Combining Geostatistical Indicator Kriging with Bayesian Approach for Supervised Classification

No-Wook Park[1)], Kwang-Hoon Chi[1)], Wooil M. Moon[2), 3)], Byung-Doo Kwon[4)]

1) National Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources,
30 Kajung-dong, Yusung-ku, Daejeon 305-350, Korea
Email : nwpark@kigam.re.kr
2) ESI[3] Lab., School of Earth and Environmental Sciences, Seoul National University, Korea
3) Geophysics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
4) Department of Earth Sciences, Seoul National University, Korea

Abstract

In this paper, we propose a geostatistical approach incorporated to the Bayesian data fusion technique for supervised classification of multi-sensor remote sensing data. Traditional spectral based classification cannot account for the spatial information and may result in unrealistic classification results. To obtain accurate spatial/contextual information, the indicator kriging that allows one to estimate the probability of occurrence of classes on the basis of surrounding observations is incorporated into the Bayesian framework. This approach has its merit incorporating both the spectral information and spatial information and improves the confidence level in the final data fusion task. To illustrate the proposed scheme, supervised classification of multi-sensor test remote sensing data set was carried out.

1. Introduction

During the last decade, advances of various data acquisition techniques have made it possible to compile large volumes of spatial data from different sources. New methods for spatial data fusion have been proposed and applied to handle the greater diversity and volume of modern data sets for more complex information extraction in the geosciences (e.g. Bayesian probabilistic approach, fuzzy logic, evidential reasoning, neural network, etc). In remote sensing fields, these approaches have been more widely tested and refined with more rigorous mathematical backgrounds (Lee *et al.*, 1987; Serpico *et al.*, 1996; Solberg *et al.*, 1996; Solaiman *et al.*, 1999).

A key difference feature of a geoscience data set is the fact that each observation relates to a particular location in space. Quantitative knowledge of an attribute value is of little interest unless the location of the observation (measurement) is exactly known and accounted for throughout the analysis. The main limitation of traditional spectral classifiers is that the proximity of a pixel to a class is computed only in the spectral space, without any consideration for the spatial coordinates of the pixel. As a result, classifications often display noisy or unrealistic features, such as isolated pixels assigned to a particular class. To overcome this type of drawbacks, various parametric methods (e.g. Gibbs-Markov random field) have been developed based on the basis of statistical properties of the data, such as the mean, the standard deviation, covariance, and texture (Solberg *et al.*, 1996; Tso and Mather, 1999). However, since the method contains many parameters which are difficult to interpret, it is difficult to formulate an effective method which can correctly infer the parameters for the given model.

Geostatistics can provide a collection of deterministic and statistical tools, so that we can better understand the problem and model the spatial variability (Isaaks and Srivastava, 1989). It was originally devised to estimate statistical properties on unsampled points for delineating ore deposits. But these days those tools are used not only for estimation of unsampled points but also for inference for local and spatial uncertainty estimation (Goovaerts, 1997). In the remote sensing data processing, geostatistics can be applied for incorporating the spatial and temporal coordinates of observations in data processing and assessing the uncertainties.

In this paper, we apply a geostatistical method to Bayesian multi-sensor data fusion. Geostatistics is applied for incorporating the spatial coordinates of observations in data. Of particular interest is the indicator kriging that allows one to estimate the probability of occurrence of classes on the basis of surrounding observations. This scheme has its merit to incorporate various independent information (e.g. spectral information and location information) afterward and improves the confidence of data fusion stage. Supervised land-cover classification using multi-sensor remote sensing data was applied to illustrate application of this methodology.

2. Methodologies

For supervised land-cover classification, we employed the Bayesian data fusion method based on smoothed kernel method and geostatistical indicator kriging (Fig.1). In this section, we will briefly review the concepts of applied methods.

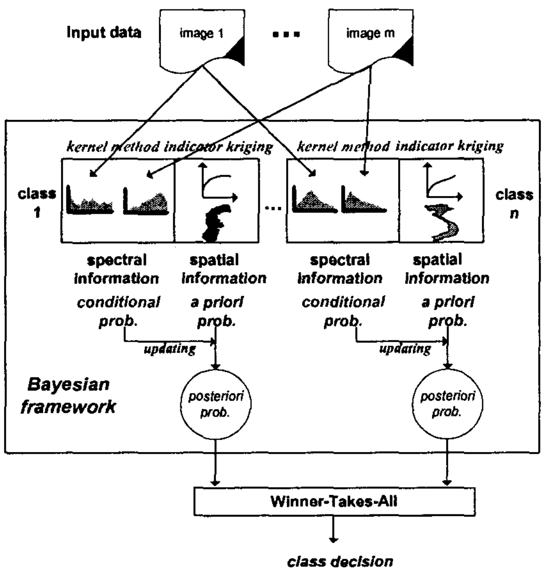

Fig. 1 Applied methods in this study.

2.1 Bayesian Framework for Supervised Classification

Among various data fusion methods reported in the remote sensing literature for the classification of multi-sensor image data, we focus on the statistical approach, especially the concepts of Bayesian methods for data fusion.

Bayesian probabilistic approach can provide us with a formalism for reasoning about partial beliefs under conditions of uncertainty. In this formalism, propositions are given numerical parameters signifying the degree of belief accorded them under some body of knowledge and the parameters are combined and manipulated according to the rules of probability theory (Moon, 1993; Pearl, 1997).

In a general multi-sensor data fusion, one can suppose that we have a set of data from m sensors $X_s(i,j)$, $s \in \{1,\cdots,m\}$ for each pixel (i,j). The final goal for classification is to assign each pixel into one of the predefined information classes $\omega_1,\cdots,\omega_n$. In a Bayesian formalism, the relationship between the measurements and *a priori* probabilities is represented by

$$P(\omega_c \mid X_1(i,j),\cdots,X_m(i,j)) = \frac{P(X_1(i,j),\cdots,X_m(i,j) \mid \omega_c)P(\omega_c)}{P(X_1(i,j),\cdots,X_m(i,j))}, \quad c \in \{1,\cdots,n\}$$

where, $P(\omega_c)$ is *a priori* probability of class ω_c and $P(\omega_c \mid X_1(i,j),\cdots,X_m(i,j))$ is the conditional probability (a *posteriori* probability) that ω_c is the correct class, given the data $X_1(i,j),\cdots,X_m(i,j)$. The image formation model $P(X_1(i,j),\cdots,X_m(i,j) \mid \omega_c)$ is the conditional probability that $X_1(i,j),\cdots,X_m(i,j)$ is the observed data, given that ω_c is the correct class. Each pixel is assigned to the class C which maximizes $P(\omega_c \mid X_1(i,j),\cdots,X_m(i,j))$.

In most cases, it is very difficult to estimate the joint conditional distribution $P(X_1(i,j),\cdots,X_m(i,j) \mid \omega_c)$. If we assume the measurements from different sensors be conditionally independent, then we can simplify the mathematical analysis and computations and the image formation model can be expressed as a product of sensor-specific conditional distributions:

$$P(X_1(i,j),\cdots,X_m(i,j)\mid\omega_c)=\prod_{i=1}^{m}P(X_i(i,j)\mid\omega_c)$$

2.2 Smoothed Kernel Method

In the parametric approach, to get the conditional probabilities $P(X_1(i,j)\mid\omega_c)\cdots P(X_m(i,j)\mid\omega_c)$, we require certain statistical assumptions. The maximum likelihood technique is based on the assumption that digital values to be processed are multi-dimensional normally distributed. However, as many researchers have criticized before, it becomes especially more difficult in the multi-sensor/source data fusion than in a single source data processing. An alternative is to adopt a non-parametric method such as the kernel method or *k*-nn method that does not make any assumption about the distribution (Duda *et al.*, 2000).

In this paper, instead of the parametric method, we have employed a non-parametric method based on the smoothed kernel method (Silverman, 1986), which is an improved version of histogram approach. Given a set of N samples X_i drawn from a statistical distribution $p(X)$, the kernel method provides a consistent estimate of the related probability density function $\overline{p}(X)$ by using an appropriate "kernel function" $k(\cdot)$ which is applied to each sample considered, i.e.,

$$\overline{p}(X)=\frac{1}{N}\sum_{i=1}^{N}k(X-X_i)$$

The shape of the distribution using the kernel method depends only on the shape of the reference rectangle, or kernel. In the smoothed kernel method, a Gaussian "bell-shaped" curve, instead of a rectangle, is used as a kernel to construct the empirical probabilistic density function.

2.3 Geostatistical Indicator Kriging

The most difficult and important aspect of the Bayesian frame is to estimate *a priori* information. To estimate *a priori* probability, the simplest approach is to assume that all classes have a same *a priori* probability. Besides, if the training data set is assumed to be representative of the proportion of different classes within the study area, the priors can be set proportional to the sample size. In this data fusion context, one can use *a priori* information in various formats, which may include, e.g. information about sensor-specific noise characteristics or information about the weather conditions at the time the images were acquired (Solberg *et al.*, 1996).

In this paper, we adopted a geostatistical indicator kriging to construct *a priori* information proposed by Goovaerts(2002). In this approach, *a priori* probabilities are replaced by the *posterioriori* probabilities related to the spatial information obtained by the geostatistical indicator kriging. One can thus combine the spatial information with spectral information obtained by smoothed kernel method utilizing the indicator kriging.

Geostatistics is largely based on a random function model, whereby the set of unknown values is regarded as a set of spatially dependent random variables. Spatial correlation allows making predictions about the property at unsampled locations from sample data. Kriging, basically a form of generalized linear regression, is a name for a spatial estimation technique, that uses the variogram as a model of geological continuity and estimates unsampled locations on that basis. Unlike continuous variables, categorical attributes such as land-cover classes cannot be estimated as a mere linear combination of neighboring observations. In many situations, the unsampled location is simply allocated to the same category as the nearest observation. Qualitative and/or quantitative information about spatial correlation between categories can be handled by an indicator algorithm, as long as it is coded into indicator values, say 1 if the category is present and 0 otherwise (Goovaerts, 1997). Then, the indicator kriging is used to estimate the probability for each indicator state of the certain class as a linear combination of neighboring indicator data.

If we suppose that $u(i,j)$, n_c denote the data locations and the number of surrounding training data, respectively, the indicator kriging system may be wriiten as follows;

$$P(\omega_c)\approx P(\omega_c\mid u(i,j))=\sum_{\alpha=1}^{n_c}\lambda(u_\alpha;\omega_c)\cdot i(u_\alpha;\omega_c)$$

where $i(u_\alpha;\omega_c)$ is the indicator transformed value (e.g. 0 or 1) of certain class ω_c at training data location u_α.

Using models of spatial dependency, the weights $\lambda(u_\alpha;\omega_c)$ are determined by solving the ordinary kriging system under the unbiasedness condition:

$$\sum_{\alpha=1}^{n_c}\lambda(u_\alpha;\omega_c)=1$$

To satisfy the order relations, any probability outside the interval [0,1] is reset to the closest bound, 0 or 1. Subsequently, the estimates are standardized by their sum.

3. Experiments

3.1 Data Set Description

To illustrate the proposed method, we applied the technique to multi-sensor remote sensing data set publicly posted by the IEEE GRSS Data Fusion Committee. It consists of multi-sensor remote sensing images related to an agricultural area near the village of Feltwell (UK) and is the same as considered in Serpico *et al.*(1996) and Giacinto *et al.*(2000). A section (250 by 350 pixels) of a scene acquired by an optical sensor (an Airborne Thematic Mapper scanner) and a radar sensor (a NASA/JPL synthetic aperture radar) was used. As for the optical sensor data, the six ATM channels corresponding to Landsat TM channels in the visible and in the infrared spectrum band were included. As for the radar data, the nine SAR channels in the PLC frequency bands and full HH, HV, and VV polarizations were included. The images, 15 in total, are filtered and normalized. To compare the experimental results with previous results, we selected 10,944 pixels belonging to five agricultural classes (e.g. sugar beets, stubble, bare soil, potatoes, and carrots) and the data set contains 5,124 training and 5,820 reference pixels.

3.2 Results

In our approach, the training data were used for two purposes. First, they were used to calculate a priori probabilities for 5 classes, assuming that we have no spectral information, until now. The second use was to calculate the conditional probabilities based on spectral information for 5 classes, assuming that we have spectral information, which is the same one as traditional spectral based classification.

To obtain the *a priori* probabilities for 5 classes, we performed the indicator kriging. First, each of the 5,124 training data was coded into a vector of five indicators and then five experimental indicator variograms were calculated. Then, we obtained the spatial based probabilities at the whole data locations using indicator kriging.

After obtaining *a priori* probabilities based on the indicator kriging, we applied the smoothed kernel method for conditional probabilities based on the spectral information, a value 2% of data range of the spread parameter in the Gaussian kernel function was selected experimentally as a result of the training phase. The *posteriori* probabilities for 5 land-cover classes were estimated by combining and updating the *a priori* probabilities with conditional probabilities. After this step, each pixel is assigned to certain class which maximizes a *posteriori* probabilities.

The classification estimated from the traditional spectral information based approach and the geostatistics based classification results are shown in Fig. 2. The kernel and geostatistics based classification results show more homogeneous result in each class region than the result based only on kernel information. The results obtained here were estimated with the consideration of spatial information, in addition to the spectral information. However, there exist classification errors, particularly, on boundary pixels, due to mixed pixel information along the boundaries between neighboring fields, as discussed in Serpico *et al.*(1996).

To quantitatively compare the geostatistics-based classification results with traditional spectral information based classification result, we computed producer's accuracy of each class, overall accuracy and kappa coefficient (Table 1). With respect to all classes, classification accuracies show considerable improvements when one combined the spatial information based on geostatistical indicator kriging with the spectral information based on the kernel method. Improvement is substantial for the carrots class. Combining the spatial information increases the overall accuracy from 0.875 to 0.923, while the Kappa coefficient rises from 0.837 to 0.901. The superiority of the geostatistics based classification over the purely spectral one is clearly demonstrated in this study. Compared to previous research results including the probabilistic neural network and *k*-nn method (Seripico *et al.*, 1996), the overall accuracy and Kappa coefficient show higher classification performances. Especially, our proposed method shows the improvement for the bare soil class, which is important in geological remote sensing. The number of training and test data for bare soil class is the smallest one of all classes. Therefore, this could have resulted in a poor classification performance, if we used only the spectral information (e.g. kernel based method, PNNs, *k*-nn).

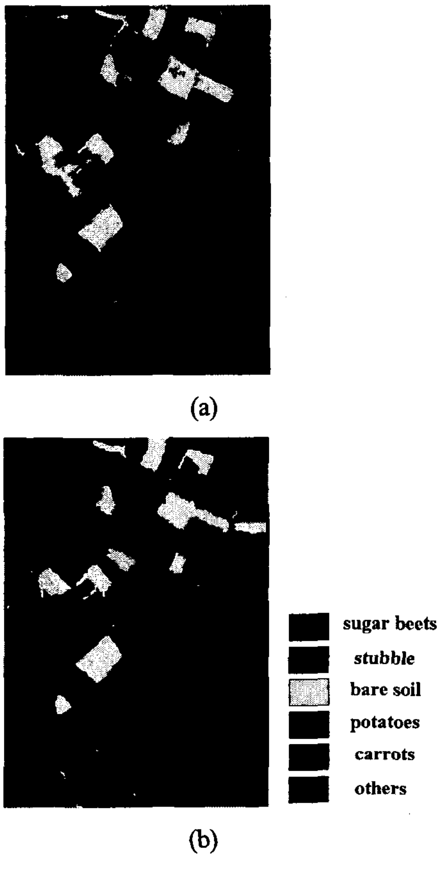

Fig. 2 (a) Kernel based classification result, (b) kernel and geostatistics based classification result.

Table 1. Class-by-class accuracies, overall accuracy, kappa coefficient in the classification of test pixels (PNNs, *k*-nn results are quoted from Seripico et al., (1996)).

Class	Number of pixels	Kernel based	Kernel and geostatistics based	PNNs	*k*-nn
Sugar beets	2043	0.952	0.944	0.978	0.974
Stubble	1371	0.838	0.903	0.824	0.884
Bare soil	555	0.774	0.896	0.796	0.760
Potatoes	884	0.798	0.939	0.818	0.864
Carrots	967	0.817	0.935	0.893	0.871
Overall accuracy		0.861	0.923	0.886	0.898
Kappa coeff.		0.818	0.901	0.850	0.869

4. Discussion and Conclusion

In this paper, we applied the geostatistical indicator kringing to Bayesian data fusion scheme with multi-sensor remote sensing data. Geostatistical indicator kriging provides us with tools for modeling the spatial distribution of categorical variables and estimating probabilities of occurrence of classes based on surrounding observations. Also, the smoothed kernel based non-parametric method is applied to obtain the spectral information. This paper proposed a procedure to combine the geostiatics based spatial information with the smoothed kernel based spectral information. As a result, the proposed method shows considerable improvements in the classification accuracy, compared to other purely spectral information based methods. This paper thus proposes a procedure to combine the geostatistics based spatial information with the smoothed kernel based spectral information.

To strengthen the applicability of proposed schemes, extensive experiments will be applied in several study area and following topics will be included. The performances of the smoothed kernel method depend on the spread parameter, which plays the role of a smoothing parameter. Though the choice of the spread parameter is not main focus of the present work, we need examine the optimization of the spread parameter and influence on overall classification accuracy.

The test area chosen for this study was an agricultural land and consists of homogeneous rectangular land-cover classes.

However, in many geological applications and urban land-use applications, there can exist irregularly shaped land-cover classes (e.g. irregularly exposed rock outcrops and meandering river). In such cases, most natural patterns are inter-related and may have multiple values over a complex non-rectilinear template of locations. Multiple-point geostatistics has been proposed (Strebelle, 2002) for such complex problems. The theory of multiple-point geostatistics builds on the established concepts of traditional geostatistics and can extend into the image analysis and image reconstruction and furthermore carries the potential for new data fusion research.

Acknowledgement

The provision of the IEEE-GRSS Data Fusion Committee test data sets used in this research is gratefully acknowledged. Participation of WMM in this research is partially funded by the NSERC of Canada operating grant A-7400 to WMM.

References

A.H. Solberg, T. Taxt and A.K.Jain, 1996, A Markov random field model for classification of multisource satellite imagery, *IEEE Trans. Geoscience and Remote Sensing*, 34(1):100-113.

B. Solaiman, L.E. Pierce and F.T. Ulaby, 1999, Multisensor data fusion using fuzzy concepts: application to land-cover classification using ERS-1/JERS-1 SAR composites, *IEEE Trans. Geoscience and Remote Sensing*, 37(3):1316-1326.

B.C.K. Tso and P.M. Mather, 1999, Classification of multisource remote 'sensing imagery using a genetic algorithm and Markov random fields, *IEEE Trans. Geoscience and Remote Sensing*, 37(3):1255-1260.

B.W. Silverman, 1986, *Density estimation for statistics and data analysis*, London, Chapman and Hall.

E.H. Isaaks and R.M. Srivastava, 1989, *An introduction to applied geostatistics*, Oxford University Press.

G. Giacinto, F. Roli and L. Bruzzone, 2000, Combination of neural and statistical algorithms for supervised classification of remote-sensing images, *Pattern Recognition Letters*, 21:385-397.

J. Pearl, 1997, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers.

P. Goovaerts, 1997, *Geostatistics for Natural Resources Evaluation*, Oxford Univ. Press, New-York, 483p.

P. Goovaerts, 2002, Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data. *Journal of Geographical Systems*, 4(1):99-111.

R.O. Duda, P.E. Hart and D.G. Stork, 2000, *Pattern classification*, Wiley-Interscience.

S. Strebelle, 2002, Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical Geology*, 34(1):1-22.

S.B. Serpico, L. Bruzzone and F. Roli, 1996, An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images, *Pattern Recognition Letters*, 17:1331-1341.

T. Lee, J.A. Richards and P.H. Swain, 1987, Probabilistic and evidential approaches for multi-source data analysis, *IEEE Trans. Geoscience and Remote Sensing*, 25(3):283-292.

W.M. Moon, 1993, On mathematical representation and integration of multiple spatial geoscience data sets, *Canadian Journal of Remote Sensing*, 19:251-255.