambda_1>2$, $\lambda_2>2$ will provide satisfactory outcomes. The same author reported that where the difference between the dimensions of the compared models is one, and the required level of significance is $\alpha = 0.0456$, then, $\lambda_1=2$ and $\lambda_2=4$ would be reasonable values to use.

Geometric MDL (GMDL)

Shortly after introducing GAIC, Kanatani derived GMDL [14] specifically for geometric fitting (see GAIC). GMDL, similar to MDL (described previously), relies on knowing the scale of noise. This criterion has the following form:

$$GMDL = \sum_{i=1}^{N} r_i^2 - (Nd + P)\delta^2 \log(\delta / L)^2$$

where L is the reference length and can be determined exactly or as Kanatani suggests can be approximated by a practical scale such as the image size.

Surface Selection Criterion (SSC)

Recently, the authors have also proposed a model selection criterion named SSC [3,10], which is an attempt to combine both the geometrical and physical characteristics of the data in order to identify the true underlying model. To achieve this, the authors have devised a scheme that generally favours models that store lesser amount of strain energy (as a measure of roughness) while closely follows the data. The proposed criterion is as follows:

$$SSC = \sum_{i=1}^{N} r_i^2 / N\delta^2 + P \frac{E_{Bending+Twist}}{E_{\max}} .$$

$E_{\text{Bending}+Twist}$ is the strain energy, which is computed as:

$$E_{Bending+Twist} = \iint_s \frac{1}{2}\left\{ \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right)^2 - 2(1-\nu)\left[\frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2} - \left(\frac{\partial^2 w}{\partial x \partial y}\right)^2\right]\right\} dxdy$$

where w is the surface we fit and ν is Poissonís ratio and it should be very small since the twisting energy, in comparison with the bending energy, is small. In our experiments we assume $\nu = 0.01$. Experiments have shown that the performance of SSC is not sensitive to this value. $E_{\max}$ is the strain energy of the model with the highest number of parameters in a set of nested models.

3 Evaluation of Different Model Selection Criteria

To evaluate and compare the performance of various model selection criteria (explained previously), we have chosen two important computer vision applications in which model selection plays an important role. Those applications are:

- Detecting the true underlying motion model in optic flow calculation (2D motion segmentation).
- Detecting the true surface model for 3D range data measurements.

In order to evaluate the performance of the proposed model selection criteria in each of these applications, we generated a number of different sets of synthetic data and implemented all of the criteria by calculating their mathematical expressions. The residuals of each model remain the same for all of the criteria. Details of our experiments as well as their results are presented in the following two sections.

It is important to note that since almost all criteria (except CP and SSC) are derived based on the MLE technique, the residuals also need to be calculated using MLE. However, during our experiments, we noticed that the performances of these criteria were deteriorated when we attempted to calculate the residuals based on the MLE. This is mainly due to the fact that the objective functions of the MLE technique are non-linear and the residuals cannot be accurately calculated[1]. Therefore, to make the computation feasible, we have used the algebraic residuals for evaluating all the criteria. The same scale of noise for all the criteria was also used and computed according to $\delta^2 = \sum_{i=1}^{N} r_i^2 /(N - P_h)$ where N is the number of data points and P_h is the number of parameters of the highest model in the library. The reason that we use the scale of noise for the highest surface (as Kanatani [12] described) is that the scale of noise for the correct model and the scale of noise for the higher order models (higher than the correct model) must be close for the fitting to be meaningful.

To measure the success of every criterion, we have divided the number of correct predictions of the underlying model by the total number of different (synthetic) data sets used in evaluating each technique.

Optic Flow Calculation and Motion Segmentation

To compare the performance of different model selection criteria for motion segmentation purposes, we have generated three sets of synthetic image sequences in which the underlying motion of their intensity patterns are known. Following Barron et al. [4], we also used a sinusoidal texture for our synthetically generated image sequences. The models of motion we used were: Affine, Partial-Quadratic and Quadratic as shown in Table 2. These models are chosen because they are commonly used to describe the motion of man-made objects in video sequences. For example, a Partial-quadratic model expresses the rigid motion of planar surface while the rigid motion of a curved surface can be approximated by a Quadratic model. Affine model of motion is also very popular as it adequately explains common movements of camera such as pan and zoom.

Model	Horizontal and Vertical Velocities
Affine	$U=ax + by + c$ & $V= cx + dy + e$
Partial-Quadratic	$U=ax^2 + bxy +cx +dy+e$ & $V=by^2 + axy + fx + gy + h$
Quadratic	$U=ax^2 + bxy +cy^2 +dx + ey + f$ & $V= gx^2 + gxy +iy^2 +jx + ky + l$

Table 1-The motion models used in our evaluation experiments.

In our experiments, we have randomly changed the parameters of every model 100 times and applied the different model selection criteria to test how well they can identify the true underlying model. Figure 1 shows the success rate of each criterion in identifying the true underlying model. Since the performance of every criterion can be affected by the size of each synthetic image [8], the experiments were repeated for images of various sizes (from

[1] While calculating algebraic residuals takes less than 1 minute, finding the MLE based residuals for only 50 different sets of data can take up to several days.

21x21 to 71x71 pixels). The image size for each experiment is also shown along the horizontal axis.

Results depicted in Figure 1 clearly show that there is a significant reduction in performance of each criterion where small images were used (smaller than 35*35). The performance deterioration is worse in criteria such as: GMDL, GAIC, GBIC and specially GMDL. SSC, GAIC and Mallowís CP appear to perform better for small images. The significant reduction of GMDLís performance can be the effect of using inappropriate *ì reference lengthî* (to make a fair comparison, a fixed reference length is used in all the experiments).

Although there is not any criterion that is superior in all cases, Figure1 shows that the SSC has a reliable and consistent performance even when it is applied to small images in which finding the appropriate model is challenging. GMDL, SSD, CAIC, GBIC also appear to work well for motion segmentation task. Their success can be attributed to the fact that they usually penalise the higher order models more than the other model selection criteria. Considering the mathematical expression of each criterion, in general when N is much greater than 25, the relationship between the penalty terms proposed by each criterion is as follows: (penalty term of) GBIC> CAIC> SSD>MDL>CP.

The penalty term of GMDL highly depends on the scale (δ) of noise and its magnitude cannot be generalised as above. The penalty term of GAIC is much larger than GBIC however as we described before the influence of the number of parameters is small. It is interesting (but not surprising) to note that the two criteria that are theoretically considered equivalent: GCP and GAIC (see [12] for details) performed similarly in our experiments.

In order to provide a comparative measure of success, we calculated and plotted (shown in Figure 2) the average success rate of different model selection criteria over the different image sizes (from 21x21 to 71x71 pixels). It can be seen from this figure that SSC outperforms the other criteria on average and its success rate is about 10% more than the others. Other competitive model selection criteria: SSD, GMDL, GAIC and GBIC have similar (average) rate of success.

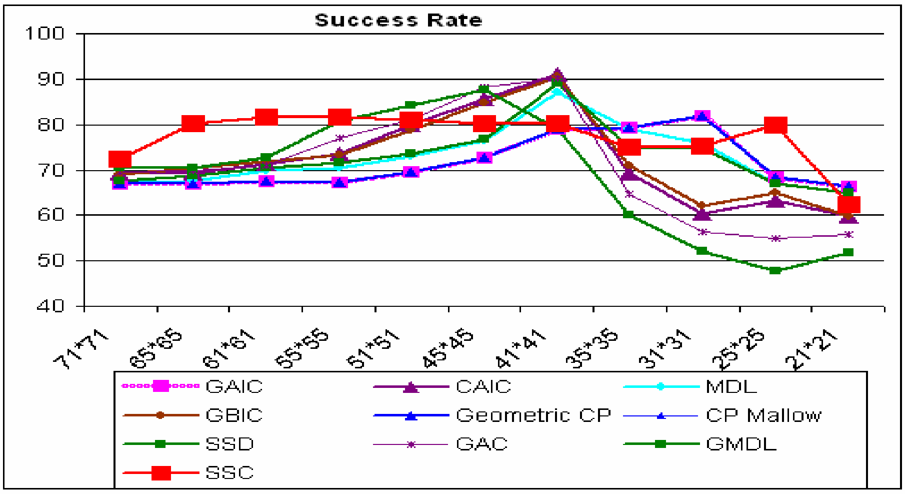

Figure 1-The success rates for different model selection criteria.

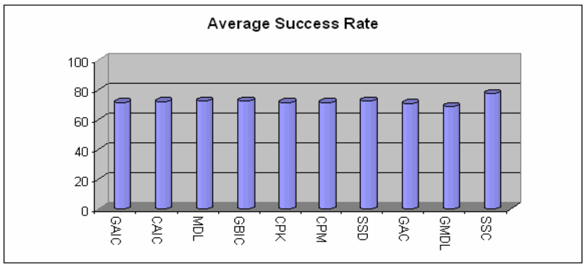

Figure 2-Average of correct prediction for different model selection criteria. We averaged the success rate of these criteria applied on image sizes of 21*21 to 71*71.

Another important piece of information associated with any model selection criterion is to discover whether the criterion is likely overestimate or underestimate the dimension of the underlying model when it fails. To calculate such a measure, we applied each criterion to 300 images of size 51*51 (100 images for each of three motion models) and recorded the number of times when each criterion has overestimated or underestimated the dimension of the underlying model. The results are shown in Figure 3. As can be seen from this figure, except for SSC almost all the other criteria tend to overestimate the dimension of the underlying motion model. The difference between the behaviour of SSC compare to others is not unexpected as there are substantial differences between the underlying logic of SSC and other model selection criteria.

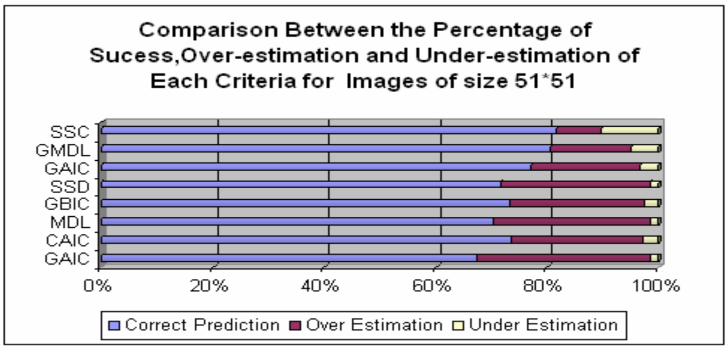

Figure 3-Various bar colours represent the percentage of success, underestimation and overestimation of model dimensions for every criterion. The size of images used is 51x51 pixels.

Range Segmentation

Range data segmentation is one of the fundamental problems of computer vision and has been studied for many years. Since model selection is a crucial part of any range segmentation scheme, we have chosen this problem as a means to evaluate and compare all the previously described model selection criteria. In our experiments, we expect a model selection criterion to be able to identify (from a library of known models) the true underlying surface model of a set of range data. In addition to all the criteria we explained and evaluated previously, we have also added the Modified AIC (MAIC ñ developed by Boyer et al. [5]) in our list of model selection techniques. MAIC is developed exclusively for range segmentation application and is based on the assumption that the error has a *t* distribution.

To perform our evaluation experiments, we first created eight synthetic data sets according to the surface models in Surface Library 1 and randomly changed the parameters of each data set 100 times. We also added 1% normally distributed additive noise. We then applied different model selection criteria on each synthetic data set to determine the true surface model. The success rate of every criterion in correctly recovering the underlying model of data is shown in Figure 4.

To provide more realistic measure of how useful a criterion might be, we then proceeded to examine the success rate of different model selection criteria on real range images. We chose 50 different images of range data measurements of various objects containing both quadratic and planar surfaces and applied different model selection criteria to the whole set. The results of those experiments are also shown in Figure 4. As it can been seen from Figure 4, for synthetic data, almost all criteria have similar performances except SSC which has higher success rate compare to other criteria. For real data, as shown in Figure 4, it appears that only SSC and MAIC perform well and the rest of criteria have little success in selecting the true surface model of real range data. Our experiments also showed that SSC is considerably better in choosing the right model when it is applied to our experimental set of real range data. An important point to make is that there is a huge difference between the performance of statistically based model selection criteria on real and synthetic data. This can be due to the fact that most assumptions used to derive these criteria are not realistic and thus, the success rate for real data experiments are very poor.

Model 1	$ax^2 + by^2 + cz^2 + dx + ey + fz = 1$
Model 2	$ax^2 + by^2 + cx + dy + eyx = 1$
Model 3	$ax^2 + bz^2 + cx + dz + exz = 1$
Model 4	$az^2 + by^2 + cz + dy + fyz = 1$
Model 5	$ax^2 + by^2 + cx + dy = 1$
Model 6	$ax^2 + bz^2 + cx + dz = 1$
Model 7	$ay^2 + bz^2 + cy + dz = 1$
Model 8	$ax + by + cz = 1$

Surface Library 1: From the most general model to the simplest model.

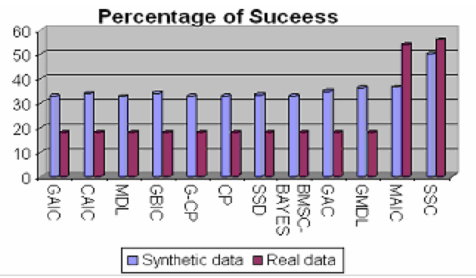

Figure 4: Success rate of various model selection criteria for synthetic (light bar) and real range data (dark bar). The size of all images used is 101x101 and the scale of additive noise is 1%.

4 Conclusion

An extensive survey of model selection techniques for computer vision applications is presented. A number of controlled experiments using synthetic and real data were used to compare the performance of different model selection criteria for motion estimation and range segmentation. It is shown here that although many model selection techniques work well for motion estimation task, few can distinguish between planar and quadratic surfaces of range data measurements. Our newly proposed model selection criterion, SSC, has been shown to outperform the other existing criteria in almost all cases.

References

1. Akaike, H., A New Look at the Statistical Model Identification, IEEE Trans. on Automatic Control, vol. AC-19, pp. 716-723, 1974.

2. Bab-Hadiashar, A., Gheissari, N., and Suter, D., Robust Model Based Motion Segmentation, ICPR 2002, Quebec, Canada, pp. 753-757, Aug. 2002.
3. Bab-Hadiashar, A. and Gheissari, N., Parametric Robust Curved Objects Range Data Analysis: Model Selection and Segmentation, Technical report, Swinburne University.
4. Barron, J. L., Fleet, D. J., Beauchemin, S. S., and Burkitt, T. A., Performance of Optical Flow Techniques, Proceedings of CVPR'92, Washington, pp. 236-42, 1992.
5. Boyer, K. L., Mirza, M. J., and Ganguly, G., The Robust sequential Estimator: A General Approach and Its Application to Surface Organization in Range Data, IEEE PAMI, vol. 16, pp. 987-1001, 1994.
6. Bozdogan, H., Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions, Psychometrica, vol. 52, pp. 345-370, 1987.
7. Bubna , K. and Stewart, V. C., Model Selection and Surface Merging in Recognition Algorithms, ICCV 98, Bombay, India, pp. 895-902, 1998.
8. Bubna, K. S. C. V., Model Selection Techniques and merging Ruls for Range Data Segmentation Algorithms, Computer Vision and Image Understanding: CVIU, vol. 80, pp. 215-245, 2000.
9. Chickering, D. and Heckerman, D., Efficient Approximation for the Marginal Likelihood of Bayesian Networks with Hidden Variables, Machine Learning, vol. 29, no. 2-3, pp. 181-212, 1997.
10. Gheissari, N. and Bab-Hadiashar, A., Motion Analysis: Model Selection and Motion Segmentation, Proceedings of 12th International Conference on Image Analysis and Processing, pp. 442-448, Sep, 17-19,2003.
11. Gu, H., Shirai, Y., and Asada, M., MDL-Based Segmentation and Motion Modeling in a Long Sequence of Scene With Multiple Independently Moving Objects, The Journal of the Pattern Recognition Society, vol. 28, pp. 781-793, 1995.
12. Kanatani, K., What Is the Geometric AIC? Reply to My Reviewers,1987(unpublished).
13. Kanatani, k., Evaluation and Selection of Models for Motion Segmentation. Memoirs of the Faculty of Engineering, Okayama University, vol. 36, pp. 91-106, Dec, 2001.
14. Kanatani, K., Model Selection for Geometric Inference, The 5th Asian Conference on Computer Vision, Melbourne, Australia, pp. xxi-xxxii, January 2002.
15. Mallows, C. L., Some Comments on CP, Technometrics, vol. 15, pp. 661-675 , 1973.
16. Rissanen, J., Universal Coding, Information, Prediction and Estimation, IEEE Trans. Inf. Theory, vol. 30, pp. 629-636, 1984.
17. Rissanen, J., Modeling by Shortest Data Description, Automatica, vol. 14, pp. 465-471, 1978.
18. S., T. P. H., Geometric motion segmentation and model selection, Phil. Trans. R. Soc. Lond. A, vol. 356, pp. 1321-1340, 1998.
19. Schwarz, G., Estimating the Dimension of a Model, The Annals of Statistics, vol. 6, pp. 461-464, 1978.
20. Tissainayagam, P. and Suter, D., Visual Feature Tracking with Automatic Motion Model Selection, Proceedings of JCIS, N.C. USA , pp. 322-325, Nov. 1998.
21. Torr, P. H. S., An assessment of information criteria for motion model selection, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 47-53, 1997.
22. Torr, P. H. S., Model Selection for Two View Geometry: A Review, Model Selection for Two View Geometry: A Review, Microsoft Research, USA, Microsoft Research, USA, 1998.
23. Zhu, S. C. and Yuille, A. L., Region Competition: Unifying Snakes/Balloon, Region Growing, and Bayes/MDL/Energy for Multi-Band Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, pp. 884-900, 1996.