Korean Journal of Remote Sensing, Vol.21, No.5, 2005, pp.383~392

Improving Urban Vegetation Classification by Including Height Information Derived from High-Spatial Resolution Stereo Imagery

Soojeong Myeong †

Program in Environmental and Resource Engineering

SUNY College of Environmental Science and Forestry, Syracuse, NY 13210, USA

Abstract : Vegetation classes, especially grass and tree classes, are often confused in classification when conventional spectral pattern recognition techniques are used to classify urban areas. This paper reports on a study to improve the classification results by using an automated process of considering height information in separating urban vegetation classes, specifically tree and grass, using three-band, high-spatial resolution, digital aerial imagery. Height information was derived photogrammetrically from stereo pair imagery using cross correlation image matching to estimate differential parallax for vegetation pixels. A threshold value of differential parallax was used to assess whether the original class was correct. The average increase in overall accuracy for three test stereo pairs was 7.8 %, and detailed examination showed that pixels reclassified as grass improved the overall accuracy more than pixels reclassified as tree. Visual examination and statistical accuracy assessment of four test areas showed improvement in vegetation classification with the increase in accuracy ranging from 3.7 % to 18.1 %. Vegetation classification can, in fact, be improved by adding height information to the classification procedure.

Key Words : High-spatial resolution imagery, stereo pair, image matching, parallax, height information, urban vegetation classification.

1. Introduction

Conventional spectrally-based land cover classification has limited success in separating grass and tree classes in urban vegetation because of spectral confusion (Myeong *et al.*, 2003; Zhang, 2001). Therefore, utilizing information other than spectral reflectivity could improve classification results for these two classes. One way to improve separation of these two classes is to use image texture, especially in imagery with high-spatial resolution. Adding texture information improved the classification results for color infrared digital imagery of an urban study area (Ryherd and Woodcock, 1996; Stefanov *et al.*, 2001; Berberoglu *et al.*, 2000; Myeong *et al.*, 2003). However, there still remains room for improvement in classification between vegetation classes; for example, shaded grass often misclassified as tree cover. Another way to improve

Received 31 August 2005; Accepted 10 October 2005.
† Corresponding Author: S. Myeong (smyeong@syr.edu)

classification is to incorporate vegetation height. The difference in vegetation height should help to differentiate trees and grass. Three-dimensional information can be obtained using LIDAR (LIght Detection And Ranging) or photogrammetric processing of stereo imagery. Photogrammetric processing has the advantage of deriving height information from less expensive imagery, thereby avoiding the expense of collecting LIDAR data. Automated processing to obtain height information is advancing and becoming useful (Schenk, 1999; Wolf and Dewitt, 2000). To estimate tree canopy height, stereo digital imagery is processed to extract differential parallax. Then, this parallax information can be used to resolve the confusion of the vegetation classes. This is due to the fact that differential parallax is greater for high objects (trees) than for low ones (herbaceous vegetation) (Avery, 1977; Wolf and Dewitt, 2000; Lillesand and Kiefer, 2000).

This paper reports on the study that classification of grass and tree classes is improved when height information is incorporated. The objective of this study was to test whether misclassified pixels can be identified and corrected when height of vegetation is considered. Specifically, this study will:

1. develop a method of automatically generating local height information from digital, high-spatial resolution, stereo imagery, and

2. determine the feasibility and potential of using this local height information to help distinguish vegetation classes in urban areas.

2. Materials

1) Study Area

Three areas with relatively flat terrain located in Syracuse, New York, were examined in this study using three stereo pair images. The average size of these areas was around 1 km^2. However, stereo pair 1 area has relatively hilly terrain compared to other two areas. The study areas were selected to encompass diverse land use types such as residential areas, cemeteries or parks to include enough area of trees and grass. In addition, four test areas were selected from these three stereo pairs for detailed accuracy assessment and visual inspection.

2) Imagery

Digital high-spatial resolution stereo pair images were used for this study (Emerge, 2003). The imagery was acquired on July 13, 1999, using a modified Kodak DCS 460 camera, using three different spectral bands (near infrared, red, and green). The imagery was collected with about 60 % overlap to provide full stereo capability for three-dimensional analysis.

The classification procedure for the originally classified imagery incorporated NDVI (Normalized Difference Vegetation Index) and texture information of three spectral bands, using the maximum likelihood decision rule (Myeong *et al.*, 2003). The confusion between trees and grass on the classified imagery revealed a limitation of this method. To test the concept of using height to improve classification, three independent stereo pairs were classified using the original methods to construct three classes: tree, grass, and non-vegetation. The classified images of these three stereo pairs were used in this study.

3. Method

The general procedure followed in this study was: 1) match the stereo pair imagery, 2) compute parallax, 3) estimate differential parallax, 4) correct misclassified pixels using differential parallax, and 5) assess the results through visual examination and statistical evaluation of test areas.

1) Image Matching

Estimating parallax for a pixel requires finding the conjugate location for that pixel in stereo images. Automated determination of conjugate points is accomplished by statistically matching the gray level distribution of a subset in one image to its counterpart in the other image (Wolf and DeWitt, 2000). Practically, the subset in one image (the "template") must be compared with numerous possible regions in the conjugate image (the "search array"). In other words, the search array is larger than the template because parallax and other geometric conditions will cause variable displacement of the conjugate point in the second image. However, computation costs and other practical considerations require that the search array be centered reasonably close to its conjugate location and that it be kept to a reasonable size.

As a first approximation to locating the search array, a simple affine transformation was used to predict the location of the conjugate point in the search image. A larger search array is then defined about this predicted location, and statistical comparison begins. The size of template was determined as 9 by 9, and that of search array was determined as around 15 by 15 based on how well each stereo pair matches after applying affine transformation. Matching occurred by comparing the gray level distribution of the template to the equal size patches in the search array. The comparison was made repetitively as the patch moved sequentially through every location in the search array. A common and straightforward method for performing the comparison is to perform cross correlation. Cross correlation is a measure of the linear relationship between two random variables (Kreyszig, 1979). The near infrared band of the imagery was used for image matching.

At each position inside the search area, the correlation value between the template and the corresponding part of the search area was computed. The coefficient ρ at pixel (i, j) is computed using the following equation (Schenk, 1999):

$$\rho = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2}$$

where the σ_{12} is the covariance of image patches 1 and 2, σ_1 is the standard deviation of image patch 1 (template), and σ_2 is the standard deviation of image patch 2 (patch in the search array). Correlation ranges from -1 to +1, with a correlation of 1 meaning a perfect match, which rarely exists due to image noise (Wolf and DeWitt, 2000). After computing the coefficient for every pixel in the search array, the correlation coefficients were compared and the most highly correlated pixel was selected as the matched point between the two images.

2) Parallax Computation

Parallax exists for all pixels imaged in the overlapping area of stereo images, proportionate to the elevation of each point. Once a stereo pair is matched, parallax can be found by measuring the location of the matched pixel in the first image and again in the second image. For example, the parallax for matched point a, P_a is defined as:

$$P_a = x_a - x_a'$$

where x_a, and x_a' are the photo coordinates (along the axis of the flight line) of the first image and the second image of the stereo pair respectively. After matching stereo pairs, the values of x_a and x_a' are obtained for each pixel and its conjugate pixel and used to compute a parallax for that pixel. In the same way, parallax was computed for every pixel in the study area.

3) Differential Parallax

Differential parallax refers to the change in parallax value for points at different elevations. For this study and for objects like trees and buildings, differential parallax is the difference in parallax between the base and top of the object. Therefore, after finding parallax for every pixel, differential parallax was computed for each vegetation class as an indicator of its height. However, automating this process required a method for estimating the parallax value for the base and top of objects.

This automation was implemented by finding a reasonable minimum parallax within a local window defined for each vegetation pixel. This local minimum value was assumed to represent the parallax of the ground. Then, this value was subtracted from the parallax of the target pixel as a means to estimate differential parallax.

Before finding the differential parallax, the majority class was identified first. An 11 by 11 window was used for this process after testing with diverse window sizes and also considering the canopy size of trees in the study area. Then, the local minimum parallax was found. Through empirical observation and experimental trials, it was determined that a different local window was more appropriate for tree pixels than for grass pixels. When the majority class within a local window was tree, a 21 by 21 window was used to find the minimum parallax value. A 21 by 21 window was large enough to include the canopy of a typical large urban tree or clump of trees and would include pixels near the edge of the crown or in a gap between crowns. When the majority class type was grass, a 5 by 5 window was used to find the minimum parallax. This smaller window helped to avoid noise that often exists in the grass class because grass is generally very homogeneous in terms of image values. This homogeneity limited the ability to find clear match points using cross correlation, making the parallax results incorrect and causing the local minimum to be unreasonably low.

4) Reclassification of the Imagery

The differential parallax layer was registered to the classified image layer and a model was developed to search pixels that were likely to have been misclassified. For pixels with correlation coefficients above 0.82, the model considered the differential parallax value associated with the pixel. These corrections were made only when the correlation values from matching were higher than 0.82 because low correlation values generally indicate pixels that are not correctly matched points (Wolf and DeWitt, 2000). The model looked for grass class pixels that had differential parallax higher than a threshold (1.8 pixels), and then changed the class of these pixels to be tree. Pixels labeled as tree but with differential parallax lower than the threshold were corrected to be grass. All threshold values were determined through experimental trials.

5) Accuracy Assessment

In this study, accuracy assessment was employed at three stages. The first accuracy assessment was performed to provide general information about overall accuracy for classification of the overlapping area of the three stereo pairs. A stratified random sample of 50 reference sampling points was selected in both the tree and grass classes (based on the reclassified imagery) to estimate accuracy with reasonable precision (Stehman, 1999). The same points were assessed for imagery of both before and after reclassification for consistency. Because the size of each stratum was different, the error matrices were constructed using the proper weighting of data from each stratum (Stehman, 1995).

The second accuracy assessment was performed to find detailed information about changes in classification. The accuracies for changed pixels of three stereo pairs were tested by sampling only the pixels that changed after adding height information to the classification procedure. A total of 200 pixels (100 pixels for each of tree and grass classes) from each stereo pair were sampled. The sampling was implemented only for pixels that had changed.

The third accuracy assessment was conducted on four small subset images from the three stereo pairs, to check the detailed effect of height information on improving classification. At the second stage, both visual examination and statistical accuracy assessment of four test areas were conducted (Figure 1). The evaluation compared results from before and after adding the height information. The test areas were selected to have mostly trees and grass. Statistical accuracy was assessed by enumerating all the pixels of vegetation class in the four test areas. The enumeration was based on polygons for each class that were drawn separately by photo interpretation and field visits and then treated as the correct reference map. This reference map was then compared with the classification based on the original methods and the classification modified using height information. Since revisions were made only to grass and tree pixels, changes in accuracy occurred only for these classes.

4. Results

The combined accuracies for vegetation classes for the three stereo pairs increased from 72.99 % to 79 %, from 72.99 % to 81.43 %, and from 76.07 % to 85.05 % respectively after adding height information to the classification procedure (Table 1). The accuracy increased by 6.01 %, 8.44 % and 8.98 % respectively and the average increase was 7.81 %. The user's accuracies for both tree and grass classes and the producer's accuracy for grass class increased greatly. The producer's accuracy for the tree class decreased slightly in pair l but increased a little in the other two pairs. As reclassification did not affect the non-vegetation class, the user's accuracy for the non-vegetation class remained the same 88 %, 88 % and 94 % respectively and the producer's accuracy of non-vegetation class for all three areas were 100 %.

The results of accuracy assessment of three stereo pairs were encouraging. The overall accuracy increased, and both producer's and user's accuracy increased after reclassification except producer's accuracy of the tree class for stereo pair 1.

The percentages of changed pixels from the grass class to the tree one are 3.3 %, 3.9 %, and 4.6 %, and the percentages of changed pixels from tree class to grass class are 8.7 %, 6 %, and 6.3 % respectively in the three stereo pair images. The detailed accuracies for changed pixels after adding height information to the classification procedure were summarized in Table 2. While pixels that changed into the tree class after reclassification show slight improvement, pixels changed into the grass class show great increase in accuracy for all the three stereo pairs.

Generally speaking, the correlation coefficients for the grass class were low while those of the tree class were high due to the fact that the tree class had texture and the grass class lacked the matching features. In this study, grass pixels, with tree shadow which was likely to be misclassified as the tree class, had high correlations because the tree shadow plays the role of the matching features. Therefore, these pixels were corrected as the right class (Figure 1). Since many misclassifications occurred on the grass class with the tree shadow, these helped greatly to improve the classification accuracy.

For all the four test areas, the classification improved greatly with the use of height information. Figure 1 shows visual comparisons of before and after adding the height information for the four test areas. Accuracy especially improved for shadow areas and small grass patches in the middle of forested areas.

Statistical accuracy assessment of the four test areas revealed that the improvement in classification varies depending on the characteristics of the test area. Table 3 shows the statistical results for each of

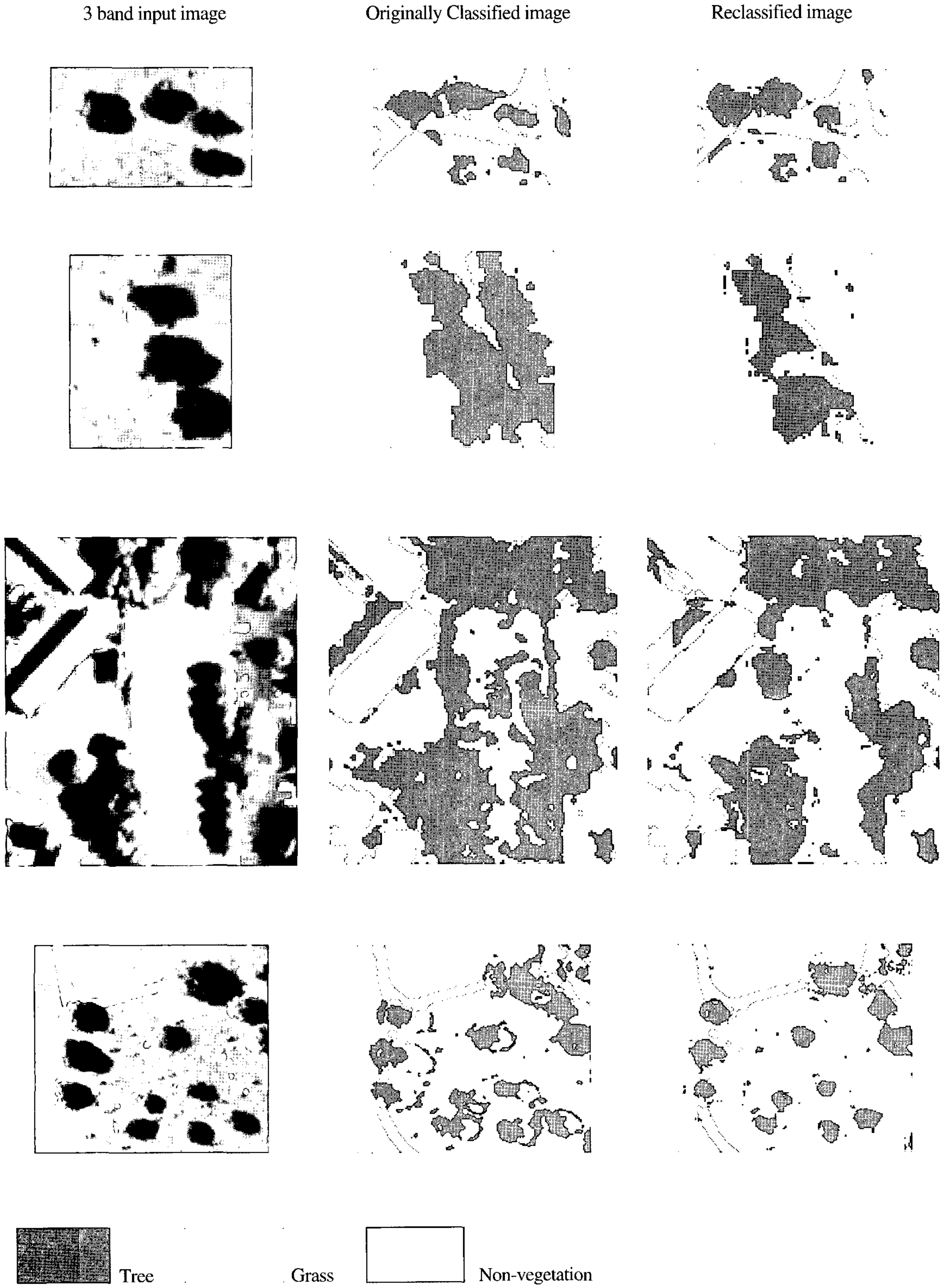

Fig. 1. Four test areas that show classification improvement after adding height information in classification.

Table 1. The overall error matrix of each stereo pair for both before and after reclassification using height information.

	Class	User's accuracy		Producer's accuracy		Overall	
		Without Height	With height	Without Height	With height	Without Height	With height
Pair 1	Tree	70.69 %	80.00 %	80.54 %	78.59 %	72.99 %	79.00 %
	Grass	76.19 %	78.00 %	65.10 %	79.44 %		
Pair 2	Tree	67.24 %	80.00 %	61.10 %	64.16 %	72.99 %	81.43 %
	Grass	76.19 %	82.00 %	80.70 %	91.06 %		
Pair 3	Tree	74.51 %	84.00 %	75.80 %	84.47 %	76.07 %	85.05 %
	Grass	77.55 %	86.00 %	76.33 %	85.57 %		

Table 2. Summary of classification results of 100 sampled pixels from each vegetation class for three stereo pairs. Sampling was based on the changed pixels after adding height information. Table input unit is number of correctly classified pixels out of 100 sampled pixels in each class.

Stereo pair	Reclassified vegetation class		Tree + Grass Correct pixels
	Grass → Tree	Tree → Grass	
1	52	68	120
2	61	72	133
3	63	76	139
Average (%)	58.2 %	72 %	65.1 %

the test areas. Although the specific producer's accuracy and the user's accuracy did not always increase, the overall accuracy of all four test areas always increased.

5. Discussion and Conclusion

This study tested the potential for improving classification by adding the height information extracted photogrammetrically from digital stereo imagery. The results confirmed the potential of incorporating this type of information in improving image classification of urban areas.

Adding height information improved vegetation classification in terms of accuracy, especially, the user's accuracy of the tree class increased greatly (Table 1). In this study, pixels reclassified as the grass class demonstrated greater improvement than pixels reclassified as the tree class (Table 2). However, this trend was not completely reflected on the overall accuracy assessment because the percentage of changed pixels is small. The reason why pair 1 presented lower improvement is potentially because of the difficulty in image matching due to its hilly terrain compared to other pairs (Table 2).

It is hard to generalize the trend of classification improvement because the amount of classification improvement varies with the test area. However, it is clear that considering height information and correlation during the classification process improves accuracy noticeable even by visual inspection. The results of four test areas revealed that classification improvement occurred especially on the grass pixels in the tree shadow area and on the tree pixels that included areas with low texture. Generally, the greatest improvements occurred in the grass class that had been misclassified as the tree class because they were in the shadow of tree. Also, most trees were rendered more correctly in the classified result and the grass class usually showed an appropriate revision to a more homogeneous condition. Finally, some areas that were clusters of trees showed improvement by better revealing crowns and gaps in the tree canopy.

Although this study showed that adding height information could improve the classification results, there are still several unsolved problems. Imperfect image matching could be the main problem, which in

Table 3. Accuracy assessment of the four test areas in Fig. 1. The first error matrix is before adding height information and the second one is after adding height information.

Test area 1

Class Name		Reference					
		Without Height			With Height		
		Tree	Grass	User's Accuracy	Tree	Grass	User's Accuracy
Map	Tree	0.16	0.14	**52.3 %**	0.19	0.12	**61.0 %**
	Grass	0.06	0.64	**91.4 %**	0.02	0.66	**96.7 %**
	Producer's Accuracy	**71.8 %**	**82.0 %**	Overall **79.8 %**	**89.6 %**	**84.3 %**	Overall **85.4 %**

Test area 2

Class Name		Reference					
		Without Height			With Height		
		Tree	Grass	User's Accuracy	Tree	Grass	User's Accuracy
Map	Tree	0.23	0.30	**43.3 %**	0.20	0.09	**67.0 %**
	Grass	0.01	0.46	**98.0 %**	0.04	0.67	**94.6 %**
	Producer's Accuracy	**96.1 %**	**60.6 %**	Overall **69.1 %**	**83. 9 %**	**88.2 %**	Overall **87.2 %**

Test area 3

Class Name		Reference					
		Without Height			With Height		
		Tree	Grass	User's Accuracy	Tree	Grass	User's Accuracy
Map	Tree	0.39	0.23	**63.1 %**	0.37	0.10	**78.9 %**
	Grass	0.03	0.35	**91.8 %**	0.06	0.48	**89.3 %**
	Producer's Accuracy	**92.8 %**	**59.9 %**	Overall **73.9 %**	**86.6 %**	**82.9 %**	Overall **84.5 %**

Test area 4

Class Name		Reference					
		Without Height			With Height		
		Tree	Grass	User's Accuracy	Tree	Grass	User's Accuracy
Map	Tree	0.15	0.08	**66.0 %**	0.13	0.02	**86.1 %**
	Grass	0.05	0.72	**93.5 %**	0.07	0.77	**91.5 %**
	Producer's Accuracy	**75.5 %**	**90.0 %**	Overall **87.0 %**	**64.8%**	**97.3 %**	Overall **90.7 %**

turn presents incorrect information for parallax. This study was based on the cross correlation method, which is a simple and widely used algorithm for image matching. This method works well when geometrical and radiometric distortions are small. However, repetitive patterns can make it difficult to find the correct conjugated pixels. Also if the matching entity is not unique enough, ambiguity occurs. Inaccurate matching can result from the difference of spectral and view angles between the first image and the second image and geometrical and radiometric distortions (Gruen, 1985).

The correlation coefficient, which is very important in interpreting the result of image matching, also has limitations. The main problem with this approach is that the highest correlation factor does not always indicate a true match (Toth and Krunpnik, 1994). Pixels located at distinct features demonstrate high correlations, but pixels in homogeneous areas such as grass fields result in low correlations. Those pixels often turned out to have incorrectly matched data. This trouble with low correlation is likely due to repetitive patterns and lack of distinct matching features. Therefore, pixels with low correlations should not be used for image reclassification. To avoid this problem, this study used threshold values and small local window sizes for differential parallax computation.

Further research might use other matching methods such as least squares or feature-based approaches to improve the matching results. Another alternative way to improve the results is multiband image matching. It has been reported that image matching techniques applied on multiband images provide higher accuracies (Gruen and Baltsavias, 1988; Heipke, 1992; Rosenholm, 1987). With multiband image matching, the accuracy of parallax would be higher than that obtained from single band imagery, thus improving the elevation and positional accuracies of mapped features (Saleh and Scarpace, 1995). Therefore, multiband image matching with various image matching strategies deserves more attention than that it has received.

This study used the affine transformation to make the programming simple while locating search areas in the conjugate images. Methods that use epipolar geometry or projective transformation should enhance the capability or improve processing efficiency or both. Therefore, further study may well experiment with more sophisticated methods for defining search areas.

The exploratory nature of this project dictated the use of simplest methods to derive height information. The goal was to evaluate the potential of utilizing height information before devoting substantial effort in a more robust implementation. If height information improved classification results with these straightforward approaches, it should be even more helpful with more accurate methods. The results are encouraging, providing evidence that automated generation of local height information from digital stereo imagery is possible. Furthermore, height and related information such as correlation improve classification of urban vegetation. Further research should be conducted using other sources of acquiring three-dimensional information techniques for a wide variety of land cover areas.

This study found that the photogrammetric approach using stereo imagery is a viable alternative method to extract height information. This study also proved that vegetation classification can be improved by adding height information to the classification procedure. The overall accuracy of all the test areas increased.

Acknowledgements

This research was sponsored by the United States Forest Service through cooperative agreement number 01CA11242343078. The use of trade names in this article is for the information and convenience of the reader. It does not constitute an official endorsement or approval by the United States Department of Agriculture or Forest Service of any products or service to the exclusion of others that may be suitable. This paper is dedicated to late P. F. Hopkins.

References

Avery, T. E, 1977. *Interpretation of Aerial Photographs*. 3rd Ed. Burgess publishing company, Minneapolis, Minnesota. 52 p.

Berberoglu, S., Llyoid, C. D., Atkinson, P. M., and P. J. Curran, 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean, *Computers and Geoscience*, 26: 385-396.

Emerge, 2003. www.emergeweb.com.

Gruen, A. W., 1985. Adaptive least squares correlation: a powerful image matching technique, *J. of Photogrammetry, Remote Sensing and Cartography*, 14(3): 175-187.

Gruen, A. W. and E. P. Baltsavias, 1988. Geometrically Constrained Multiphoto Matching, *Photogrammeteric Engineering and Remote Sensing*, 54(5): 633-641.

Heipke, C., 1992. A Global Approach for least-squares image matching and surface reconstruction in object space, *Photogrammetric Engineering and Remote Sensing*, 58(3): 317-323.

Kreyszig, E., 1979. *Introductory Mathematical Statistics: Principles and Methods*, John Wiley, New York, 470 p.

Lillesand, T. M. and R. W. Kiefer, 2000. *Remote Sensing and Image Interpretation*, John Wiley & Sons, Inc. New York.

Myeong, S., D. J. Nowak, P. F. Hopkins, and R. J. Brock, 2003. Urban cover classification using digital, high-spatial resolution aerial imagery, *Urban Ecosystems*, 5: 243-256.

Rosenholm, D., 1987. Multi-Point Matching using the Least-Squares technique for evaluation of three-dimensional models, *Photogrammetric Engineering and Remote Sensing*, 53(6): 621-626.

Ryherd, S. and C. Woodcock, 1996. Combining spectral and texture data in the segmentation of remotely sensed images, *Photogrammetric Engineering and Remote Sensing*, 62(2): 181-194.

Saleh, R. and F. L. Scarpace, 1995. Multi-dimensional matching techniques for stereophotogrammetric mapping, *Proc. of the ASPRS/ACSM Annual Conference*, Charlotte, NC. Feb 27 - Mar 2, 1995. Vol 2, 450-452.

Schenk, T., 1999. *Digital Photogrammetery*, TerraScience. Laurelville, OH.

Stehman, S., 1995. Thematic map accuracy assessment from the perspective of finite population sampling, *International Journal of Remote Sensing*, 16(3): 589-593.

Stehman, S., 1999. Basic probability sampling designs for thematic map accuracy assessment, *International Journal of Remote Sensing*, 20(12): 2423-2441.

Stefanov, W. L., M. S. Ramsey, and P. R. Christensen, 2001. Monitoring urban land cover change an expert system approach to land cover classification of semiarid to arid urban centers, *Remote Sensing of Environment*, 77: 173-185.

Toth, C. and A. Krupnik, 1994. Experiences with automatic image patch matching. *Proc. of the ASPRS/ACSM Annual Conference*, Reno, NV, April 25-28, 1: 644-651.

Wolf, P. R., 1983. *Elements of Photogrammetry*. McGraw-Hill, Inc., New York.

Wolf, P. R. and B. A., DeWitt, 2000. *Elements of Photogrammetry with Applications in GIS*. 3rd Ed. McGraw-Hill, Inc. New York.

Zhang, Y., 2001. Texture-integrated classification of urban treed areas in high-resolution color-infrared imagery, *Photogrammetric Engineering and Remote Sensing*, 67(12): 1359-1365.