FEATURE SELECTION METHODS FOR OBJECT-BASED CLASSIFICATION OF SUB-DECIMETER RESOLUTION DIGITAL AERIAL IMAGERY

A. S. Laliberte [a], D.M. Browning [b], A. Rango [b]

[a]Jornada Experimental Range, New Mexico State University, Las Cruces, NM 88003, USA – alaliber@nmsu.edu
[b]USDA-Agricultural Research Service, Jornada Experimental Range, Las Cruces, NM 88003, USA – (dbrownin, alrango)@nmsu.edu

KEY WORDS: object-based image analysis (OBIA), feature selection, high resolution, aerial photography, vegetation, classification, accuracy

ABSTRACT:

The availability of numerous spectral, spatial, and contextual features renders the selection of optimal features a time consuming and subjective process in object-based image analysis (OBIA). While several feature selection methods have been used in conjunction with OBIA, a robust comparison of the utility and efficiency of approaches could facilitate broader application. In this study, we tested three feature selection methods, 1) Jeffreys-Matusita distance (JM), 2) classification tree analysis (CTA), and 3) feature space optimization (FSO) for object-based classifications of rangeland vegetation with sub-decimeter digital aerial imagery in the arid southwestern U.S. We assessed strengths, weaknesses, and best uses for each approach using the criteria of ease of use, ability to rank and/or reduce input features, and classification accuracies. For the five sites tested, JM resulted in the highest overall classification accuracies for three sites, while CTA was highest for two sites. FSO resulted in the lowest accuracies. CTA offered ease of use and ability to rank and reduce features, while JM had the advantage of assessing class separation distances. FSO allowed for determining features relatively quickly, because it operates within the eCognition software used in this analysis. However, the feature ranking in FSO is unclear and accuracies were relatively low. While all methods offered an objective approach for determining suitable features for classifications of sub-decimeter resolution aerial imagery, we concluded that CTA was best suited for this particular dataset. We explore the limitations, assumptions, and appropriate uses for this and other datasets.

1. INTRODUCTION

The selection of appropriate spectral bands or image features is a crucial step in any image analysis process. Using a set of optimal features ensures that the classes in question are discriminated effectively and with sufficiently high accuracy, and that the dimensionality is reduced for efficient use of training samples (Jensen, 2005). In object-based image analysis (OBIA), the determination of optimal features can be a time consuming process due to the availability of numerous spectral, spatial, and contextual features. Feature selection techniques range from graphic methods to statistical approaches involving class separation distances. Several feature selection methods have been used in conjunction with OBIA. Herold et al. (2003) and Carleer and Wolfe (2006) used the Bhattacharyya distance, while Nussbaum et al. (2006) and Marpu et al. (2006) employed the Jeffreys-Matusita distance for feature selection. Johansen et al. (2009) evaluated feature space plots, box plots, band histograms, and feature space optimization. Classification tree analysis for selection of optimal features was successfully applied by Chubey et al., (2006), Yu et al., (2006), Laliberte et al. (2007), and Addink et al. (2010).

The above mentioned studies all used high resolution satellite images (QuickBird, Ikonos, SPOT) or aerial photography (0.3-1.25m resolution). In recent years, the use of digital mapping cameras has greatly increased, and examples for use of these images include mapping benthic habitats (Green and Lopez, 2007), land use/land cover mapping (Rosso et al. 2008), and border monitoring and change detection (Coulter and Stow, 2008). Digital airborne imagery can be acquired at sub-decimeter resolution and exhibits great potential for mapping rangeland vegetation mapping (Laliberte et al., in press) despite multiple challenges, such as high spatial frequency, the effect of shadows, viewing geometry, illumination, and the necessity for mosaicking multiple images for analysis. Optimal features for classification may be scale dependent, and features used in the analysis of coarser resolution imagery may not be applicable to finer resolution data. Determination of appropriate features for very high resolution imagery, and a robust comparison of the utility and efficiency of various feature selection methods could facilitate broader use of sub-decimeter aerial imagery for vegetation mapping.

The objectives of this study were to 1) determine the optimal features for fine-scale vegetation mapping, and 2) evaluate three feature selection methods (i.e., Jeffreys-Matusita distance, classification tree analysis, and feature space optimization), in the context of object-based classification of rangeland vegetation with digital aerial imagery with a 6 cm ground resolved distance. Evaluation criteria for the feature selection methods included efficiency and ease of use, ability to rank and reduce features, and classification accuracies.

2. METHODS

Study area and images

The study sites were located at the Jornada Experimental Range and the Chihuahuan Desert Rangeland Research Center in southwestern New Mexico, USA (32º34’11”W, 106º49’44”N). Average elevation is about 1200 m, and rainfall amounts and distribution are highly variable, with more than 50% of the mean annual precipitation of 241 mm occurring during July, August, and September. Much of the historic semi-desert grassland has experienced marked increase in shrub abundance and distribution, although some grass-dominated areas remain. For this study, imagery was acquired over five 150 m x 150 m plots in five vegetation communities, three shrub-dominated: creosote (CGRAV), tarbush (TEAST), mesquite (MWELL), and three grass-dominated: grassland (GIBPE), and playa (PCOLL).

The images were acquired on 13 September, 2009 with an UltraCamL large format digital mapping camera at a flying height of approximately 480 m above ground. The camera acquires multispectral data in the red (580-700 nm), green (480-630 nm), blue (410-570 nm), and near infrared (690-1000 nm) bands. Five overlapping images were acquired over each site, orthorectified using interior and exterior orientation values, and mosaicked using Leica Photogrammetric Suite (Erdas, 2009). The image mosaic (6 cm ground resolved distance) was subsequently co-registered to an orthorectified QuickBird image using Erdas Imagine AutoSync™ with an average of 230 automatically generated tie points and an average RMS error of 5 cm. Images were clipped to the plot boundaries for further analysis.

Field Data

At each of the five plots, training and test samples for the dominant vegetation species were collected by digitizing vegetation boundaries over the displayed image using ArcPad on a tablet PC. This method eliminated potential errors associated with GPS data collection due to the discrepancy between image resolution and GPS error. On average, 520 species-level samples were collected per plot (at least 20 samples per class), and classes/plot ranged from 6 to 10. Half of the samples were used as training sites, and half were retained for accuracy assessment.

Image Analysis

We used eCognition Developer 8 (Definiens, 2009) for the object-based image analysis. The images were segmented at two scales, a fine scale multiresolution segmentation with scale parameter 100, and a coarser scale spectral difference segmentation with a maximum spectral difference of 1500. All bands were weighted equally for the segmentation; color/shape was set to 0.9/0.1, and smoothness/compactness was set to 0.5/0.5. All classifications were done at the coarser segmentation scale. A rule-based approach was used to classify shadow, bare ground, and vegetation, followed by a nearest-neighbor classification for detailed species-level classes using training samples. We tested the three feature selection methods for the species-level classification only.

We started the analysis with 31 spectral, spatial, and texture features. Spearman's rank correlation analysis was used to eliminate features with correlation coefficients >0.9 to reduce data dimensionality. As a result, 12-18 features remained per plot (Table 1). Three feature selection methods were tested, 1) Jeffreys-Matusita distance (JM), 2), classification tree analysis (CTA), and 3) feature space optimization (FSO). Inputs for all three methods consisted of the feature values or object statistics of the training objects for each class. For JM, we used the SEaTH tool (Nussbaum et al., 2006) due to its compatibility with eCognition's exported object statistics. SEaTH calculates the class separability for every feature and two-class combination and outputs individual text files for each two-class comparison. The data were compiled in Excel to determine the largest average JM distance for every possible 4-10 feature combination. We then selected the feature combination that resulted in the largest JM distance for the least separable pair of classes. For CTA, we used CART® (Salford Systems) which produces a decision tree and optimum features that are ranked based on variable importance scores of the primary splitters in the tree. FSO is a tool available in eCognition and it calculates an optimum feature combination based on class samples. FSO evaluates the Euclidean distance in feature space between the samples of all classes and selects a feature combination resulting in the best class separation distance, which is defined as the largest of the minimum distances between the least separable classes. We assessed classification accuracies for the species-level classes only by creating an error matrix to determine users, producers, and overall classification accuracies, and Kappa statistics (Congalton and Green, 1999). Statistical differences between classifications were assessed with McNemar's test, a non-parametric test of contingency tables (Foody, 2004).

3. RESULTS

3.1 Feature Selection

The features selected with the three methods tested are shown in Table 1. The lowest number of features was selected by CTA in three of the five plots. FSO consistently selected the largest number of features, in two plots twice as many features as JM, and in two plots three times as many features as CART, one noted exception was with the CGRAV plot. We observed more consistency in feature selection between JM and CTA than between FSO and either of the other two methods. For example, in CGRAV, FSO did not select NDVI or Mean NIR, while both features were selected by JM and CTA. A similar observation was apparent with the ranking of features, which was similar for JM and CTA, and less similar for FSO compared to JM or CTA. For JM, the ranking is based on the average JM distance of all two-class comparisons for a particular feature space. For CTA, the ranking was obtained by using the variable importance scores of the primary splitters in the decision tree. The FSO ranking is based on the order of selection within the FSO tool. Overall, spectral features were more likely to be selected than spatial or texture features, and on average the highest ranking features were Mean Red, NDVI, and Ratio Red.

3.2 Classification Accuracy

The highest overall classification accuracies and Kappa values were obtained using JM for three plots, followed by CTA for two plots. Classifications with features from FSO had the lowest overall accuracies and Kappa values (Figure 1). P-values from McNemar's tests for all two-method comparisons were<0.001 except for the CTA vs. FSO comparisons in MWELL (p=0.052) and GIBPE (p=0.029). In general, Kappa values for individual classes showed more consistency across methods if the class was highly separable from other classes (based on the JM distances). Conversely, less separable classes had greater variation in Kappa values for the three methods. An example is shown in Figure 2 with the plot MWELL. The species ATCA (Fourwing saltbush) and YUEL (Yucca), which were confused with each other, showed a large variation in Kappa values for the methods, while highly separable species such as GUSA (Broom snakeweed) and PRGL (Honey mesquite) had comparable Kappa values for all methods. However, there were exceptions to this observation related to the size class of certain shrubs. While large PRGL (avg. diameter>0.6m) were highly separable from other species (as in

Plots:	**CGRAV**			**MWELL**			**TEAST**			**GIBPE**			**PCOLL**		
Method:	**JM**	**CTA**	**FSO**	**JM**	**CTA**	**FSO**	**JM**	**CTA**	**FSO**	**JM**	**CTA**	**FSO**	**JM**	**CTA**	**FSO**
Features															
NDVI	2	2		1	1	7	3	2	11	2		13	2	3	14
Mean Blue															
Mean Green				6	5	12									
Mean NIR	4	4			3		6		16	6	3	10	6		13
Mean Red	1	1	6	3	8	10	2	1	3	4	1	6	5	1	
Max difference					6	13	8		13	7	4	15	7		6
Standard deviation Blue															
Standard deviation Green						1									
Standard deviation NIR			4			4			6			1			2
Standard deviation Red		6	2						15			5			9
Ratio Blue	3	5				2	7	5	1			4			
Ratio Green				7	2	5		4	7			11	3		1
Ratio NIR							4		8						
Ratio Red							5		2	1	2		1	2	8
Mean Diff. to neighb. Blue													8	4	5
Mean Diff. to neighb. Green				5	9	6									
Mean Diff. to neighb. NIR			1										9		11
Mean Diff. to neighb. Red			7						10		5	8			
Area				2	4		1	3	5	3	6		4		
Compactness												14			7
Density			3			8			4			2			3
Roundness					7	9			9						
Shape index															
GLCM Homogeneity						11			14			7			12
GLCM Contrast												9			
GLCM Dissimilarity									12						10
GLCM Entropy	7		5												15
GLCM StdDev															
GLCM Correlation						3						3			4
GLCM Ang. 2nd moment	5	7								5		12	10		
GLCM Mean	6	3		4											
Total uncorrelated		12			16			17			17			18	
Total selected	7	7	7	7	9	13	8	5	16	7	6	15	10	4	15

Table 1. Features selected from 31 input features using three feature selection methods (JM, CTA, FSO) for five plots. Numbers in columns represent feature ranks for JM by largest JM distance, for CTA by primary splitter in decision tree, and for FSO by order of selection in the FSO tool. Uncorrelated features per plot have correlation coefficients <0.9.

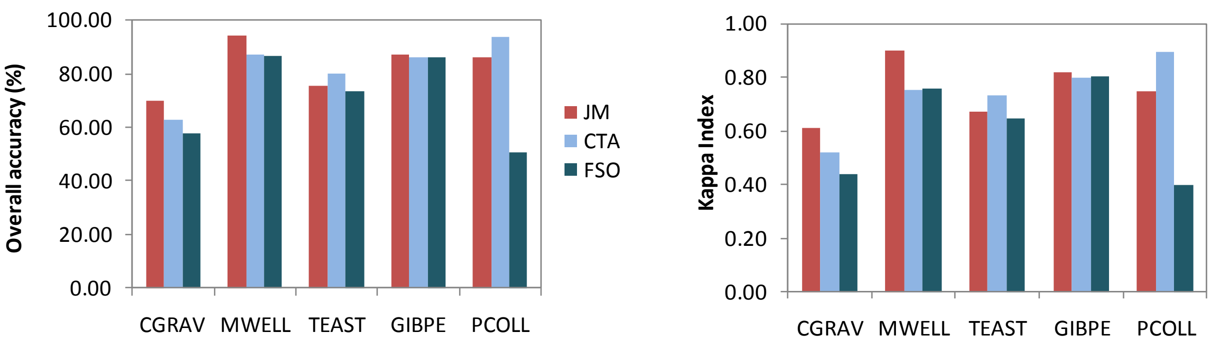

Figure 1. Overall classification accuracies (left) and Kappa indices (right) for classifications of five plots with three feature selection methods (JM, CTA, FSO). P-values from McNemar's tests for all two-method comparisons were <0.001 except for the CTA vs. FSO comparisons in MWELL (p=0.052) and GIBPE (p=0.029).

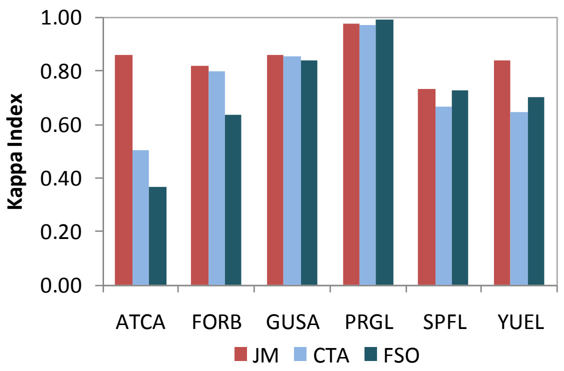

Figure 2. Kappa indices for individual classes (four-letter symbols) in plot MWELL for three classification methods (JM, CTA, FSO).

plot MWELL), small PRGL (avg. diameter<0.6m) were often confused with other small shrubs due to their higher reflectance. This resulted in Kappa values around 0.5 for PRGL in the plot CGRAV (not shown). An example classification of plot TEAST with features determined with the CTA method shows the fine detail acquired with this imagery (Figure 3).

3.3 Processing Time

The three methods for feature selection resulted in a range in processing times due to the different steps required for each method. In all plots, CTA was faster than JM, which was faster than FSO for obtaining an optimum feature combination. However, FSO was entirely dependent on how many texture features were included, because texture determination is CPU demanding. In this analysis, FSO chose between one and four texture features for each plot, slowing down the operation. For example, feature space optimization took on average 10 seconds for a test using 10 features without texture. With four texture features, this task required up to 2 hours. If no texture features are included, FSO has the potential to be the fastest feature selection method.

Both CTA and JM required exporting the object statistics from eCognition and the use of another program for determining feature parameters. CTA was more time efficient, because CART® processing times from import to outputs of results required only seconds. Feature rankings based on variable importance scores of the primary splitters in the tree were easy to interpret. JM required the most data manipulation, because several steps were required for analyzing object statistics using the SEaTH tool, compiling individual output files for each two-class combination (up to 45), and obtaining average separation distances for every possible 4-10 feature combination before determining the optimum combination for the least separable pair of classes. Classification times in eCognition were also highly dependent on the number of texture features included. Plots without texture features were classified in minutes, while plots with four texture features took up to 9 hours to classify.

4. DISCUSSION

In order to assess the suitability of a particular feature selection approach, classification accuracy is an important criteria, because ultimately the analyst desires a result with as high accuracy as possible. However, other criteria have to be considered as well. If multiple images have to be analyzed, ease of use,

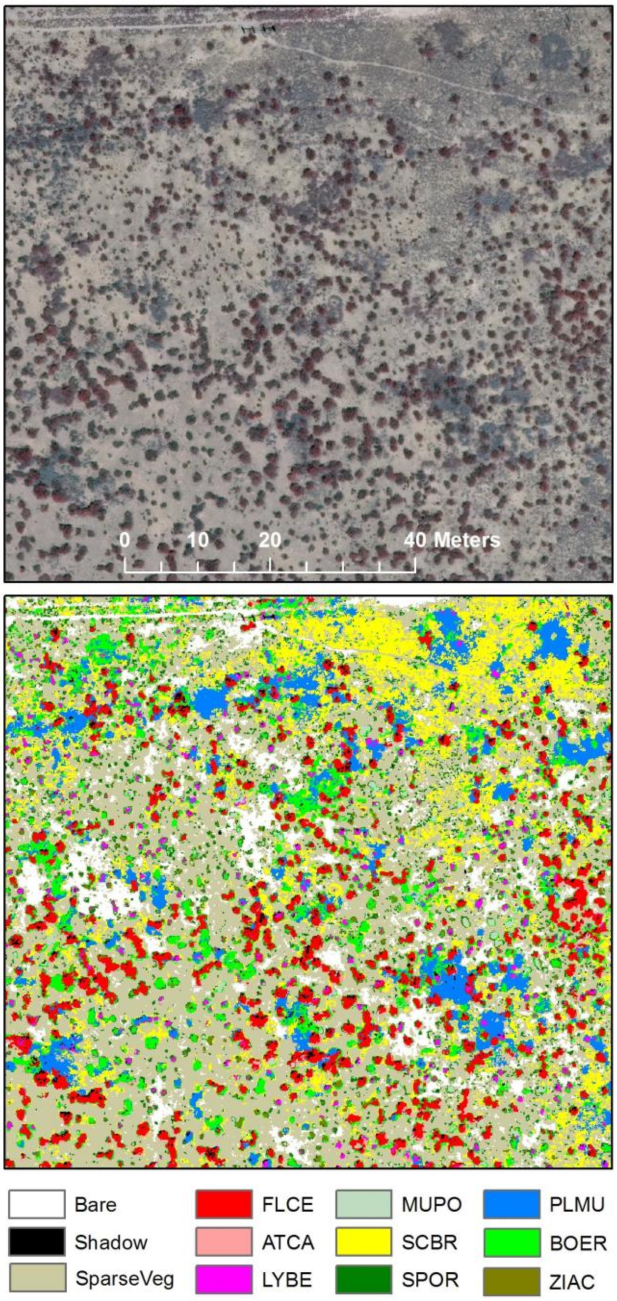

Figure 3. UltraCam X image with 6 cm ground resolved distance (top) and classification (bottom) for a portion of plot TEAST. Classification was performed using CTA. FLCE, ATCA, and LYBE are shrubs (four-letter symbols), ZIAC is a forb, and the remaining classes are grasses.

efficiency, and processing times are equally important, and an efficient workflow may be preferable to higher overall accuracy, especially if the difference in accuracy consists of a few percentage points. The ability to clearly interpret the results of a feature selection method also has to be taken into consideration. Finally, a robust feature selection technique has to be capable of ranking and reducing a potentially large number of input features. In most situations, training samples are costly and therefore limited. Using a reduced number of features for a given set of training samples reduces dimensionality and prevents what is known as the Hughes phenomenon, the deterioration of classification accuracy due to addition of unnecessary features (Kim and Landgrebe, 1991).

Each feature selection method had advantages and disadvantages, and while we determined that CTA was best suited given the mapping objectives, image type and resolution in this study, we acknowledge that either of the other methods may be preferable in different studies. CTA proved to be an excellent feature reduction and ranking tool, it required few steps, and the results were easy to interpret. Moreover, a decision tree is also non-parametric. While in some datasets the assumption of normality underlying the JM distance may not be met due to limited sample sizes, this method has the advantage that class separation distances for specific classes can be determined. In this study, class separation distances offered new insights into the influence of size classes of certain species. While larger mesquite shrubs were clearly separable from other shrubs, small mesquites had a tendency to be confused with other shrubs. In terms of efficiency, if many classes and/or many features are involved, JM can be time-consuming, because the analysis is essentially based on multiple two-class comparisons.

FSO was the lowest ranking method in this study. While it can be applied directly in eCognition and can potentially be the fastest approach without texture features, it consistently selected the largest number of features of all three methods. In addition, the feature selection approach appeared to be a “black box”, and it provided the lowest classification accuracies. However, FSO does provide class separation distances based on the Euclidean distance in feature space between the samples of all classes. Although we did not select features based on class separation distances from FSO, we compared the separation distances obtained with FSO and JM for all two-class combinations for each plot. The correlations for the two measures were relatively high and all were significant at the 0.05 alpha level (Table 2). While the maximum separation distances from FSO varied greatly from plot to plot, JM distances have the advantage of being scaled from 0-2, which makes is convenient for comparing separability of the same classes in different plots.

Plot	Correlation coefficient	p-value	n
CGRAV	0.88	<0.001	45
MWELL	0.81	<0.001	15
TEAST	0.82	<0.001	36
GIBPE	0.41	0.03	28
PCOLL	0.76	<0.001	21

Table 2. Correlations between Jeffreys-Matusita distances and class separation distances obtained with FSO tool for all two-class comparison for five plots.

The strengths, weaknesses, and suggested best uses for the three feature selection methods demonstrate that classification objectives and logistical constraints are best considered a priori (Table 3). In this study, we only tested nearest-neighbor classification using samples. However, two of the methods, CTA and JM, also provide threshold values that can be used for rule-based classification. The SEaTH tool used for determination of the JM distance provides threshold values for separating only two classes, while CTA outputs thresholds based on the entire classification tree.

	JM	CTA	FSO
Strengths	1) JM distances and rules for 2-class comparisons, 2) compatible with eCognition export	1) feature reduction and ranking, 2) non-parametric, 3) can obtain features or specific rules, 4) fast analysis	1) feature reduction within eCognition, 2) class separation distances, 3) fast without texture features
Weaknesses	1) no initial feature reduction, 2) requires multiple steps for feature selection, 3) assumes normality	1) no class separation distances, 2) potential for overfitting decision tree	1) “black box” feature selection approach, 2) unclear feature ranking, 3) no rules
Best uses	1) feature selection for NN or rule-based classification when separation distances are needed, 2) with limited number of classes and features	1) feature reduction and ranking for NN or rule-based classification, 2) with many classes and/or features, 3) features with non-normal distributions	1) feature reduction for NN classification

Table 3. Strengths and weaknesses of, and best uses for three feature selection methods for object-based image analysis of very high resolution digital aerial photography. The feature selection methods are Jeffreys-Matusita distance (JM), applied using the SEaTH tool, classification tree analysis (CTA), and feature space optimization (FSO).

5. CONCLUSIONS

In this study, we assessed three feature selection methods for object-based classification of rangeland vegetation using sub-decimeter resolution digital aerial imagery. All methods offered an objective approach for feature selection. JM resulted in the highest classification accuracies for three of the five sites; however, CTA was deemed to be the best method for this particular set of imagery, because of the efficient workflow, the ability to both rank and significantly reduce input features, and the lack of parametric assumptions for normality. Even though CTA ranked second in terms of accuracy, we determined that it was the best method if multiple images were to be analyzed. FSO was ranked third based on feature reduction capability, the “black box” nature, and lowest accuracies. JM is an attractive method if class separation distances are of interest. FSO also provided class separation distance, but they are not scaled for easy comparisons. Further studies will investigate the validity of these findings when the methods are applied to classifications of larger areas.

REFERENCES

Addink, E. A., S.M. de Jong, S.A. Davis, V. Dubyanskiy, L.A. Burdelow, and H. Leirs, 2010. The use of high-resolution remote sensing for plague surveillance in Kazakhstan. *Remote Sensing of Environment*, 114, pp. 674-681.

Carleer, A. P., and E. Wolff, 2006. Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, 27(6), pp. 1035-1051.

Chubey, M. S., S.E. Franklin, and M.A. Wulder, 2006. Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering and Remote Sensing*, 72(4), pp. 383-394.

Congalton, R. G., and K. Green, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, Boca Raton, FL, 137 pp.

Coulter, L.L. and D.A. Stow. 2008. Assessment of the spatial co-registration of multitemporal imagery from large format digital cameras in the context of detailed change detection. *Sensors*, 8, pp. 2161-2173.

Definiens, 2009. eCognition Developer 8.0 User Guide. Definiens AG, Munich, Germany.

Erdas, 2009. Erdas 2010, Leica Photogrammetric Suite User Guide. Erdas Inc., Norcross, GA, pp. 454.

Foody, G. M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70, pp. 627-633.

Herold, M., X. Liu, and K.C. Clarke, 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing*, 69(9), pp. 991-1001.

Johansen, K., S. Phinn, C. Witte, S. Philip, and L. Newton, 2009. Mapping banana plantations from object-oriented classification of SPOT-5 imagery. *Photogrammetric Engineering and Remote Sensing*, 75(9), pp. 1069-1081.

Kim, B., and A. Landgrebe, 1991. Hierarchical classifier design in high-dimensional numerous class cases. *IEEE Transactions on Geoscience and Remote Sensing*, 29(4), pp. 518-528.

Laliberte, A. S., D.M. Browning, J.E. Herrick, and P. Gronemeyer, In press. Hierarchical object-based classification of ultra high resolution digital mapping camera (DMC) imagery for rangeland mapping and assessment. *Journal of Spatial Science*.

Laliberte, A. S., E.L. Fredrickson, and A. Rango, 2007. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering and Remote Sensing*, 73(2), pp. 97-207.

Marpu, P. R., I. Niemeyer, S. Nussbaum, and R. Gloaguen. 2008. A procedure for automatic object-based classification. T. Blaschke, S. Lang, and G.J. Hay (Eds.), *Object-based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*. Springer, Berlin, Germany, pp. 169-184.

Neumann, K. J., 2008. Trends for digital aerial mapping cameras. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXXVII, Part B1, pp. 551-554.

Nussbaum, S., I. Niemeyer, and M.J. Canty, 2006. SEATH - A new tool for automated feature extraction in the context of object-based image analysis. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Salzburg, Austria, Vol. XXXVI-4, Part C42.

Yu, Q., P. Gong, N. Clinton, G. Biging, M. Kelly , and D. Schirokauer, 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing*, 72(7), pp. 799-811.

ACKNOWLEDGEMENTS

This research was funded by the USDA Agricultural Research Service and the National Science Foundation Long-Term Ecological Research Program, Jornada Basin IV: Linkages in Semiarid Landscapes.