A HIERARCHICAL OBJECT-ORIENTED APPROACH FOR EXTRACTING RESIDENTIAL AREAS FROM HIGH RESOLUTION IMAGERY

Juan Gu[a, b, *], Jun Chen[b] and Qiming Zhou[c]

[a] Faculty of Resource Science and Technology, Beijing Normal University, Beijing, China , julietgujuan@hotmail.com
[b] National Geomatics Center of China, Baishengcun, Zizhuyuan, Beijing, China, chenjun@nsdi.gov.cn
[c] Department of Geography, Hongkong Baptist Univertity, Kowloon Tong, Kowloon, Hong Kong, China, qiming@hkbu.edu.hk

KEY WORDS: Residential areas extraction, High resolution imagery, Hierarchical object-oriented, Segmentation, Classification, Change detection

ABSTRACT:

Change detecting and timely updating of residential areas is widely recognized as one of the most challenging tasks for an operational GIS. High spatial resolution satellite data provide an efficient source of information. This paper describes a hierarchical object-oriented approach for extracting residential areas from IKONOS and SPOT5 images acquired in 2001 and 2004, respectively. Hierarchical object networks are formed through segmenting the image into non-overlapping and homogeneous regions (i.e. image objects) with separated fine and coarse resolutions, and constructing topologic relationship between them. The extraction of residential areas is implemented in three processing steps based on this object network. First, classify the image objects into several land cover classes such as roofs, green lands, roads, barren lands and water using image objects' spectral, textural and spatial properties identified by the fine-resolution segmentation. The second step is performed based on the classification result of the first step for coarse-resolution segmentation. Only two classes are concerned here, namely, residential and non-residential. The decision to whether an image object belongs to residential is made by computing the percentage of 'roofs' image objects derived from the first step. The third step merges all adjacent image objects that belong to residential class. All adjacent image objects that represent the same, or partially same, structure are merged into one new image object representing the whole residential area. The change of residential areas between the two image dates is then analyzed and discussed.

1. INTRODUCTION

Residential areas occupy a relatively small portion of earth surface, but their exact extent, distribution and expansion have a great concern for governors and planers. In a rapid development region such as eastern china, timely updating spatial database that adequately reflect the change and expansion of residential areas is widely recognized as one of the most challenging tasks for an operational GIS. Remote Sensed data provide an efficient source of information for detecting and monitoring the change of residential areas, especially when images with high spatial resolution (e.g. IKONOS, SPOT5 and QuickBird) become readily available. However, high spatial resolution with fewer spectral bands does not always lead to more accurate extraction results because of poor classification results due to greater spectral variation within class and a greater degree of shadow (Laliberte *et al.*, 2004). Traditional pixel-based method can result in significant errors because of confusion between the spectral characteristics of reflectance from building roofs with some other land cover objects (Harris et al., 1995; Gao *et al*., 1998).

By considering spatial relationship between adjacent pixels, however, much information maybe contained including texture and shape information in high resolution imagery, which allows for identification of individual objects as opposed to single pixels (Thomas *et al.*, 2003). Image analysis leads to meaningful objects only when the image is segmented in 'homogeneous' areas (Baatz *et al*., 2000). The segmentation ensures the subsequent object based image processing. An object-oriented approach enables the user to incorporating both spectral information (tone, color) as well as spatial arrangements (size, shape, texture, pattern and association with neighboring objects) for image analysis. This method comes closer to the way that human operators interpret information visually from image and which has been successfully implemented for years in many applications. Laliberte et al. (2004) showed object-oriented multiscale image analysis method had advantages over pixel-based classification approach in assessing the rate of shrub encroachment into desert grassland in southern New Mexico from 1937 to 2003. Object-oriented approach was also used in detecting urban feature, and classification results were more homogeneous than pixel based method and easier for human to understand (Meinel *et al*., 2001; Hofmann, 2001).

The objectives of the study are twofold. The first is to develop methods for extracting residential areas from high resolution satellite imagery using a hierarchical object-oriented approach. The second objective is to analyze temporal pattern of residential area change by comparing the classification results of multitemporal images.

2. STUDY SITE AND DATA SET

The study site is Xihongmen Town, which is located in the southern part of Beijing City. It covers an area of 32 square kilometers with 50,000 permanent residents. Xihongmen Town was chosen as experimental town both at municipal level and national level in 1999 and in recent years it has accelerated its urbanization and made outstanding achievements. There are two main kinds of building style in the town. In the township are Chinese traditional bungalows with small yards with surrounding walls. These bungalows arrange in order and the density is high. In the fringe of the town, several new residential areas were built with modern high buildings of five to seven floors. The density of high buildings is lower and most of them are surrounded by trees and grassland.

(a)

(b)

Figure 1. Image data of Xihongmen Town
a: The IKONOS sub-scene (only multispectral channels)
b: The SPOT5 sub-scene (only multispectral channels)

The data used throughout this paper are two different temporal sub-images that covering the study site (shown in Fig. 1). One is the sub-image of an IKONOS scene (acquired on 26th, April 2001, local time), and another is the sub-image of a SPOT5 scene (acquired on 2th, June 2004, local time). IKONOS provides multispectral data in the red, green, blue and near-infrared (NIR) channels at 4.0m spatial resolution and Panchromatic (PAN) channel at 1.0m resolution. SPOT5 provides multispectral channels (10.0m) and PAN channel (2.5m). In this paper, only multispectral channels were used for residential extraction. All channels were used in accuracy assessment phase.

Both images were converted to the Universal Transverse Mercator (UTM) coordinate system, through a second-order polynomial rectification algorithm based on common control points, extracted from a topographic map at the scale of 1:10,000, and a nearest-neighbor method was used to resample the images. Evaluation of the registration accuracy yielded root-mean square error (RMS Error) values equivalent to 0.6, 0.4, 0.8 and 0.4 of a pixel size for the IKONOS 4m image, IKONOS 1m image, SPOT 10m image and SPOT 2.5m image, respectively. A principal component method was used to merge the PAN image with multispectral image to produce 1m multispectral IKONOS image and 2.5m multispectral SPOT5 image for the reference of accuracy assessment.

3. METHODOLOGY

3.1 Multiresolution segmentation

Objects are basic elements of an object-oriented approach and refer to contiguous regions in the image (i.e. image objects). Multiresolution segmentation is the first step for generating object. The algorithm used here is a fractal net evolution approach (FNEA) developed by Baatz and Schäpe in 2000. It is a bottom-up region-growing technique starting with one-pixel objects. In numerous subsequent steps smaller image objects are merged into bigger ones (more pixels) based on the chosen scale, colour, and shape parameters, which define the growth in heterogeneity between adjacent image objects. This process stops when the smallest growth exceeds the threshold defined by the scale parameter (Benz *et al*., 2004).

The heterogeneity criterion consists of two parts: a criterion for tone and a criterion for shape. The spectral criterion is the change in heterogeneity that occurs when merging tow image objects as described by the change of the weighted standard deviation of the spectral values regarding their weightings. The shape criterion is a value that describes the improvement of the shape with regard to two different models describing ideal shapes (eCognition User guide, 2000).

The overall fusion value f is computed based on the color heterogeneity $\triangle h_{color}$ and the shape heterogeneity $\triangle h_{shape}$ as follows:

$$f = w_{color} \cdot \Delta h_{color} + w_{shape} \cdot \Delta h_{shape}, w_{color} \in [0,1], w_{shape} \in [0,1], w_{color} + w_{shape} = 1 \quad (1)$$

where w_{color} and w_{shape} can be defined according to the resolution of the image used and objective of the user want to received.

The color criterion $\triangle h_{color}$ is the weighted mean of all changes in standard deviation for each channel c. The standard deviations σ_c themselves are weighted by the objects size n_{obj}:

$$\Delta h_{color} = \sum_c w_c (n_{merge} \cdot \sigma_{c,merge} - (n_{object\_1} \cdot \sigma_{c,obj\_1} + n_{object\_2} \cdot \sigma_{object\_2})), w_c \in [0,1] \quad (2)$$

Where: n is object size; σ is standard deviation of the object; and w_c is weight of c band defined.

The shape criterion $\triangle h_{shape}$ consists of two parts: smoothness $\triangle h_{smooth}$ and compactness $\triangle h_{compt}$ criterion:

$$\Delta h_{shape} = w_{compt} \cdot \Delta h_{compt} + w_{smooth} \cdot \Delta h_{smooth}, w_{compt}, w_{smooth} \in [0,1], w_{compt} + w_{smooth} = 1 \tag{3}$$

Where: w_{compt} and w_{smooth} can be defined according to data used and objective desired.

The change in shape heterogeneity caused by merge is evaluated by calculating the difference between the situation after and before the merge. This results in the following method of computation for smoothness and compactness:

$$\Delta h_{smooth} = n_{merge} \cdot \frac{l_{merge}}{b_{merge}} - (n_{object\_1} \cdot \frac{l_{object\_1}}{b_{object\_1}} + n_{object\_2} \cdot \frac{l_{object\_2}}{b_{object\_2}}) \tag{4}$$

$$\Delta h_{compt} = n_{merge} \cdot \frac{l_{merge}}{\sqrt{n_{merge}}} - (n_{object\_1} \cdot \frac{l_{object\_1}}{\sqrt{n_{object\_1}}} + n_{object\_2} \cdot \frac{l_{object\_2}}{\sqrt{n_{object\_2}}}) \tag{5}$$

Where: n is the object size, l is the object perimeter and b is the perimeter of the bounding box.

Throughout the segmentation procedure, the whole image is segmented and image objects are generated based upon several adjustable criteria of homogeneity or heterogeneity in color and shape. Adjusting the so-called scale parameter indirectly influences the average object size: a larger value leads to bigger objects and vice versa. Additionally the influence of shape as well as the image's channels on the object's homogeneity can be adjusted. During the segmentation process all generated image objects are linked to each other automatically (eCognition User Guide, 2000).

In this application, two different level approaches were used to enable the description of super and sub- objects relationships for both two images. For multispectral IKONOS (4m) data, first segmenting image with scale parameter 100, color 0.7/ shape 0.3, smooth 0.5/compactness 0.5. The segmentation result shows in upper left of Fig. 2. In the image, purple segments are most of residential areas or at least have some buildings, the deeper the color, the more buildings there are, grass green segments are pure vegetation, bottle green segments are the mixture of vegetation with other ground covers such as sparse buildings, barren land, part of roads are cover be trees so the color of their segments looks the same of that vegetation, part of them are made by concrete or asphalt and their segments approximates that of residential areas. It can be seen that 90% segments present meaningful ground cover. But in the lower part of the image, the segments present most like vegetation class and in fact they should be classified to residential class. Residential areas here are constructed with sparse buildings surrounded by numerous high trees and grassland. In order to make correct segmentation here, we made a second level segmentation based on above result with scale parameter 30, colour 0.9/shape 0.1, smooth 0.5/compactness 0.5 (see upper right of Fig. 2). The result of this level could segment all buildings correctly and the details of the image are well reserved. Do the same process to SPOT5 in two scale level. The segmenting parameters and the result can be seen in Fig. 2.

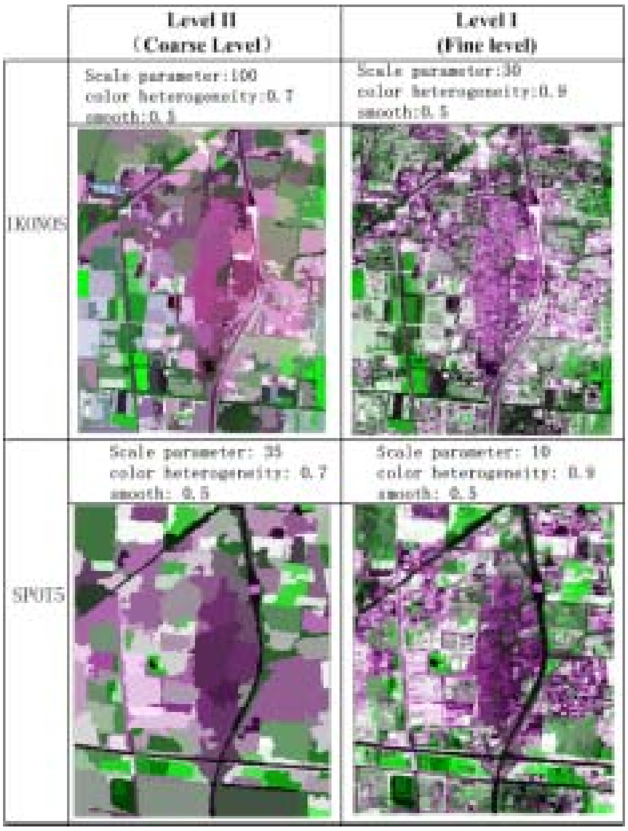


Figure 2. Multiresolution segmentation results of IKONOS and SPOT5 data

3.2 Classification

The classification of image objects can be performed by using nearest neighbor (NN) classifiers based on marking typical image objects as representative samples or by using membership functions which is based on fuzzy logic theory combined with user-defined rules. Both act as class descriptor. While the nearest neighbor classifier describes the classes to detect by sample objects for each class which the user has to determine, fuzzy membership functions describe intervals of feature characteristics wherein the objects do belong to a certain class or not by a certain degree. Thereby each feature of image objects can be used either to describe fuzzy membership functions or to determine the feature space for the nearest neighbor classifier. A class then is described by combing one or more class descriptors by means of fuzzy-logic operators or by means of inheritance or a combination of both (Hofmann, 2001). A membership function ranges from 0 to 1 for each object's feature values with regard to the object's assigned class. The fuzzy rule base defines criteria such as 'to all image objects, if its value of length/width is larger than a certain value, it may be segment of roads, and it is sure not belong to roofs. Spectral, shape, and statistical characteristics as well as relationships between linked levels of the image objects can be used in the rule base to combine objects into meaningful classes (Benz *et al.*, 2004).

Features for classification are computed on image objects obtained from segmentation, not on single pixel. Therefore, classification can address an astonishingly broad spectrum of different kind of information. Beyond spectral information there is shape information, texture information and operating – over the network of image objects – many different relational or context features. For each feature this information is computed per object considering its actual shape and size. Thus, typical failures of filter operations, especially on transitions between different types of areas, are avoided (eCognition User Guide, 2000).

In this paper, the extraction of residential areas is performing on two segmentation levels obtained. In Level I (fine scale level), classification was performed using both standard nearest neighbour classifiers and membership function. The features of image objects were used in classification procedure including:

(1) Mean value
Mean value is calculated from the layer values of all pixels forming an image object. Objects belong to the class of roofs have high mean value.
(2) Grey level co-occurrence matrix (GLCM): Homogeneity
The GLCM is a tabulation of how often different combinations of pixel grey levels occur in an image. A different co-occurrence matrix exists for each spatial relationship. Mostly, GLCM are not used directly, but features based on them are computed. These features describe some characteristics of texture, such as homogeneity, contrast and entropy. Haralick and Shanmugam (1973) suggested the use of 14 textural features. Homogeneity was select to compute textural characteristics of image objects. If the image is locally homogenous, the value is high.
(3) Shape index: Length/width

Images were classified into roofs, roads, barren land, green land and water （see Fig. 3）. Although in many cases the roofs of building were well outlined in this level, a sufficient classification of single houses based on spectral feature was impossible. The reason was that some roads, barren land and roofs of building were made in same materials such as concrete or asphalt. It must select some other features to discriminate them. Texture features (the degree of homogeneity of image objects) and shape index (Length/width) were selected to improve the classification by expressed in membership function. Constructed functions:

If *Length/width (object)> a certain value,* Then: *object* $\in$ *roads;* Else: o*bject* $\in$ *roofs.*

If *homogeneity (object) > a certain value,* Then: *object* $\in$ *barren land;* Else*: object* $\in$ *roofs*

Membership function was also used in Level II (coarse scale level) to build a rule to describe the building density of the object based on the classification of Level I. The decision to whether an image object belongs to residential is made by computing the percentage of 'roofs' image objects derived from the first step. Constructed function:

If
$$\sum Area(Object_{son} \in builings) \Big/ Area(Object_{father}) \ge a \quad certain \quad value$$
Then: $Object_{father}$ $\in$ residential area; Else: $Object_{father}$ $\in$ non-residential area

According to the procedure introduced above, the classification was performed on the two data, respectively. The classification results show in Fig. 3. In classification results of Level I, it can seen that roads were identified in SPOT5 is better than in IKONOS. That because the roads were widened and refaced and a lot of trees alongside the roads were destroyed after 2001. Grass land was classified very well in both images. The identification of roofs in SPOT5 is also better than that in IKONOS for variance of spectral characteristics of reflectance of roofs in IKONOS. In classification results of Level II, residential areas in both data were well presented and it was easy for the following change detection.

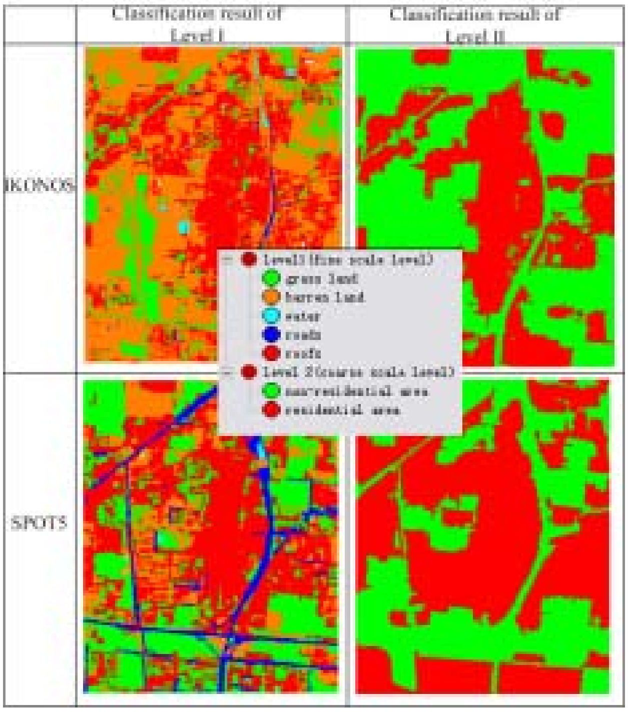


Figure 3. Classification results of IKONOS and SPOT5 data

By far, the work for extraction of residential areas was not finished. A last step still need to do to merge all adjacent image objects that classified to residential areas into new image objects representing the entire residential areas. This was done by classification-based fusion. This process used a knowledge base from a classification of the image objects. The formulation of the rules for the so called classification based fusion is carried out by editing structure groups. A structure group is a collection of classes representing the same structure in an image and can consist of classes defined for different levels in the image object hierarchy. This way, a number of image objects forming a residential area can be merged into one image object representing an entire residential area. In order to see clearly the fusion results, the fusion result of residential areas were vectorized and overlay with corresponding images. The overlay figures show in Fig 4..

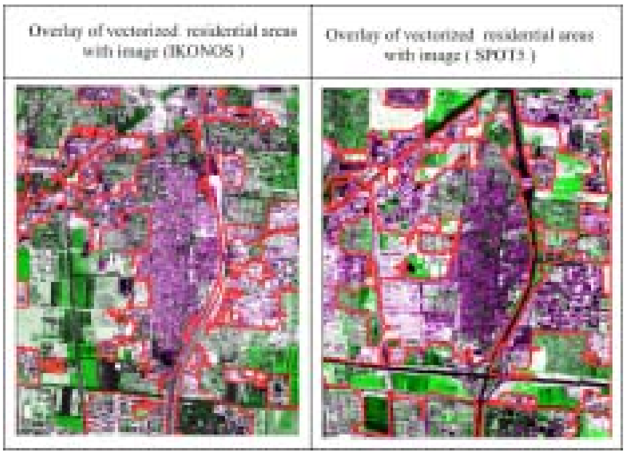


Figure 4. Overlay of vectorized residential areas with corresponding images

3.3 Accuracy Assessment of classification

The Accuracy Assessment Tool in ERDAS Imagine was used to compute the quality of classification and identification of residential areas. The final classification results were output and assess by produce 100 reference points randomly on the PAN-sharpen images. For IKONOS data, the producer's accuracy of residential areas is 75%, user's accuracy is 65%. For SPOT5 data, the producer's accuracy of residential areas is 92%, user's accuracy is 84%. The reason for the classification error is the confusion between the spectral characteristics of reflectance from building roofs with and barren land.

3.4 Change Detection

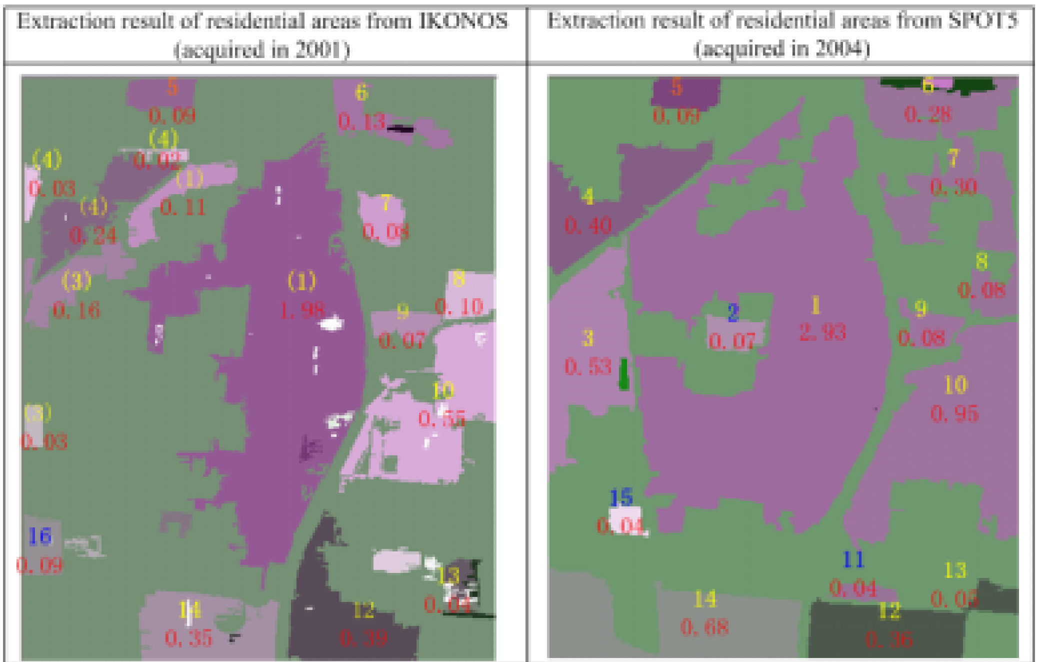
Change detection of residential areas was conducted based on the final classification results (show in Fig. 5). In 2001, the residential area of Xihongmen Town was totally 4.33 km2, and to 2004, the total residential area up to 6.99 km2. The increasing rate is 61% in three years. Compare the two classification results showing in Fig 5. , we can see that most of the residential areas are expanded only residential 12 decreased a little because new road was built through its upper part and some buildings had to be destroyed. Residential 16 was also destroyed for this reason. Among all the residential, the residential 1(the township of Xihongmen Town) got the biggest expansion of 0.95 km2. A lot of new residential districts (2, 3) were built on the place where were farmland in 2001.

4. DISCUSSION AND CONCLUSION

Hierarchical object-oriented classification systems provide us a good way to segment arbitrary images into highly homogeneous objects and extract residential areas by using complex rules based on spectral, textural and spatial characteristics. This method approximates the way human operation interpret residential areas from images visually, but has the advantage of an automated classification routine. Major problems are the generation of meaningful objects that fit to semantic classes of the object lists. Additional knowledge from thematic layers has not been used in this study so far but will be included in our further work.

REFERENCES AND/OR SELECTED BIBLIOGRAPHY

Baatz, M. and Schäpe, A., 2000, Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation, In: Strobl, J., Blaschke, T. and Griesebner, G. (eds.), Angewandte Geographische Informationsverarbeitung XII, Wichmann-Verlag, Heidelberg : 12-23



Note: red numbers in both images represent the area of the corresponding residential area (unit: km^2), yellow ones refer to those areas changed, blue ones in the left image indicate the disappearance in the right image and blue ones in the right image mean they are new residential areas.

Figure 5. Change detection of residential areas based on the classification results of the two data

Benz, U.C., Hoffmann, P., Willhauck, G., Lingenfelder, I. and Heynen, M., 2004, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58: 239-258

DEFiNiNENS AG, 2000. eCognition User Guide, *http://www.defineiens.com*

Gao, J., Skillcorn, D., 1998, Capability of SPOT XL data in producing detailed land cover maps at the urban-rural periphery, *International Journal of Remote Sensing*, 19(15): 2877-2891

Hofmann, P., 2001, Detecting urban features from IKONOS data using an object-oriented approach, *RSPS2001 Proceedings*: 79-91

Harris, P.M., Ventura, S.J., 1995, The integration of geographic data with remotely sensed imagery to improve classification on an urban area, *Photogrammetric Engineering and Remote Sensing*, 61(8): 993-998

Haralick, R.M. and Shanmugam, K., 1973, Combined spectral and spatial processing of ERTS imagery data, *Journal of Remote Sensing of Environment*, 3:3-13

Laliberte, A.S., Rango, A., Havstad, K.M., Paris, J. F., Beck, R. F., McNeely, R. and Gonzalez, A.L., 2004, Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico, *Remote Sensing of Environment*, 93: 198-210

Meinel, G., Neubert, M. and Reder, J., 2001, The potential use of very high resolution satellite data for urban areas- first experiences with IKONOS data, their classification and application in urban planning and environmental monitoring, in Jürgens, C. (ed.): *Remote Sensing of urban areas,* Regensburg: 196-205

Thomas, N., Hendrix, C. and Congalton, R.G., 2003, A comparison of urban mapping methods using high-resolution digital imagery, *Photogrammetric Engineering and Remote Sensing*, 69(9): 963-972

ACKNOWLEDGEMENTS

This project is financed by the National Natural Science Foundation of China (Contract No. 40337055). The author would like to thank Beijing Hyundai Space Eye Imaging Technology Co. Ltd. for supplying the software of eCognition and Beijing Spot Image Co. Ltd. for providing SPOT5 data.