Change Detection from SPOT- Panchromatic Imagery at the Urban-rural Fringe of Ho Chi Minh City, Vietnam

Annette M. Kim[1], Desheng Liu[2], and Peng Gong[2]

[1]Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Center for Assessment and Monitoring of Forest and Environmental Resources
Department of Environmental Science, Policy, and Management
University of California, Berkeley, CA 94720, USA

Abstract

This paper proposes a simple, automated method to detect rural to urban land use changes at the pixel level of SPOT-Panchromatic images in the developing world. The proposed method entails two tasks: (1) classification of images as either urban (built-up) or rural (non built-up) at a relatively high level of spatial detail (pixel level) in order to include the classification of houses made of natural materials. The binary classification was performed through a combined thresholding of spectral information and spatial information derived by a normalized high-pass filter. An automatic procedure was used to determine the optimal threshold (2) classified image comparison of two different dates by overlaying them to detect changes from rural to urban land use during the corresponding period. An accuracy of 82.31% was achieved for the final change map.

I. INTRODUCTION

Information about rural to urban land use conversions is crucial to city leaders in the developing world because they have the responsibility to manage the rapid urban growth in order to help promote its economic productivity and standard of living. However, these same cities usually lack the funding and the trained staff to manually collect and update data about the rapid changes. In light of this dilemma, remotely sensed images provide an ideal data source to monitor the changes because of their relatively low cost and repetitive acquisition capability. Furthermore, digital processing of these remotely sensed images can make the usually laborious change detection task automated and efficient.

A large number of change detection methods have been developed using multi-temporal remote sensing images since the 1970's. These include post-classification comparison, image differencing, image ratioing, image regression, principle components analysis, multi-date classification, change vector analysis, and artificial neural network. Reviews or comparative studies can be found in Singh (1989), Ridd and Liu (1998), Mas (1999), and Gong and Xu (2003).

Change detection with respect to the urban environment has generated considerable research interest (Riordan 1980, Jensen and Toll 1982, Martin 1989, Gong et al 1992, Gong 1993, Ridd 1995, Li and Yeh 1998, Zhang 2001, Gluch 2002). There exist some common features among these studies: 1) they defined urban land use as a composite of different land cover types (such as concrete, asphalt, trees, grass, different roofs) and utilized a land use classification strategy; 2) the spatial details of these studies were rough and the finest change unit usually consisted of a relatively large area ranging from tens of pixels to hundreds of pixels.

Although classifying such large units of land is sufficient for natural environment studies, for urban studies they are not as helpful. The types of research questions social sciences such as urban economics and urban planning ask concern the behavior of individuals and firms and how they are altering the land use patterns. Therefore, a unit closer to the size of urban land development projects is needed in order to understand the agent-based determinants of growth (Irwin and Bockstael, 2002). Parcel level data would be the ideal unit but many local governments have not had the funds or capacity to digitize and regularly update land survey and ownership data, especially in the developing world. Given the available data, a next best alternative would be rural to urban land use conversion data at a smaller unit. The pixel level scale better approximates the size of individual residential land developments, especially in the developing world. Another challenge particular to the classification of images from the developing world is that houses, especially in squatter settlements and on the periphery where the urban growth occurs, consists of a wider diversity of building types than in the US. Very small houses made of plant material similar to their surroundings proliferate rendering brightness value alone too blunt a vehicle for classification (Bertaud, 1989). If remote sensing images could be classified at a smaller unit and with greater sensitivity to low brightness value construction materials, rural to urban land use conversion data could be generated that would be practically helpful to urban planners and useful to urban scholars.

SPOT-Panchromatic (SPOT-PAN) images are valuable in this regard since they have collected 10 meter resolution images

1082-4006/04/1001-42$5.00

since 1986. Furthermore, SPOT-PAN images also have great potential in generating important time-series urban land use conversion data because they are the only commercially available satellite images at high resolution that continuously acquired images from the 1986 to the present for most of the globe, coincidentally the period of historic global urbanization. Since in the case of many Third World cities, this is the only record of the rapid urban land use changes that were taking place during this period, it is important to develop appropriate methods to interpret this data. The ability to analyze spatial urban growth trends over time would be a boon for both policymakers and social scientists alike.

However, in generating spatially detailed classification of land cover, many of the previous methods of group-pixel based texture analysis (such as frequency based methods) are inappropriate because the output of the group-pixel classifier is a composite of different parcel units. Zhang (2001) presented a new approach to detect the detailed urban housing development by fusing SPOT-PAN and TM images and performing spatial feature post-classification. In a recent study, Gluch (2002) detected the urban growth as the change of non-built pixels to built pixels from a merged TM and SPOT-PAN data. The spatial detail of the study by using two types of texture analysis is at the super pixel (3X3 pixels) level. Gluch utilized a user-defined threshold to determine whether a super-pixel is classified built or un-built. Although Gluch (2002) used both SPOT-PAN and aerial photos, she noted that SPOT-PAN alone would be sufficient to distinguish built from non-built features through texture analysis. There are few works done to detect the urban growth using SPOT-PAN alone.

This paper proposes a simple, automated method to detect rural to urban land use changes on the urban periphery of cities in developing country by testing a new algorithm that provides detailed classification using SPOT-PAN images alone.

II. STUDY AREA

Like the majority of developing country cities, urban land development at the fringe of Ho Chi Minh City (HCMC) dramatically increased during the last decade. Huge inflows of population migration and rising incomes have created a high demand for shelter and urban services which has fueled this rapid urban growth. Even by the most conservative statistics, the city's population has increased by over a million people during the 1990-2000 period (Statistical Office of Ho Chi Minh City, 2000). In order to cope with this situation, Ho Chi Minh City incorporated 5 outlying districts into the city boundaries in 1997 and has plans for future expansions. Within the city, the area that is experiencing some of the most dramatic growth is the urban periphery. While in the inner urban district areas, population grew by 21% during the 1990s, in the five new urban districts on the edge of HCMC population grew by 58% and by 73% in the rural fringe districts just outside the city boundaries.

The study area is the urban periphery of HCMC which consists of a roughly 60,000 hectare ring of land around the city's center. This area has a wet season lasting from roughly June to September and a drier season in the other months with the driest period being March and April. The topography in this area is flat with such land cover types as rivers, wetlands, irrigated farmlands, dry farmlands, wet bare soil, dry bare soil, and built up urban areas. The built up areas consist of a variety of land uses such as residential, commercial, industrial, and mixed-use.

III. DATA DESCRIPTION

SPOT-PAN images over Ho Chi Minh City were purchased in ortho-rectified and geo-referenced form. This source of data is the only record of the changes that occurred throughout the transition period while land information institutions themselves were being reformed. The first image purchased from their archives was taken in January 1994. The SPOT satellite was then commissioned to repeatedly acquire images of HCMC during January of 2001 until a cloud-free image could be taken. The 10 m resolution of the SPOT-PAN images is particularly helpful in the context of Ho Chi Minh City's building typology where the building footprint of a house is typically 5 m X 10 m and houses are clustered close together.

For accuracy assessment, we relied on some visual interpretation of the SPOT-PAN image in 1994 and relied on field data collection done during January 2001 when the second satellite image was taken. The image of 2001 was registered to the image of 1994 using a first order polynomial model with nearest neighbor resampling. The RMS error for this image-to-image registration was 0.12 pixel which is satisfactory for change detection purposes.

IV. METHODS

The 1994 and 2001 images were classified as a binary map consisting of 'built' (value = 1) and 'non-built' (value = 0) pixels. 'Built' land uses are assumed to be urban without differentiating between different classes of urban land use such as residential, industrial, or commercial. To achieve detected changes at a higher spatial detail than previous methods, a classification algorithm at the pixel level was proposed in this project. The algorithm includes two steps: 1) image enhancement with a normalized high-pass filter, and 2) combined thresholding of the original image and the enhanced image.

Pixels of most 'built' areas have brighter tones than 'non-built' areas in the image. However, in a certain range there still exists much confusion between the 'built' area and 'non-built' area (the pixels with digital number (DN) between 55 and 70 in Figure 1) especially between dark small houses with grass roofs and dry bare land. Therefore, the spectral information of

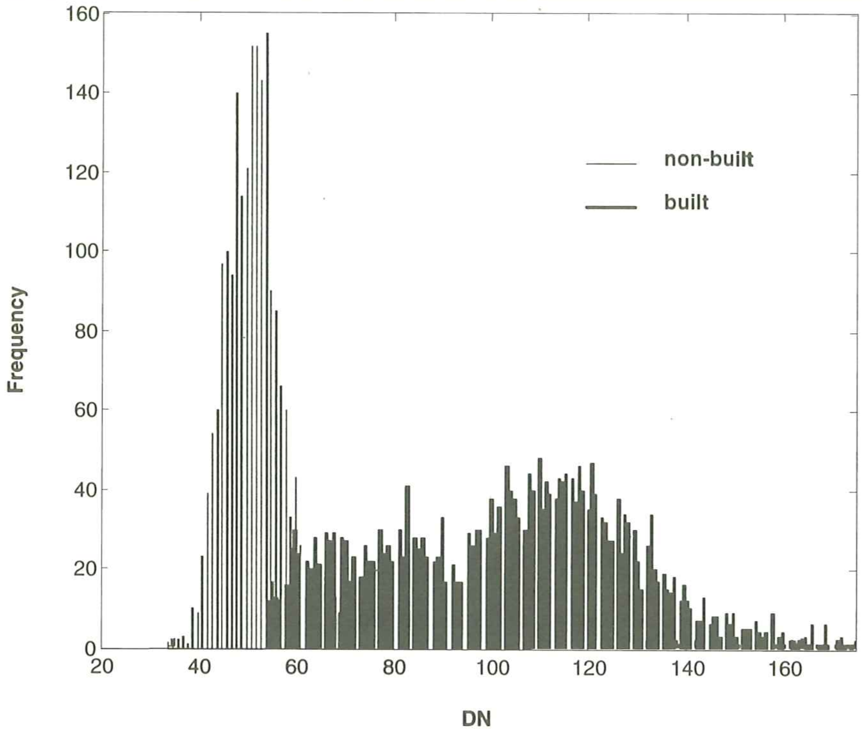

Figure 1. Histogram plot of sample 'built' and 'non-built' pixels taken from the SPOT-PAN image of the urban periphery of Ho Chi Minh City in 2001.

SPOT-PAN alone cannot separate the 'built' area from 'non-built' area completely and accurately. Spatial contextual information needs to be combined with the spectral information to improve the classification.

Image enhancement with respect to the spatial context makes good use of the high spatial resolution of the SPOT panchromatic image, which will obviously augment the ability to differentiate the built-up pixel from the non built-up. Since the previous studies' textural enhancement algorithms based on group-pixels are not suitable to this project, a normalized high pass filter will be used to enhance the spatial contextual information from the image. It is well known that high-pass filters such as the Laplacian filters have the effect of edge enhancement. In homogenous areas, no edge will be enhanced and the resultant image is uniform with a value near zero. In heterogeneous areas, gray-level contrasts in the local neighborhood will be enhanced as edges. When the filter size is greater than the size of the object, the object as a whole will be detected as an edge. This holds in the case of small houses in the urban periphery of HCMC.

A typical n X n high-pass filter can be expressed as an n X n matrix M with the central element as (n^2-1) and the others as -1. It is easy to show that the convolution result of this high-pass filter equals $(DN_{central} - Mean) \times n^2$. Here, Mean is the mean value of all DNs within the n X n window. In bright areas, the DN of the central pixel is higher than that in dark areas and the contrast between built-up and its background (non built-up) is larger than that in dark areas. Therefore, it can be observed that the convolution result for 'built' pixels positively correlates to the DN of the central pixel. Also, the convolution will increase with n. To utilize a uniform convolution measure to all the 'built' areas with respect to different filter sizes, a normalized high-pass filter is proposed as $(DN_{central}-Mean)/DN_{central} = 1-Mean/DN_{central}$. The normalization makes the optimal threshold selection much easier because the 'built' enhancement is relatively invariant to DN and filter size.

The criterion used for selecting an appropriate filter size was the enhancement of all the small and medium 'built' areas and the edge of large 'built' areas. In light of the fact that most of the urban structures in HCMC's periphery are less than 50 m X 50 m, the filter sizes used in this project include 3X3, 5X5, 7X7, 9X9, 11X11, and the optimal filter size was determined as the one achieving the best classification accuracy by comparing the resultant image with the ground data.

The proposed classification method combines thresholding performed on both the spectral and high-pass filtered images as diagrammed in Figure 2. The thresholding can be interpreted as three steps: 1) classify all pixels with DN greater than threshold 1 (t1) as 'built' no matter how low its enhancement 2) classify all pixels with DN smaller than threshold 2 (t2) as 'non-built' no matter how high its enhancement 3) classify the remaining pixels (DN between t1 and t2) with enhancement higher than threshold 3 (t3) as 'built'.

The optimal values for the three thresholds were selected by searching in the 3-D appropriate parameter space by compar-

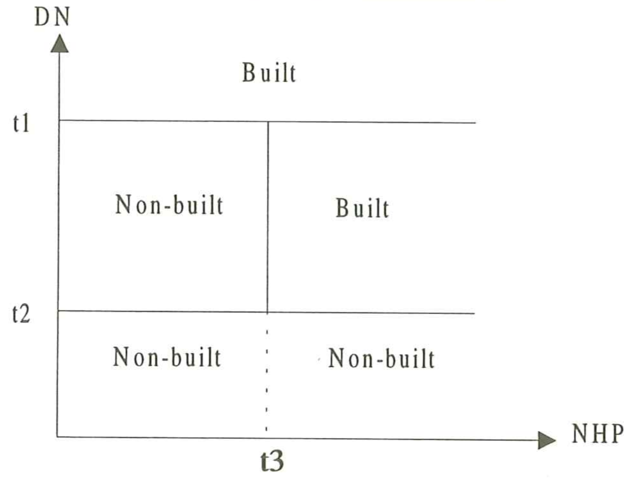

Figure 2. Schematic diagram depicting the thresholding scheme employed to combine spectral (DN) and spatial contextual (NHP: Normalized High-pass filter) data in the classification of 'built' (urban) and 'non-built' pixels.

ing the classification with the ground data. T1 is fixed as the maximum DN of 'non-built' samples. T2 starts from the minimum DN of the 'built' samples, increases 1 at each step, and stops after 30 steps (i.e. t2: DN_{min}~DN_{min}+30). T3 starts from –0.10, increases 0.01 at each step for each t2, and stops after 30 steps (i.e. t3: –0.10 ~ 0.20). For each filter size, the classification accuracy for each combination of thresholds (900 in total) was calculated and optimal thresholds were chosen as the combination maximizing the accuracy. This approach provides a threshold selection that is more objective than user specification.

With the 1994 and 2001 images classified using this method, a post-classification comparison was performed for change detection. Urban growth or the change of 'non-built' to 'built' pixel is extracted by overlaying the two classified binary images and the pixels with a value change from 0 to 1 are labeled as rural-to-urban land use change. No attempt is made to define the change from 1 to 0 since it would primarily be due to classification error.

V. RESULTS

Five normalized high-pass filters ranging from 3X3 to11X11 were applied to the two co-registered images. The effect of normalization is illustrated in Figure 3 (before normalization) and Figure 4 (after normalization). The positive correlation between DN and the enhancement has been reduced to the same range.

A random sampling and verification process with the help of visual interpretation of SPOT-PAN images, aerial photos, and field visit was performed to extract the ground data for optimal threshold selection and accuracy assessment of change detection. For the purposes of threshold selection, the sampling process is intended to include a full DN range of 'built' and 'non-built' pixels. In total, two independent samples were generated: one for the optimal threshold training and the other for accuracy assessment.

The optimal threshold selection process seeks the best balance between accurately classifying 'built' and 'non-built' pixels. For a given t1 and t2, a low t3 will favor the 'built' accuracy

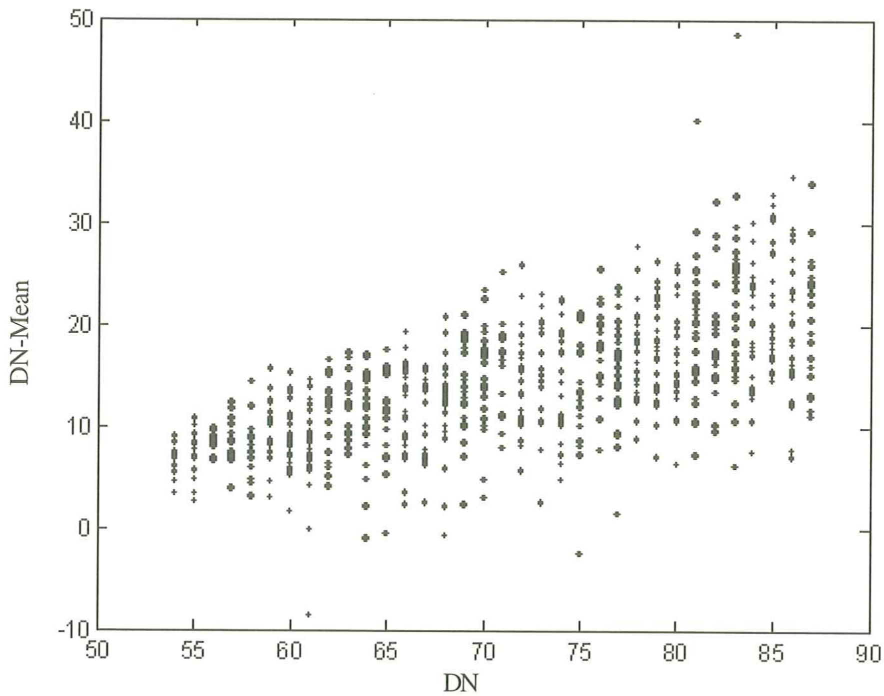

Figure 3. Scatter plot of the data's DN and enhancement before normalization

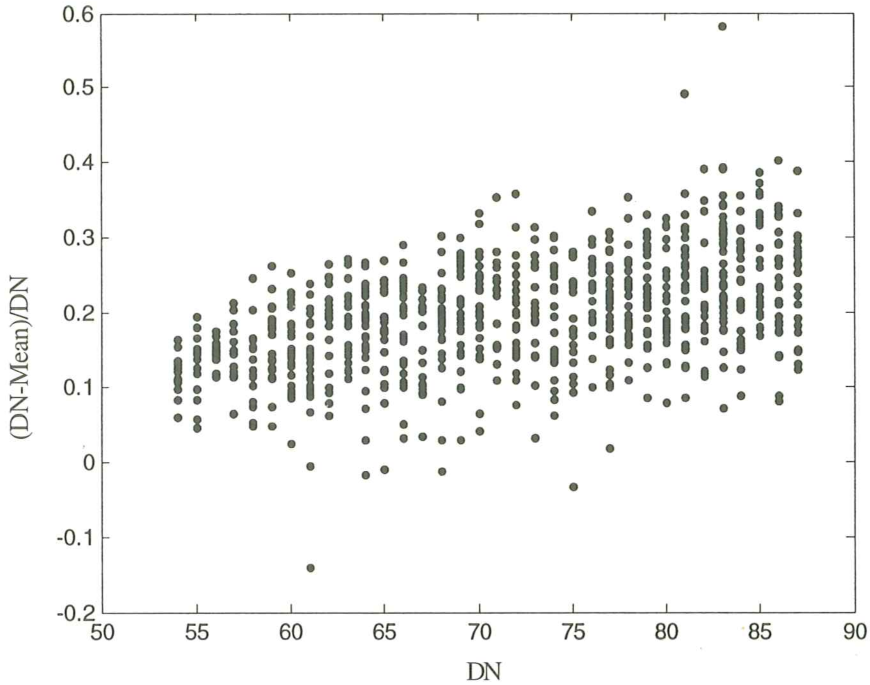

Figure 4. Scatter plot of the data's DN and enhancement after normalization.

while a high t3 will favor the 'non-built' accuracy. As a linear combination of the two accuracies, the overall accuracy will be maximized at an intermediate t3. Figure 5 shows the relationship between searching for 'built' accuracy, 'non-built' accuracy, and overall accuracy with respect to t3 for the image of 2001. As t3 increases from –0.1 to 0.3 with an increment of 0.01, the 'built' accuracy decreases from 100% to 75% while the 'non-built' accuracy increases from 12% to100%. As a linear combination of 'built' accuracy and 'non-built' accuracy, the overall accuracy starts at 29% and increases to the maximum 97.89% and then decreases to 94%. The optimal threshold 0.12 corresponds to the maximum overall accuracy.

The optimal thresholds and the corresponding maximum ac-

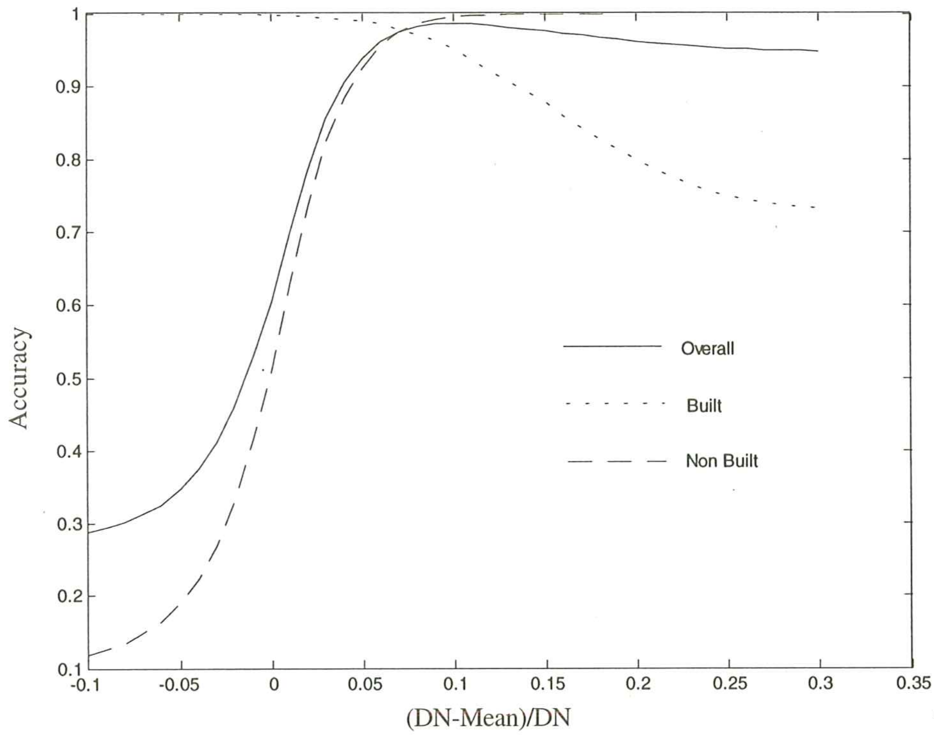

Figure 5. Classification accuracy curves of overall, built and non-built pixels with respect to t3 for a given t1, t2 and filter size

curacies for the 5 filter sizes are listed in Table 1 (1994) and Table 2 (2001), respectively. For year 1994, all five maximum training accuracies are around 96% to 98%, in which the maximum accuracy across the 5 filter sizes is 97.78% at filter size 7 X 7. All the test accuracies are around 94% to 95% with the maximum 94.97% at filter size 7 X 7. For year 2001, the maximum training and test accuracies across the 5 filter sizes are 97.89% and 94.72% at filter size 7 X 7. For each filter size, t3 in year 1994 is approximately half of that in year 2001 (except for 3 X 3). This can be explained by the fact that the SPOT-PAN image taken in 2001 has greater contrast than the 1994 image. It should be noted that although t3 is increasing with filter size, they are all in the same level so that it is possible to use the same initial value and increment. In this sense, the normalized high pass filter simplifies the threshold searching process.

The optimal thresholds (for the 1994 image, t1:DN=95, t2:DN=69, t3=0.06; for the 2001 image, t1:DN=87, t2:DN=53, t3=0.12) were applied to the whole study area. The classified images consist of binary values with '1' referring to 'built' pixels and '0' referring to 'non-built'. Change from rural to urban was detected as pixels with value '0' in 1994 and '0=1' in 2001 by comparing the two classified images. Compared to the verification sample, an accuracy of 82.31% was achieved for the final change map.

VI. CONCLUSIONS

The classification of the binary map was determined by the combined thresholding of spectral information and the spatial contrast information by the normalized high-pass filter. The normalized high-pass filter proposed here proved to have two advantageous properties: 1) it is insensitive to filter size and 2) insensitive to relative brightness. Therefore, a simplified optimal threshold procedure was successfully used to search the optimal value from a large threshold set consisting of all possible combinations. This process makes threshold determination automatic and objective and alleviates the laborious process of manually checking the image.

This method of combining spectral and spatial context information at the pixel scale was able to differentiate amongst individual pixels with low brightness value and classify the small houses of informal settlements in Ho Chi Minh City's urban fringe. As such, it has the potential to be useful to researchers and policymakers in the developing world where there has been a dearth of data about the rapid urban growth patterns that have developed during the last decade of the 20th century. In particular, developing methods to interpret the images already taken with SPOT-PAN is important because it may be the only record of these urban growth patterns as they developed during the end of the 20th century.

ACKNOWLEDGMENT

This research was funded by the Lincoln Institute of Land Policy, the University of California Vice Chancellor, and the National Science Foundation.

REFERENCES

[1] Bertaud, Marie-Agnes, 1989, The Use of Satellite Images for Urban Planning: a Case Study from Karachi, Pakistan, *World Bank Technical Note*.

[2] Gluch, R., 2002, Urban growth detection using texture analysis on merged Landsat TM and SPOT-P data. *Photogrammetric Engineering and Remote Sensing*, 68:1283-1288.

[3] Gong, P., 1993, Change detection using principal component analysis and fuzzy set theory. *Canadian Journal of Remote Sensing*, 19: 22-9.

[4] Gong, P., Ledrew, E.F., and Miller, J.R., 1992, Registration noise reduction in difference images for change detection. *International Journal of Remote Sensing*, 13: 773-779.

[5] Gong, P., and Xu, B., 2003, Remote sensing of forests over time: change types, methods, and opportunities. In *Remote Sensing of Forest Environments: Concepts and Case Studies*, edited by Woulder M. and Franklin S.E. (Amsterdam, Netherlands: Kluwer Press), pp. 301-333.

[6] Irwin, E.G., and Bockstael, N.E., 2002, Interacting agents, spatial externalities and the evolution of residential land use patterns. *Journal of Economic Geography*, 2: 31-54.

[7] Jensen, J., and Toll, D., 1982, Detecting residential land-use development at the urban fringe. *Photogrammetric Engineering and Remote Sensing*, 48: 629-643.

[8] Li, X., and Yeh, A.G.O., 1998, Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta. *International Journal of Remote Sensing*, 19:1501-1518.

[9] Mas, J., 1999, Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing*, 20: 139-152.

[10] Martin, L., 1989, Accuracy assessment of Landsat-based visual change detection methods applied to the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing*, 55: 209-

Table 1. Comparison of classification accuracy of optimal thresholds by filter size for the SPOT-PAN image taken in 1994.

Filter size	t1	t2	t3	Training Accuracy	Test Accuracy
3X3	95	72	0.04	0.9602	0.9412
5X5	95	72	0.05	0.9765	0.9427
7X7	95	69	0.06	**0.9778**	**0.9497**
9X9	95	69	0.07	0.9749	0.9469
11X11	95	69	0.07	0.9727	0.9455

Table 2. Comparison of classification accuracy of optimal thresholds by filter size for the SPOT-PAN image taken in 2001.

Filter size	t1	t2	t3	Training Accuracy	Test Accuracy
3X3	87	53	0.05	0.9688	0.9395
5X5	87	53	0.09	0.9776	0.9421
7X7	87	53	0.12	**0.9789**	**0.9472**
9X9	87	53	0.14	0.9770	0.9404
11X11	87	53	0.14	0.9708	0.9381

215.

[11] Ridd, M.K., 1995, Exploring a V-I-S (Vegetation-impervious-soil) model for urban ecosystem analysis through remote sensing: Comparative anatomy for cities. *International Journal of Remote Sensing*, 16: 2165-2185.

[12] Ridd, M. K., and Liu, J., 1998, A comparison of four algorithms for change detection in an urban environment. *Remote Sensing of Environment*, 63: 95-100.

[13] Riordan, C.J., 1980, Non-urban to urban land cover change detection using Landsat data, *Summary Report of the Colorado Agricultural Research Experiment Station*, Fort Collins, Colorado.

[14] Singh, A., 1989, Digital change detection techniques using remotely sensed data. *International Journal of Remote Sensing*, 10: 989-1003.

[15] Statistical Office of HCMC, 2001. *2000 Yearbook,* Ho Chi Minh City: Statistical Office of HCMC.

[16] Zhang, Q., Wang, J., Peng, X., Gong, P., and Shi, P., 2002, Urban built-up land change detection with road density and spectral information from multi-temporal TM data. *International Journal of Remote Sensing*, 23: 3057-3078.

[17] Zhang, Y., 2001, Detection of urban housing development by fusing multisensor satellite data and performing spatial feature post-classification. *International Journal of Remote Sensing*, 22: 3339-3355.