IMAGE CLASSIFICATION WITH A REGION BASED APPROACH IN HIGH SPATIAL RESOLUTION IMAGERY

Jong Yeol Lee [a], Timothy A. Warner [b]

[a] Korea Research Institute for Human Settlements, 1591-6, Kwanyang-Dong, Dongan-Gu, Anyang-Shi, Kyounggi-Do, 431-712 Republic of Korea - jylee@krihs.re.kr

[b] West Virginia University, Department of Geology and Geography, PO Box 6300, Morgantown, WV 26506-6300, USA - Tim.Warner@Mail.wvu.edu

KEYWORDS: Image segments, classification, object-oriented approach, multivariate analysis, high spatial resolution

ABSTRACT:

A number of studies have been carried out to find an appropriate spatial resolution to which to aggregate data in order to reduce the variation within an object, and minimize the classification error. Such approaches are pixel-based, and do not draw on the spatial variability as a source of information. The variability within an object can provide additional information that can be used for image classification. Instead of pixels, groups of pixels that form image segments, which are called "patches" in this study, were used for image classification. New methods that exploit multivariate statistics to improve the image classification are suggested. In the case of the object-based classification, patches are not expected to consist of pixels with completely homogeneous spectral radiances, but rather certain levels of variability are expected. To treat this variation within objects, multivariate normal distributions are assumed for every group of pixels in each patch, and multivariate variance-covariance matrices are calculated. A test of this approach was conducted using digital aerial imagery with a nominal one meter pixel size, and four multispectral bands, acquired over the small city of Morgantown, West Virginia, USA. Four classification methods were compared: the pixel-based ISODATA and maximum likelihood approaches, and region based maximum likelihood using patch means and patch probability density functions (pdfs). For region-based approaches, after initial segmentation, image patches were classified into seven classes: Building, Road, Forest, Lawn, Shadowed Vegetation, Water, and Shadow. Classification with ISODATA showed the lowest accuracy, a kappa index of 0.610. The highest accuracy, 0.783, was obtained from classification using the patch pdf. This classification also produced a visually pleasing product, with well-delineated objects and without the distracting salt-and-pepper effect of isolated misclassified pixels. The accuracies of classification with patch mean, and pixel based maximum likelihood were 0.735, 0.687 respectively.

1. INTRODUCTION

Perfect classification could be achieved if each spectral class were to have a unique spectral signature. However, spectral overlap between most real classes occurs as a result of noise in the system, the natural variability of objects within a specific class, and the spatial variability of radiance within each object (Swain and Davis, 1978; Price, 1994). An added complication is that the spectral structure of an image is a function of scale (Cao and Lam, 1997). Higher spatial resolution may actually lead to greater variability within classes, as additional detail is resolved.

A number of studies have been carried out to find an appropriate spatial resolution to which to aggregate data in order to reduce the variation within an object, and minimize the classification error (Pax-Lenney and Woodcock 1997; Teillet *et al.*, 1997; Latty *et al.*, 1985). Such approaches are pixel-based, and do not draw on the spatial variability as a source of information. Another problem with methods that search for an optimal scale is that real objects and classes are variable in size, and thus there is usually no single spatial resolution that suppresses all unwanted spectral variability (Marceau *et al.,* 1994a; 1994b). Studies that use image segmentation to identify single objects (Gougeon, 1995a) can overcome this problem of a single optimal scale. However, most such studies use mainly aggregated information such as average DN, and to a limited extent the variance within the image segments (Kettig and Landgrebe, 1976; Gougeon, 1995a; Meyer *et al.*, 1996). The variability within an object can provide additional information that can be used for image classification. The spectral correlation between bands, as quantified by the covariance matrix, is in fact often a key determinant in traditional maximum likelihood classification for separating classes that overlap in their univariate distributions.

2. TRADITIONAL IMAGE CLASSIFICATION

Maximum likelihood classification is a standard, pixel-based supervised approach, which classifies unknown pixel-based on multivariate probability density functions (pdf) of the classes of interest. Statistical properties of training data sets from ground reference data are typically used to estimate the pdfs of the

classes. Each unknown pixel is assigned to the class with the highest probability at the pixel location. The decision rule is as follows:

$$p(X|\omega_c)p(\omega_c) \ge p(X|\omega_i)p(\omega_i) \qquad (1)$$

where X : the spectral multivariate vector

$p(X|\omega_c)$: pdf of X, given that X is a member of class c

$p(\omega_c)$: *a priori* probability of class c in the image

i: class number among the m number of classes in the image

The resultant likelihoods (D) can be used as surrogates for probabilities.

$$D = [\ln(p(X|\omega_c)p(\varpi_c)]2\pi^{p/2} = \ln(p(\varpi_c) - \frac{1}{2}\ln(|\Sigma|)$$
$$-\frac{1}{2}(X - \mu_i)^T \Sigma^{-1}(X - \mu_i) \qquad (2)$$

Figure 1 shows the pdfs of two spectral classes, with their overlap marked with diagonal lines. The decision rule for this method is that all pixels are assigned to the class with the higher pdf for that spectral value. For example, even if a pixel with the value of "a" belongs in reality to class B, it will be classified as class A (Figure 1). This is an inevitable result of overlapping class pdfs.
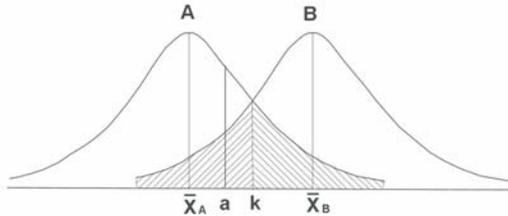
Figure 1. The decision rule of a pixel-based maximum likelihood classifier.

3. METHODS

Instead of pixels, groups of pixels that form image segments were used for image classification in this study. There are few studies that evaluate the use statistic of segmented regions for classification (Kettig and Landgrebe, 1976; Meyer *et al.*, 1996; Gougeon, 1995a; Janssen and Molenaar, 1995). However, most studies employing aggregated information focus on first order statistics and only use second order statistics to a limited extent. In this section, new methods that exploit multivariate statistics to improve the image classification are suggested.

Figure 2 represents a conceptual comparison between traditional classification and the methods developed in this study. An example of the pixel-based approach (Figure 2, left) is the traditional supervised maximum likelihood classification. Within a patch, pixels from the outliers of the class distribution are likely to be misclassified. Window-based approaches use arbitrary groupings and return the value of the window to the central pixel (Figure 2, middle). In the case of the object-based classification (Figure 2, right), patches are not expected to consist of pixels with completely homogeneous spectral radiances, but rather certain levels of variability are expected. This approach, therefore, incorporates a more realistic representation of real phenomena. The variation in an object is used as one characteristic of the object in this method, whereas it is an obstacle with traditional pixel-based classification methods. To treat this variation within objects, multivariate normal distributions were assumed for every group of pixels in each patch, and multivariate variance-covariance matrices were calculated. Two methods of exploiting this information were investigated: maximum likelihood based on the patch mean, and maximum likelihood with Gaussian pdf.
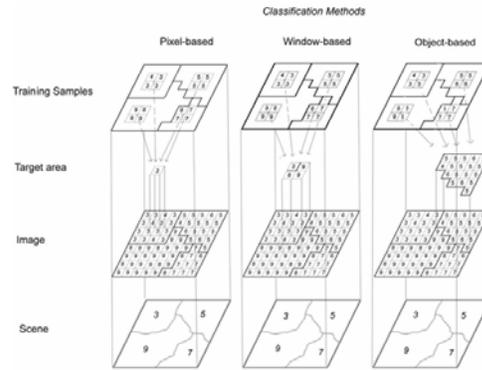
Figure 2. Comparison of object-based classification with traditional image classification approaches.

3-1. Maximum likelihood classification using the patch mean

Maximum likelihood classification with the patch mean uses a decision rule modified to use the mean vector of a group of pixels, instead of individual pixels. When the mean of the group is classified as belonging to a certain class, all the pixels in the group are assigned to that class. The decision rule is as follows:

$$p(\bar{X} | \varpi_c)p(\omega_c) \ge p(\bar{X} | \varpi_i)p(\omega_i) \qquad (3)$$

Where: $\bar{X}$: mean vector of a group

$p(\bar{X}|\omega_c)$: probability associated with the mean of the group of pixels of class c, given that the mean vector $\bar{X}$ is a member of class c

This method classifies each group of pixels as a unit. This will tend to minimize misclassification for isolated pixels with outlier spectral characteristics.

3-2. Region-based maximum likelihood classification with pdf

The method suggested in this study can be summarized as a comparison of the pdf of an unknown group with pdfs of each of the training data sets. If two samples originate from the same population, the pdfs of the two groups should be similar to each other. Significantly, the distribution of radiance values that causes misclassification in pixel-based approaches (Swain and Davis, 1978), is critical information for the method developed in this study.

To simplify the explanation, suppose two normally distributed populations have means μ_1 and μ_2, and standard deviation σ_1 and σ_2, respectively.

Figure 3 represents three different cases that could occur. If two populations are very similar, then the two pdfs almost completely overlap (Figure 3a). If it is possible to estimate the area of the overlapped region, it should be close to 1, because the sum of all possible probabilities is equal to 1. However, if two populations are very different from each other, there should only be a very small overlap area for the two pdfs (Figure 3c). Thus it can be seen that the size of the overlapped area is proportionate to the similarity of the two pdfs. If the two pdfs are identical to each other, the overlapping area is equal to 1, if completely different, then 0, and the values between are an index of similarity (Figure 3b). The area of overlap can be found by integrating the relevant overlap portions of the two pdfs:
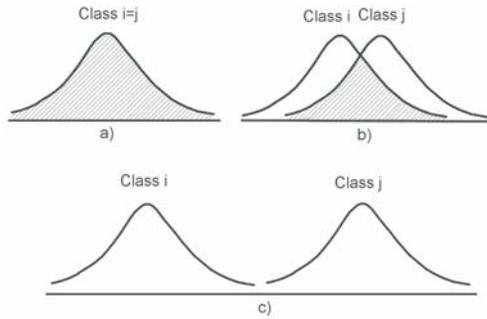

Figure 3. Likelihood measured with pdf. The areas with diagonal lines indicate the degree of similarity between two classes. (a) Two almost completely overlapping class. (b) Two partially overlapping classes. (c) Two almost completely separated classes.

$$O_{12} = \int_{-\infty}^{+\infty} \{ \frac{1}{\sqrt{2\pi}\sigma_m} \exp[\frac{-(X_m - \mu_m)^2}{2\sigma_m^2}] \} dX_m \qquad (4)$$

Where: O_{12} : likelihood index between X_1 and X_2

$m = 2$ for $X_1 \geq X_2$

$m = 1$ for $X_1 < X_2$

When the likelihood index is extended to a multivariate pdf, with p variables and multiple samples, the equation is modified as follows:

$$O_{ij} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ..... \int_{-\infty}^{+\infty} \{ \frac{1}{2\pi^{p/2}|\Sigma_m|^{1/2}} \exp[-\frac{1}{2}(X_m$$

$$- \bar{X}_m)^t \Sigma_m^{-1}(X_m - \bar{X}_m)]\} dX_1........dX_p$$

$$(5)$$

Where O_{ij} : likelihood index between X_i and X_j

i : patch id under investigation

j : training data set id under investigation

$m = j$ for $X_i \geq X_j$

$m = i$ for $X_i < X_j$

The decision rule in this study is extracted from the relationship between the likelihood and similarity as follows:

$$O_{ic} \geq O_{ij} \qquad (6)$$

Thus a patch is assigned to class c if the maximum of the likelihood index values is found for the pdf comparison of the patch and training data set C. With this method, a very stable similarity index is obtained because the variance and covariance information, as well as the class mean, are all directly used. One disadvantage of this method is that it requires much computing time.

4. APPLICATION OF THE PIXEL AND REGION-BASED CLASSIFICATION METHODS

The ADAR data is used to compare the new approach with traditional methods. The system captures images 1,000 by 1,500 pixels in size, each pixel approximately 1 x 1 meter. The ADAR system acquires four bands of data with four separate digital cameras sensitive to blue, green, red, and infrared wavelengths covering the range from 400 to 1,000 nm.

The data were acquired at 19:42:18 GMT (2:42 pm local time) on March 24 1997 (early spring, prior to tree leaf-out) from an altitude of 2,522 meters. Leaf-off data provides clearer observation of ground features, but less spectral discrimination of forest cover. The analysis procedure in this study comprises three stages. It is assumed that patches have previously been identified by image segmentation using the region growing process incorporating thresholding and region growing. In the first stage, statistics for the patches are computed. The statistics used were the same as those used in the region growing stage, including the mean vectors and variance-covariance matrices.

For the second stage, representative patches were selected to build a training data set for seven classes: Building, Road, Forest, Lawn, Shadowed Vegetation, Water, and Shadow. The patches selected as training

data were treated as independent spectral classes within each informational class. This means that the selected patches were not aggregated into composite statistics for the seven classes. The likelihood index for each patch was computed for all individual training data set patches by the divergence index, the maximum likelihood using the patch mean, and the patch pdf.

Small patches with fewer than six pixels were excluded from the region-based maximum likelihood analysis, and treated as part of the "melt pond," to use McDevitt and Peddada's (1998) term. There were two reasons for identifying melting pond pixels. Firstly, because five variables were used in this study, patches with fewer than six pixels had less than the minimum number of pixels potentially required to characterize the multivariate statistics. Secondly, the melting pond was assumed to represent objects that are not of direct interest, but rather extraneous objects such as cars, or chimneys on buildings.

In the third stage, each patch was classified into seven classes by the suggested methods. For maximum likelihood with patch pdf, the range over which the pdf was calculated was limited to three standard deviations. The pdf is very low outside of this range, and is not expected to have much significance in the calculation. Excluding pdf values greater than three standard deviations has the advantage of reducing the computing cost.

Figure 4 shows a one dimensional representation of the process. Within the pdf overlap region, the decision range was divided into ten equal cells. The probability of the center of each cell calculated for both the training and the patch classes, and the lower of the two probabilities is used for the cell height. After multiplying cell height by the width, the cell area is calculated. The total area of the overlap is then estimated by summing the cell areas (Figure 4). This procedure is modified for the multivariate case by dividing the multidimensional overlap region into 10^n cells, where n is the number of bands. For two bands a volume of the overlap region is calculated, and for three or more bands a hypervolume is calculated. For this work, five bands were used, thus, 10^5 cells were calculated for each likelihood index.
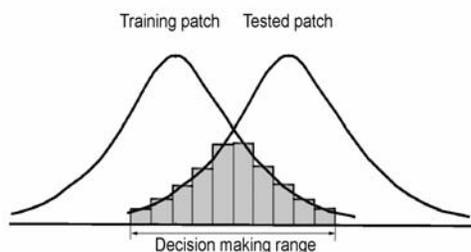

Figure 4. Maximum likelihood calculation utilizing patch pdfs.

The patch was assigned to the class with the highest likelihood after the unknown patch is compared with each patch in the training data set. In the next step of the classification, melting pond pixels are classified. These small patches are treated as noise, and therefore assigned to an adjacent class. If the patch is surrounded by a single class, it is assigned to that class. In the general case, however, the patch is adjacent to more than one class. In this case, the patch is assigned to the adjacent class with the most similar DN values in the green band (Band 2). A more sophisticated, multivariate approach was not used because of the small sample size of these patches. In the final step, adjacent patches of the same class were merged to form objects.

ERDAS Imagine was used to conduct the traditional pixel-based classifications. The unsupervised ISODATA program (Tou and Gonzalez, 1974; ERDAS, 1999) was executed with 24 clusters. After classification, the 24 clusters were assigned empirically to the most appropriate class among the seven classes based on the ground truth and knowledge of the area. For each of the supervised classification methods, the same training data sets were used.

5. RESULTS AND DISCUSSION

Figure 5 shows the results from the four previously mentioned methods. To compare the accuracy of the four methods, error matrices for the kappa index and errors of commission and omission were produced using the IDIRSI program ERRMAT (Eastman, 2003) (Table 1). Ground reference maps for the accuracy evaluation were produced using photo-interpretation and expert knowledge for three parts of the study area: Downtown Morgantown, a medium density residential area, and a forested stream valley.

Table 1. Summary accuracy statistics for 7 classes by the 4 classification methods used in this study.

		Blding	Road	Forest	Lawn
ISODATA	CERR	0.363	0.489	0.135	0.440
	OERR	0.487	0.320	0.147	0.046
MHL with pixel	CERR	0.391	0.347	0.063	0.366
	OERR	0.202	0.412	0.133	0.214
MLH with patch mean	CERR	0.309	0.291	0.104	0.304
	OERR	0.194	0.292	0.042	0.503
MLH with patch pdf	CERR	0.225	0.243	0.101	0.294
	OERR	0.170	0.249	0.045	0.480

		Shd Veg	Water	Shadw	Kappa
ISODATA	CERR	0.567	0.024	0.062	0.610
	OERR	0.373	0.981	0.352	
MHL with pixel	CERR	0.440	0.123	0.137	0.687
	OERR	0.297	0.308	0.359	
MLH with patch mean	CERR	0.360	0.095	0.048	0.735
	OERR	0.691	0.232	0.398	
MLH with patch pdf	CERR	0.306	0.042	0.054	0.783
	OERR	0.603	0.089	0.234	

The overall kappa value of the supervised pixel-based classifications was 0.687. The lowest accuracy, 0.610, was obtained with the unsupervised pixel-based
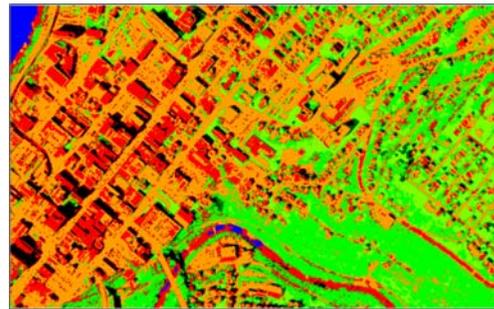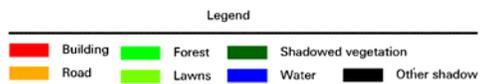
classifier ISODATA. The maximum likelihood classifier using the patch mean resulted in a relatively high kappa value of 0.735. Maximum likelihood classifier with pdf produced the overall best accuracy, 0.783.

Looking at the results in more detail, the unsupervised classifier resulted in many isolated pixels and small clusters, as expected (Figure 5 a). The Water class in the region of the stream was almost completely misclassified as Building with this method. The stream has exposed and shallow covered rock that is apparently spectrally similar to the materials from which buildings are constructed. Building was also misclassified as Road, and consequently the Building omission error was relatively high (Table 1). Pixel-based supervised classification (Figure 5 b), like the unsupervised classification, resulted in a rather noisy classification. The classes of Buildings and Roads were extensively confused, resulting in high errors of commission and omission for both classes. However, compared to the unsupervised classification, the confusion between Building and Water was dramatically reduced for the pixel-based maximum likelihood classification.
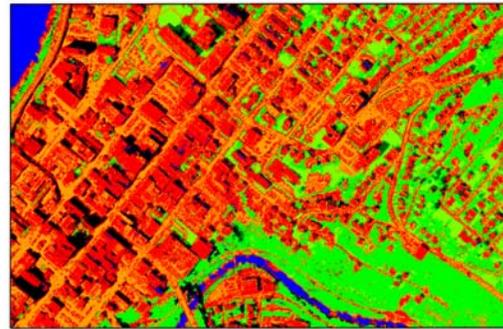
The maximum likelihood classifier using the patch mean (Figure 5 c) yielded a visually pleasing classification, and the second best overall accuracy. The higher classification accuracy of the maximum likelihood classification with patch pdf is most likely a result of the incorporation of differences in the kurtosis of classes through the variance-covariance matrix data. When only the patch mean is used in the classification, such differences are suppressed. The particular classes that were less well classified in the maximum likelihood using the patch mean, compared to the patch pdf, were the Building and Road classes. But the computing cost for classification with the mean was much lower than with the pdf. Thus, the classifier with the patch mean is an efficient alternative to classification with pdf.

The maximum likelihood classification with pdf produced higher accuracy than any other classifier (Table 1). The segmentation suppresses isolated pixels and small clusters (Figure 5 d), and thus classification error resulting from high within object variance was efficiently controlled by this method.
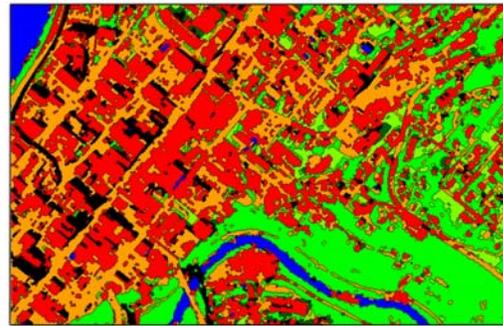
However, a number of cases of confusion arose between Building and Road, and Lawn and Forest. The confusion between Lawn and Forest can be related to segmentation. Although these two classes generally had sufficient spectral difference between them for good classification, in some cases the low

Legend

| | Building | | Forest | | Shadowed vegetation |
| | Road | | Lawns | | Water | | Other shadow |

Figure 5. Results of the classifications. (Above) Legend. (Right) a): ISODATA from ERDAS Imagine. b): Maximum likelihood classification from ERDAS IMAGINE. c): Maximum likelihood classifier with patch mean. d): Maximum likelihood classifier with patch pdf.

a)

b)

c)

d)

contrast boundaries between Lawns and Forest areas resulted in these regions being merged into a single patch. The confusion between Building and Road was not a result of segmentation as generally these two classes were well delineated. However, confusion occurred because the spectral radiances of the two classes were sometimes very similar. This arises because materials such as asphalt, stone and concrete are used for both building roofs and roads.

As part of the classifications carried out using maximum likelihood, all pixels were assigned to the class with the highest likelihood. This is a relative, not an absolute measure. Thus even classes that result in very low likelihood when compared to all the training data sets are classified. It is possible that a region is not represented by any of the training data sets, and this should be identified. In future work, it may be desirable to establish an absolute minimum maximum likelihood for classification. Patches that fail to meet the minimum value would be flagged as unknown.

6. CONCLUIONS

This study produced a region-based classification approach specifically designed for high spatial resolution imagery. The new classification method resulted in improved results at both the image object scale and a richer attribution at the aggregate land cover scale. This research made a contribution to the growing field of analysis of high spatial resolution imagery.

The methods developed in this research are important not just because they produce more accurate results that show the spatial patterns more clearly because of their lack of distracting high frequency noise. The delineation and attribution of image objects, rather than classified pixels, is an important step toward integrating remote sensing with GIS. The object-based approach resulted in a pleasing simplicity of spatial structure compared to the noisy patterns of traditional pixel-based classification.

Acknowledgements

References

Cao, C., and N. S. Lam, 1997. Understanding the scale and resolution effects in remote sensing and GIS. In: Scale in Remote Sensing and GIS (D. A. Quattrochi, and M. F. Goodchild, editors), Lewis Publishers, New York, NY, pp. 57-72.

Eastman, J., R., 2003. IDRISI Kilimanjaro Guide to GIS and Image Processsing, Clark University, Worcester, MA.

ERDAS, 1999. ERDAS Field Guide, ERDAS, Atlanta, Georgia, 672 p.

Gougeon, F. A., 1995. Comparison of possible multispectral classification schemes for tree crowns individually delineated on high spatial resolution MEIS images. *Canadian Journal of Remote Sensing* 21(1): 1-9.

Janssen, L. L. F., and M. Molenaar, 1995. Terrain objects, their dynamics and their mornitoring by the integration of GIS and Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing* 33 (3): 749-758.

Jensen, J. R., 1996. Introductory Digital Image Processing, Prince Hall, Upper Saddle River, New Jersey. York, NY, pp. 3-78.

Kettig, R. L., and D. A. Landgrebe, 1976. Classsification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics* GE-14(1): 19-26.

Latty, R. S., R. Nelson, B. Markham, D. Williams, D. Toll, and J. Irons, 1985. Performance comparisons between information extraction techniques using variable spatial resolution data. *Photogrammetric Engineering and Remote Sensing* 51 (9): 1495-1470.

Lillesand, T. M., and R. W. Kieffer, 1994. Remote sensing and image interpretation, John Wiley & Sons, Inc. New York, NY, pp. 169-178.

Marceau, D. J., D. J. Gratton, R. A. Fournier, and J. Fortin, 1994a. Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. *Remote Sensing of Environment* 49: 93-104.

Marceau, D. J., D. J. Gratton, R. A. Fournier, and J. Fortin, 1994b. Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment* 49: 105-117.

Mcdevitt, R. J., and S. D. Peddada, 1998. An automated algorithm for cleaning and ordering the boundary points of a one-dimensional curve in a segmented image. *IEEE Transactions on Geoscience and Remote Sensing* 36 (1): 307-312.

Meyer, P., K. Staenz, and K. I. Itten, 1996. Semi-automated procedures for tree species identification in high spatial resolution data from digitized colour infrared-aerial photography. *ISPRS Journal of Photogrammetry & Remote Sensing* 51: 5-16.

Pax-Lenney, M., and C. E. Woodcock, 1997. The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote sensing of Environment* 61: 210-220.

Price, J. C., 1994. How unique are spectral signatures? *Remote Sensing of Environment* 49: 181-186.

Swain, P. and S. Davis, 1978. Remote Sensing: The Quantitative Approach, McGraw Hill, N.Y. 369 p.

Teillet, P. M., K. Staenz, and D. J. Williams, 1997. Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. *Remote Sensing of Environment* 61: 139-149.

Tou, J. T. and R. C. Gonzalez, 1974. Pattern Recognition Principles, Addison-Wesley Publishing Co, Reading, MA.