A NEW IMAGERY CLASSIFICATION METHOD USING SPATIAL COVARIANCE INFORMATION

JAMES A SHINE
George Mason University and
U.S. Army Topographic Engineering Center
CEERD-TR-G, 7701 Telegraph Road
Alexandria, VA 22315-3864

DANIEL B. CARR
George Mason University
Fairfax, VA 22030-4444

ABSTRACT: Classical and modern statistical methods offer a wide variety of approaches to classification of data in general and classification of imagery in particular. None of these approaches explicitly use spatial information. Spatial covariance structures have been used for data prediction, but not directly for classification. This paper describes a classification method using the spatial covariance information in imagery to directly classify images in a supervised approach. A series of thresholds are measured with training data for each class, and a model is then fitted. Each pixel is measured for its fit for each class, and the class with the best fit is chosen. A framework is also described for using multiple bands of information and classifying from the combined bands.

INTRODUCTION: Classification in general involves choosing the correct value of y when there are k finite choices, given a number of observations **x,** where **x** is a p-dimensional variable. Classification of imagery in particular involves choosing the correct terrain category for a given square pixel given the input from one or more imagery bands and information from neighboring pixels. Accurate classification of remotely sensed imagery and other spatial data is important for topographic support and other efforts to protect the USA's security structure.

There are many different existing methods for classification. Some of these include discriminant analysis, clustering, neural networks, support vector machines, decision trees, nearest neighbors, and various ensemble approaches such as bagging and boosting. In general, these approaches do not use spatial information in the classification process. Even inherently spatial processes such as PRISM (1) and linear kriging are prediction procedures rather than classification procedures. Two approaches use spatial information in the overall classification process (2,3), but no known classification methods use spatial information to directly classify data.

SPATIAL STATISTICS: Spatial data is often spatially correlated. Points closer together have less variation than points farther apart. Imagery data is almost always spatially correlated. This spatial correlation/covariance information can be empirically quantified by graphs such as the variogram. To plot a variogram, the average variance (squared differences) of a set of points a certain distance apart is plotted against the distance. Generally, the plot will start with low values and gradually the variance will increase over distance until at some point, the variance will level off. A typical variogram obtained from imagery is shown in Figure 1.

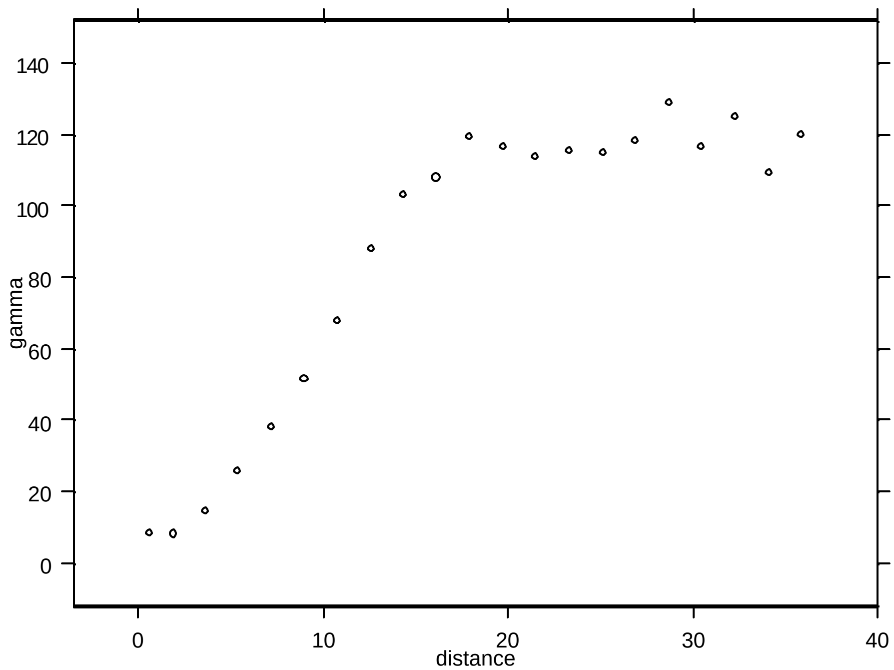

Figure 1: A typical imagery variogram

The variogram must be checked for changes in values based on different directional orientation, known as anisotropy; however, anisotropy is rarely significant in imagery variograms.

Once a variogram is obtained, the spatial covariance structure can be quantified by choosing a variogram model. Usually a model is chosen to minimize a fitting criterion such as sum of squares. Four common variogram models are shown in Figure 2. The equations for these models are:

Nugget model: gamma(h) = c for h > 0

Linear model: gamma(h) = ch for h > 0

Spherical model: $gamma(h) = c\{\frac{3h}{2a} - \frac{h^3}{2a^3}\}, h < a; gamma(h) = c, h \geq a$

Exponential model: gamma(h) = c{1-exp(-h/r)} (

where c are constants that denote the highest values of gamma (the "sill"), and a or r denote the distance where the variogram effectively becomes level (the "range"). Figure 3 shows a spherical model with the sill, c, equal to 300, and the range, a, equal to 60. The exponential model does not have an explicit range since its model is asymptotic, but the effective range is generally considered to be approximately 3r; at this distance, gamma is approximately 95% of the asymptotic sill (4).

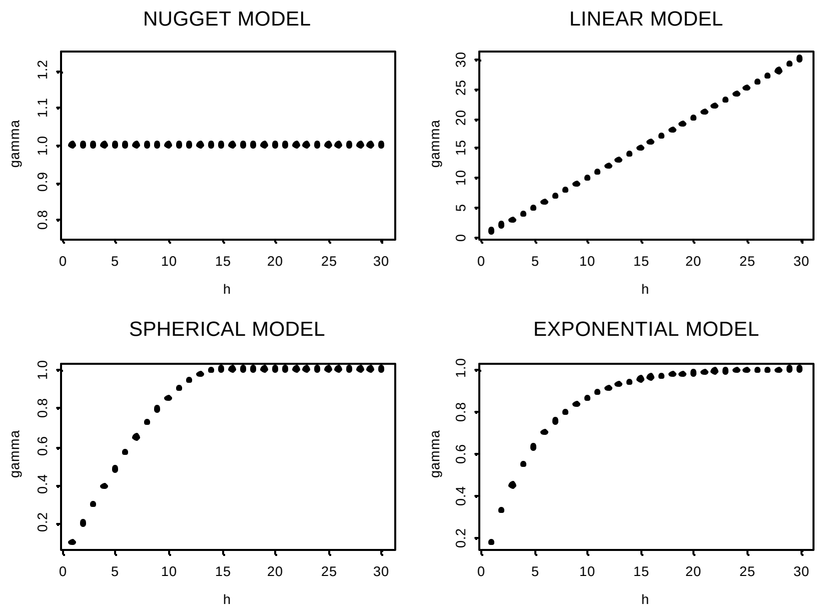

Figure 2. Four common variogram models

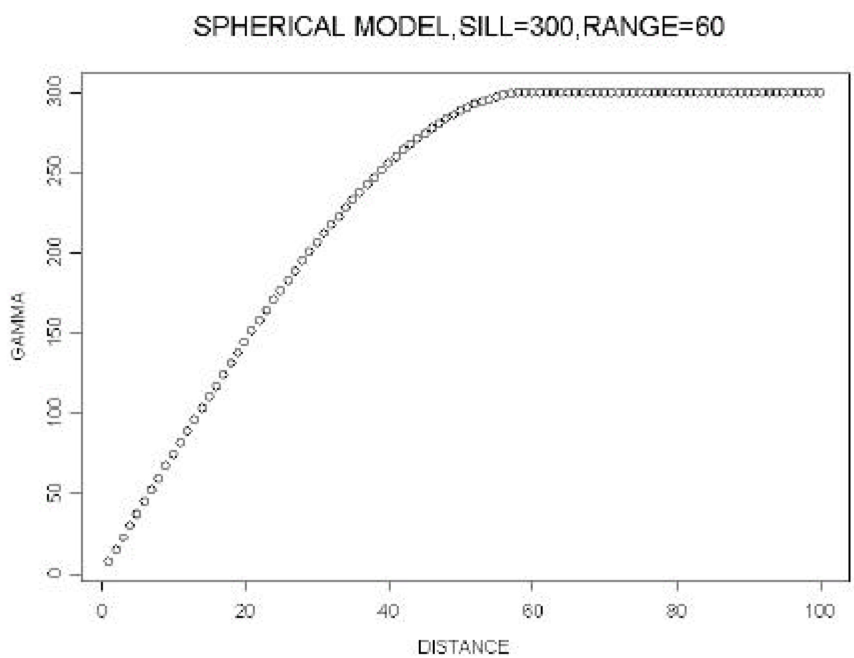

Figure 3: A spherical model with a=60 and c=300

The chosen model can then be used to fit unknown data observations using a weighted average of neighboring observations, with the chosen model specifying the weighting scheme. As mentioned earlier, this approach (sometimes known as kriging) allows prediction of unknown values, but not direct classification.

Variables on an image can also be converted to indicator variables by application of a threshold. For example, if 5 pixel values are 35, 130, 76, 91 and 49, and the threshold is set at 70, the new indicator values will be 0,1,1,1 and 0 respectively. Indicator variables with spatial components may undergo the same variogram-model-predict process just as regular variables.

METHODOLOGY: The two new approaches described in this paper are supervised approaches. The data set is divided into training and testing sets if there are sufficient observations; if not, a bootstrap or cross-validation approach is used.

For the first approach, the spatial coordinates (x and y) and the value of one or more of the data observations (one or more bands in the imagery case) are examined. A total of k classes, which are known, are assumed. For each of the k classes, indicator variograms are obtained at L thresholds. For each of the L thresholds, the data (one band of a remotely sensed image in this case) is converted to indicator data; each pixel becomes 1 if its value is greater than or equal to the threshold, and 0 if it is less than the threshold. The choice of thresholds can be arbitrary, or it can be based on summary statistics such as quantiles obtained from the empirical distribution function. The result is L variograms for each class. For any given class, only those pixels from that class are used to determine the L variograms. For k classes, there will be L x k variograms. For each variogram, a model is plotted. The usual fitting approach minimizes the sum of nonlinear least squares between the theoretical model and the actual variogram values. Using the L variograms, a single model is derived for each of the k classes. The simplest approach is to take an average of the parameters of each of the L models. For example, if L=4 and the result is 4 spherical models with the distance parameter, a, equal to 100,120, 130 and 110 respectively, the single model for the class would be a spherical model as described in Equation 5.4 with a equal to 115. This approach assumes that all L models are of the same form; however, it is possible that one or more of the L models might have a different form. For example, the best fits might come from L-1 spherical models, and 1 exponential model. It is very difficult to combine two separate models. One option would be to make the spherical model an exponential model; another would be to average the spherical models and ignore the exponential model; another would be to combine the 4 variogram data points for each lag and fit an average model directly. Other approaches such as ensemble methods are also worth consideration.

If the models have significantly different objective functions (one threshold fits its model significantly better than others) the model or models with better fits could be given more weight in the decision process.

The single chosen model for each class is then used over the entire data set (image in this case) to create a prediction image for each class. Any given element in the prediction image will take a value between 0 and 1. The process is repeated for each of the k classes, resulting in k prediction images. Each pixel now has k prediction values associated with it, one for each class. Classification is then performed by choosing the largest prediction value and then assigning the pixel the class associated with that value. For example, if k is 6, and performing this process on pixel (27,81) yields prediction image values .3, .7, .4, .5, .2, and .45 respectively, this pixel is classified as category 2.

With multivariate data (such as the four-band multispectral images examined throughout this dissertation) a pixel can be classified as described for each variable or band, and then the multiple decisions can be combined in some form to make an overall decision. A simple approach would be a majority rule, but this might ignore other useful information. For example, certain bands may interact. Also, a prediction might be strong in one band (large difference between the prediction values for the chosen class and the second highest class) and weak in another band (a small difference). Weighting the voting might make sense in such scenarios.

This method is obviously a complex model. First, L x k variograms must be computed and modeled. Next, for each class, L different models must be combined into one single model. Next, k different prediction images must be computed and a classification decision made for each pixel. Finally, for multivariate data, this process is repeated for each variable and then another decision process from multiple inputs is necessary.

For the second approach, only the spatial information is used. Here indicator variables are again created, but for the classes themselves rather than for thresholds. For example, if k=6 and a training pixel or data observation is in class 3, the values for the 6 new indicator class variables will be 0,0,1,0,0 and 0 respectively. A variogram model (only one model) is fitted for each of the k classes. This model and the spatial information from the training pixels are then used to predict each of the test pixels for that class. As in the first method, the class that has the largest value for its prediction is chosen as the class for that pixel. For this method, not only is spatial information used directly in the classification process, it is all that is used in the classification process. While it is a very simple approach, and does not seem to have been studied to date, it would seem that by not using the information in the bands, the accuracy might be diminished.

Another limitation of both of these approaches is that each class must contain enough points to construct a valid variogram model. Generally, this requires a minimum of 30 observations, with 50 to 100 observations more desirable (4).

RESULTS: Work is ongoing to test the accuracy of this approach with different data..

CONCLUSIONS: A new classification method performs classification directly using spatial information. It offers a new tool to be used in classification when spatial data is being processed.

REFERENCES:

(1) Daly, C., Neilson, R.P., and Phillips, D.L., "A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountanious Terrain", Journal of Applied Meteorology, Volume 33, pp. 140-158, 1994.

(2) Lark, R.M., “Geostatistical description of texture on an aerial photograph for disciminating classes of land cover”, International Journal of Remote Sensing, Volume 17, Number 11, pp. 2115-2133, 1996

(3) Carr, J.R., and Miranda, F.P., “The semi-variogram in comparison to the co-occurrence matrix for classification of image texture”, IEEE Transactions on Geoscience and Remote Sensing, Volume 36, Number 6, pp. 1945-1952, 1998.

(4) Oliver, M.A., and Webster, R., “Geostatistics for Environmental Scientists”, John Wiley and Sons, 2001.