International Journal of Remote Sensing
Vol. 28, No. 12, 20 June 2007, 2777–2800

Per-field urban land use classification based on tax parcel boundaries

S. WU*, J. SILVÁN-CÁRDENAS and L. WANG
Department of Geography, Texas State University–San Marcos, Texas 78666, USA

(*Received 11 August 2005; in final form 30 August 2006*)

The objective of this study is to test a per-field approach for classifying detailed urban land use, such as single-family, multi-family, industrial and commercial. Tax parcel boundaries are used as the field boundaries for classification. Twelve attributes of parcels, such as parcel sizes, parcel shape, building counts and building heights, are used as the discriminant factors between different land use types. For our study area that consists of 33 025 parcels, we first derived parcel attributes from geographic information system (GIS) and remote sensing data. We then converted the parcel vector data to an image of 12 bands with pixel values from parcel attributes. After that, we performed a standard supervised classification to classify the image into nine land use types. The best classification result with a decision tree classifier had an overall accuracy of 93.53% and a Kappa Coefficient of 0.7023. This study shows the feasibility of applying a per-field approach based on tax parcel boundaries to classify detailed urban land use.

1. Introduction

Conventional land use classification in remote sensing assigns classes by pixels based on their spectral, textual, or contextual properties. In contrast to per-pixel classification, per-field classification classifies land use by pre-determined field boundaries, with the assumption that each field belongs to a single, homogeneous class (Pedley and Curran 1991, Aplin *et al*. 1999, Erol and Akdeniz 2005). Per-field classification is developed to overcome the weakness of per-pixel classification. Specifically, per-pixel classification commonly produces pixelly results, which look noisy and sometimes require post-processing to improve the outlook and classification accuracies. Furthermore, per-pixel classification has difficulties in classifying spectrally heterogeneous land use classes and identifying class boundaries when using window-based texture measures. Using high-spatial-resolution data, such as IKONOS images and LIDAR (light detection and ranging) data, may improve classification accuracies of per-pixel classification but is still limited by the weakness of per-pixel classification.

In addition to overcoming the problems of per-pixel classification, per-field classification has the advantage of allowing the incorporation of a variety of field attributes, such as the size, shape and perimeter of the field, as classification criteria. Past studies have reported improved classification results by using a per-field approach or by mixing a per-field approach with a per-pixel approach for land use and/or land cover classification (Pedley and Curran 1991, Lobo *et al*. 1996, Aplin *et al*. 1999, Dean and Smith 2003, De Wit and Clevers 2004, Erol and Akdeniz 2005).

*Corresponding author. Email: sw1020@txstate.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160600981541

The objective of the current study is to apply a per-field approach to classify detailed urban land use, such as single family, multi-family, industrial and commercial.

In per-field classification, field boundaries are pre-determined. Past per-field studies mostly relied on existing vector polygon data as field boundaries. Examples include Erol and Akdeniz (1996), Aplin *et al.* (1999), Zhan *et al.* (2000), Aplin and Atkinson (2001), Smith and Fuller (2001), and Erol and Akdeniz (2005). Some studies used image segmentation techniques to partition images into fields of homogeneous spectral or spatial characteristics. Examples include Janssen and Molenaar (1995), Lobo *et al.* (1996), Fuller *et al.* (2002), Hill *et al.* (2002), and Geneletti and Gorte (2003). Some studies delineated field boundaries by visually interpreting and manually digitizing homogeneous fields from hard-copy maps or digital images. Examples include Pedley and Curran (1991), Lobo *et al.* (1996), Berberoglu *et al.* (2000), Dean and Smith (2003), De Wit and Clevers (2004), and Lloyd *et al.* (2004). Within the above three approaches for determining field boundaries, utilizing existing vector data are preferred in that the data are ready to use. Furthermore, since existing vector data usually come from field surveys and/or photo interpretation, they provide a satisfactory degree of accuracy and precision, and more meaningful field boundaries than segmentation techniques in this case.

After field boundaries are determined, researchers can utilize field boundaries for classification in two ways. The first is to utilize field boundaries to derive field attributes in a pre-classification stage. Researchers obtain field attributes, such as image spectral or texture statistics within fields, to use as discriminant criteria for classification. Examples include Pedley and Curran (1991), Erol and Akdeniz (1996), Lobo *et al.* (1996), Smith and Fuller (2001), Dean and Smith (2003), Lloyd *et al.* (2004) and Erol and Akdeniz (2005). The second way is to utilize field boundaries in a post-classification stage. After an initial per-pixel classification, the majority class of pixels within a field is assigned to all pixels within the field. Examples include Janssen and Molenaar (1995), Aplin *et al.* (1999) and Aplin and Atkinson (2001). Berberoglu *et al.* (2000) compared the two approaches and argued that utilizing field boundaries in a post-classification stage provided a better result.

In addition to image statistics within fields, various field attributes can be used for classification. For instance, Weiler and Stow (1991) used field size as a parameter to characterize different types of urban land use. De Wit and Clevers (2004) used field areas and shapes as the criteria to reassign land classes in a post-processing stage. Based on the percentages of build-up area, green space, and water body within fields that were initially classified from SPOT images, Zhan *et al.* (2000) classified different types of land use. Smith and Fuller (2001), Fuller *et al.* (2002) and Hill *et al.* (2002) utilized field attributes in both pre-classification and post-classification stages. They first obtained the mean spectral reflectance statistics within fields and classified land use according to the statistics. They then conducted a knowledge-based correction to modify land use classes based on other field statistics. The field statistics they used include the class probability, the classes of surrounding fields, the mean elevation, the modal slope, the modal aspect, the building area percentage, the building height, the canopy height and the terrestrial/marine cover types. Many of them were derived from datasets other than remote sensing images, such as elevation data, buildings data, and terrestrial/marine data.

Geneletti and Gorte (2003) combined per-pixel and per-field classifications to maximize the classification accuracy. They first performed different levels of image segmentation to establish different levels of field boundaries. They then performed a regular per-pixel land cover classification based on the spectral reflectance of pixels. After that, in a sequential manner from coarse to fine scales of segmentation, they reassigned the majority class within fields to pixels based on certain class percentage thresholds. The final result was that some pixels were not reassigned to the majority class but remained in their initial classes from per-pixel classification.

2. Methodology

In the previous section, we reviewed three common sources of field boundaries for per-field classification, from existing vector data, from image segmentation, and from manually digitizing. In the present study, we also used a per-field approach for urban land use classification. The Travis County Appraisal District (TCAD) tax parcel boundaries were used as the field boundaries. The advantage of using tax parcel boundaries is that individual tax parcels are relatively small and always contain the same type of land use. In contrast, common image segmentation techniques are incapable of identifying fields of homogeneous urban land use, and manually digitizing is too time-consuming to be a practical approach for delineating homogeneous fields in an urban environment. Nevertheless, a limitation of using tax parcel boundaries is that they do not cover city street surface areas. As a result, our land use classification only focused on classifying land use types from the tax parcel areas, and street surfaces were not considered.

Our per-field classification is different from past per-field studies in a number of ways. First, the goal of the current study is to classify detailed urban land use classes. In the past, most per-field studies classified spectrally homogeneous land cover types, and a few of them classified residential from non-residential urban land use, yet none of them classified detailed urban land use classes, such as single family, multi-family, industrial and commercial.

Second, our per-field classification of land use is conceptually different from standard land use classification in remote sensing. It focuses on the classification of urban land use from tax parcels instead of remote sensing images. In other words, the parcel data are the primary data, and the imagery data are the ancillary data used to calculate parcel attributes. In addition, rather than emphasizing imagery parameters for classification, our per-field classification emphasizes using geographic information system (GIS) parameters as discriminant criteria for classification.

Third, in order to classify detailed urban land use, we test a variety of field attributes derived from GIS and image data, such as building count, building shape and closeness to major roads, most of which have not been used in past studies of per-field classification.

3. Study area and data sources

The City of Austin, the capital of Texas, USA, provides a suitable environment for exploring the proposed land use classification method. The city is not too large in land area (approximately 250 square miles in 2005), yet with a variety of residential and non-residential land use, old and new neighbourhoods, and housing patterns. Its land use and housing patterns have changed considerably during the past 15 years. During the period from 1990 to 2000, the city has grown by 41%, from a population of 465 622 in 1990 to 656,562 in 2000, with an annual growth rate of 3.5%. During the second period from 2000 to 2005, the annual growth rate slows down to 1.20%. For the next 30 years, the city plans to maintain a steady annual growth rate between 1.20% and 2.00% (City of Austin 2005e).

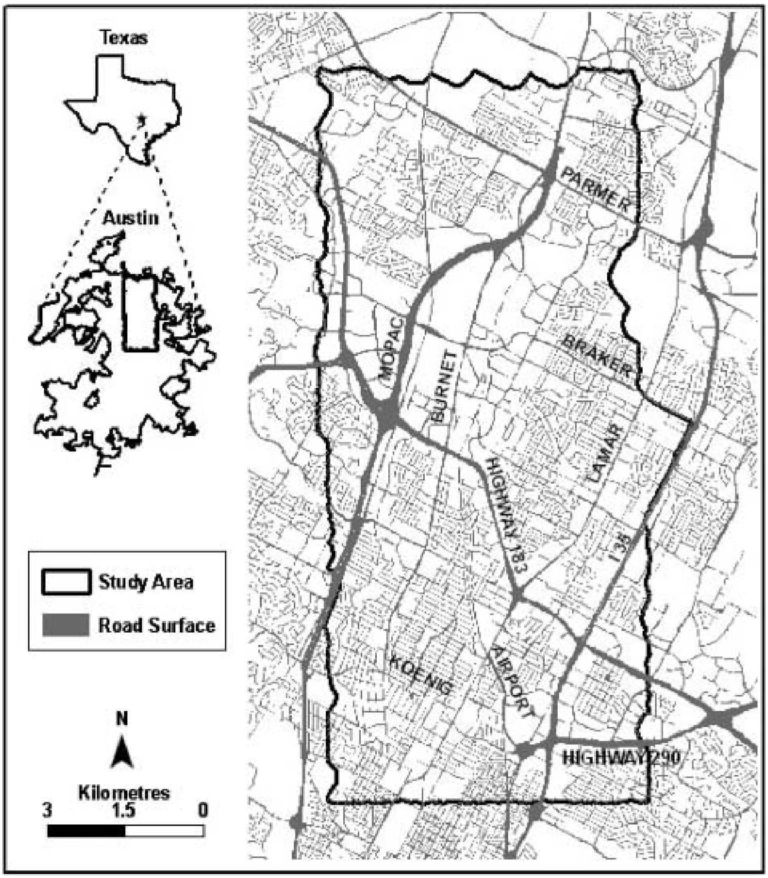

Figure 1. Study area in Austin, Texas.

In order to test per-field land use classification, we selected an area of approximately 6 by 14 km in the north central part of the City of Austin (figure 1). The city's three main thoroughfares, IH-35, MoPac, and Highway 183 run through this area. In general, non-residential land use areas are close to major roads and residential land use areas are in between (figure 2).

We obtained tax parcel boundaries data, four ancillary datasets that would be used for deriving parcel attributes, and ground truth land use data from the City of Austin Neighborhood Planning and Zoning Department (NPZD), either directly downloaded from their FTP server (City of Austin 2005a) or acquired through personal contact. The four ancillary datasets, building data, elevation data, street data and image data, are for the purpose of calculating parcel attributes. The ground truth land use data are for selecting training samples and for evaluating classification results.

The tax parcel boundaries data are in vector polygon format and is up-to-date till the year 2005. The building data are building footprints in vector polygon format and contains information of the average altitude for individual building roofs. The elevation data are 0.61 m (2 ft) contour lines, which measures the elevation for ground surface. The building data and the elevation data were generated by the Analytical Surveys Incorporation (ASI) contracted with the City during the year of 2003. ASI first manually digitized building footprints from aerial photographs, then

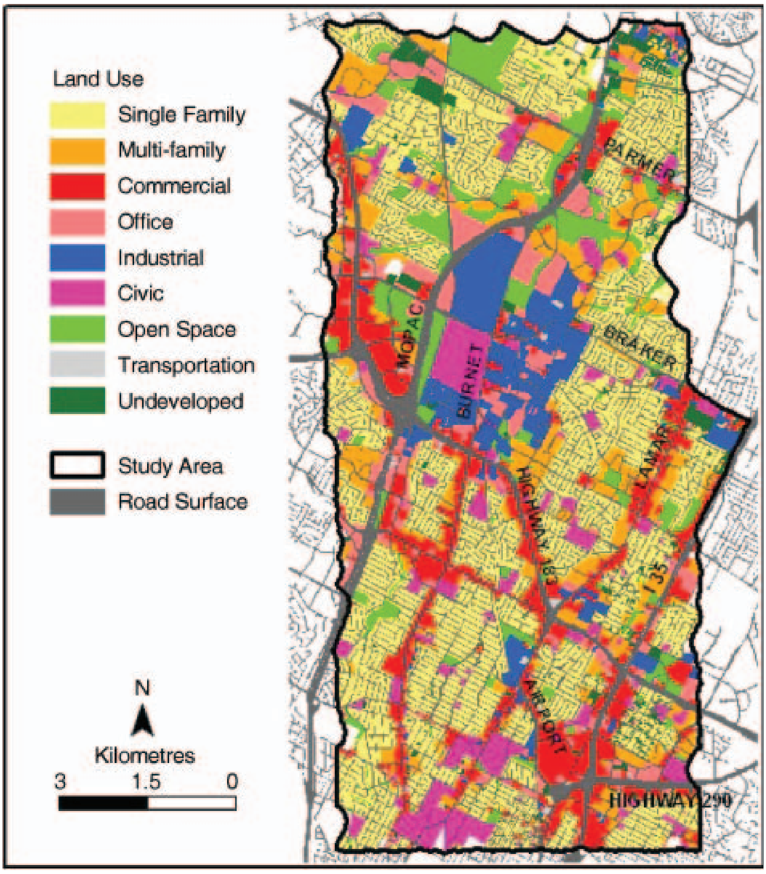

Figure 2. Land use in the study area.

estimated the altitude for individual building roofs by referencing with LIDAR elevation data. From the LIDAR data, the ground surface elevation data were also generated

The street data are in vector line format. It is originated from the US Census 2000 Topologically Integrated Geographic Encoding and Referencing (TIGER) street data, and the City has updated it till the year 2004. The street data contain the census feature class codes (CFCCs), which represent street categories with different levels or sizes of streets (U.S. Census Bureau 2000). The highest level of streets has a CFCC starting with a number of one and includes primary highways. The second level of streets has a CFCC starting with a number of two and includes primary roads. The third level of streets has a CFCC starting with a number of three and includes secondary roads. The fourth level of streets has a CFCC starting with a number of four and includes local, neighbourhood, and rural roads.

The image data are 0.61 m spatial resolution, three band (green, red and near-infrared) colour infrared (CIR) digital orthophotos. The source aerial photographs were taken by ASI during the year of 2003.

The ground truth land use data are in vector polygon format, which is up-to-date till the year 2003. The City updates land use data based on a variety of sources of information, including historical land use data, TCAD tax parcel data, city parcels database, natural preserves GIS data, aerial photographs, building footprint data and field check (City of Austin 2005b). NPZD divides the City's land use into 16 general land use types, which are further divided into 37 detailed subtypes (City of Austin 2005c). We merged the 16 general land use types to a classification scheme of nine land use types (table 1). Some land use types merged into larger land use categories are either rare in the study area (such as land use of mobile homes, large-lot single family, mining, and utilities) or are not applicable to the tax parcel areas (such as land use of streets, water, and unknown).

4. Processes

We tested 12 parcel attributes derived from the ancillary datasets as discriminant factors for land use classification. The 12 parcel attributes are parcel size, parcel shape compactness, the number of buildings, the maximum building's area, the standard deviation of the building's area, the total building-area percentages, the maximum building's height, the standard deviation of the building's height, the maximum building shape compactness, the standard deviation of building shape compactness, the highest category of streets within 50 m and the impervious cover percentage. The shape compactness measure of buildings or parcels was calculated by dividing the area with the square of the perimeter (Schalkoff 1989). The more curved shape thus has a smaller compactness measure.

The 12 parcel attributes were selected for theoretical and/or empirical reasons. After observing land use parcel maps (figures 3 to 11) and relevant statistics (table 2 and figures 12 and 13), we can see that each parcel attribute is distinctive for one or more land use types. For example, single-family land use parcels are usually small and rectangular, and they mostly contain only one building (figure 3). Therefore, the attributes of parcel size, parcel shape compactness, and the number of buildings may separate them from others. Compared with buildings in civic land use parcels,

Table 1. Land use classification scheme of nine land use types.

Land use type	Descriptions
Single family (Sf)	Mobile homes, large-lot single family, single-family detached and two-family attached, duplex
Multi-family (Mf)	Three/fourplex, apartment/condo, group quarters, and retirement housing
Commercial (Com)	Retail and general merchandise, apparel and accessories, furniture and home furnishings, grocery and food sales, eating and drinking, auto related, entertainment, personal services, lodgings, building services
Office (Off)	Administrative offices, financial services (banks), medical offices, research and development
Industrial (Ind)	Manufacturing, warehousing, equipment sales and service, recycling and scrap, animal handling, mining facilities
Civic (Civ)	Semi-institutional housing, hospital, government services, educational facilities, meeting and assembly facilities, cemeteries, day care facilities
Open space (Open)	Parks, recreational facilities, golf courses, preserves and protected areas, water drainage areas and detention ponds
Transportation (Tran)	Railroad facilities, transportation terminal, aviation facilities, parking facilities, utilities facilities
Undeveloped (Und)	Vacant land and land under construction

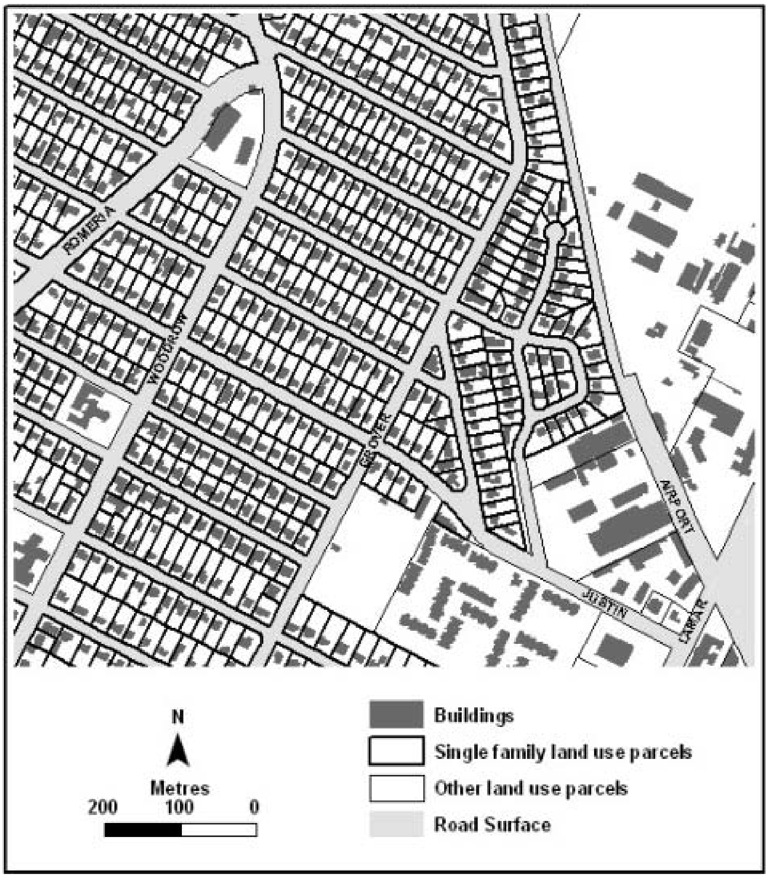

Figure 3. Examples of single-family land use parcels.

buildings in multi-family land use parcels are relatively small and uniform in size (figures 4 and 5). Moreover, multi-family land use parcels averagely have more building areas than do civic land use parcels (table 2). Therefore, the attributes of the maximum building's area, the standard deviation of the building's area, and the total building-area percentage may be used to distinguish one land use from the other. Commercial land use and office land use parcels have similar statistics of the average number of buildings (1.5 and 1.42 respectively), yet buildings in commercial land use parcels usually have lower and more uniform heights than those in office land use parcels (table 2, figures 6 and 7). Therefore, the attributes of the maximum building's height and the standard deviation of the building's height may be used to distinguish one land use from the other. Multi-family land use and industrial land use parcels have similar parcel sizes (table 2, figures 4 and 8), yet multi-family land use parcels have higher standard deviation of building shape compactness and higher values of the maximum building shape compactness. It is partly because they usually have many buildings within parcels. Commercial land use parcels are usually close to major roads and have the highest category of street within 50 m (table 2). Open-space land use, undeveloped land use and transportation land use parcels usually have no buildings (table 2), yet transportation land use parcels usually have higher impervious cover percentages than the other two types of land use parcels.

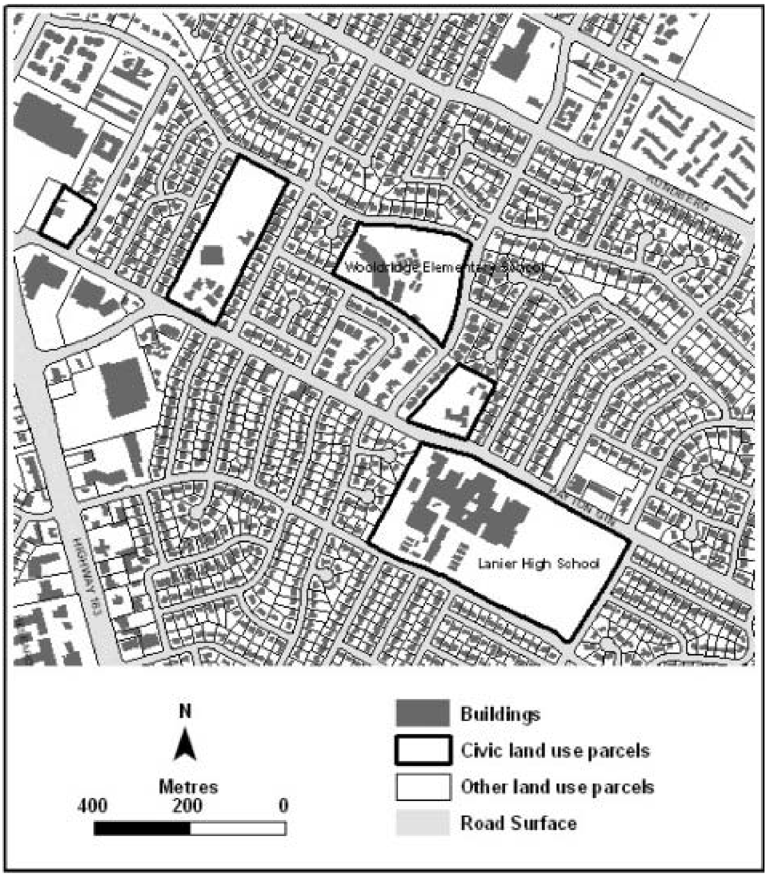

Figure 4. Examples of civic land use parcels.

We derived most of the parcel attributes by overlaying parcel boundaries with the ancillary datasets in a GIS. To calculate the building height parameter, we first estimated the mean building-floor elevations for individual buildings from the elevation data, then subtracted building-floor elevations from respective building-roof altitudes to obtain height statistics for individual buildings.

Based on the CFCCs of street data, we calculated the street-category statistic for parcels. The statistic has a number ranging from one to four based on the highest category of streets within 50 m buffers of parcels. The distance was decided subjectively by considering the relevant distances between parcels and streets as well as the relevant sizes of parcels.

One of the reasons to consider the parameter of impervious cover percentage is that the City of Austin has site development regulations regarding the maximum impervious cover allowed for specific zoning districts (City of Austin 2005d). Therefore, it is expected that this parameter would help characterize different land use types. We classified the impervious cover versus pervious cover from aerial photographs in a two-step procedure. In the first step, we computed the normalized difference vegetation index (NDVI) for the entire area. A threshold of 0.3 was chosen to discern between vegetated and non-vegetated areas. We then categorized the vegetated areas as the pervious cover. In the second stage, the non-vegetated

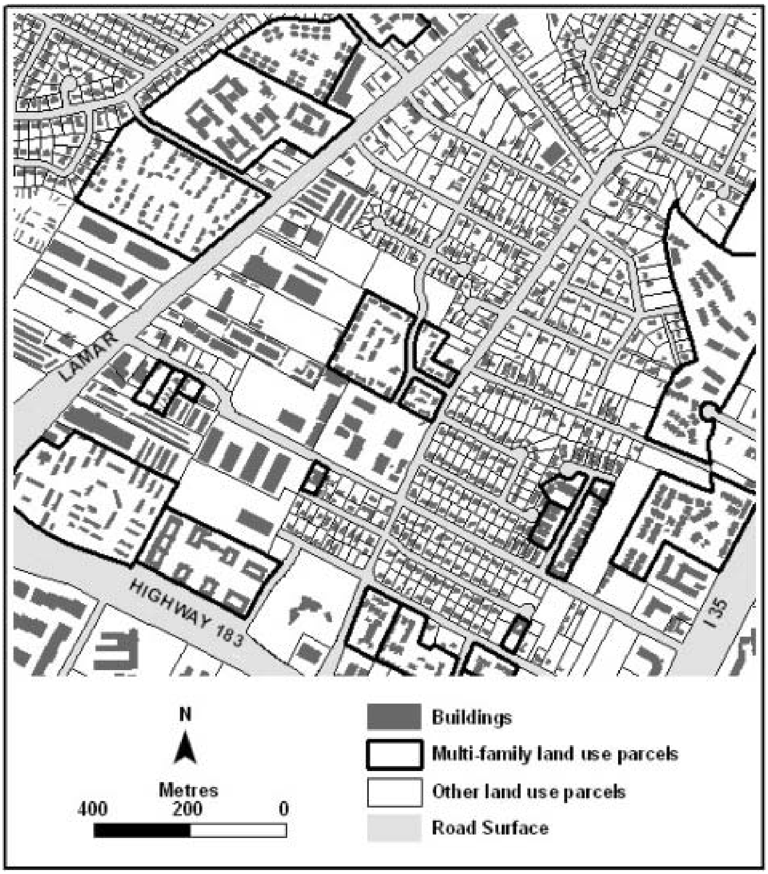

Figure 5. Examples of multi-family land use parcels.

areas were grouped into 32 clusters using the ISODATA unsupervised clustering algorithm. We then manually classified the 32 clusters into either pervious or impervious classes. Using a large number of clusters allowed the separation of bare land, which was spectrally similar to some impervious cover but should be categorized as pervious cover.

It is hypothesized that the 12 parcel attributes are related to land use types. Our goal, therefore, is to test how well these parcel attributes collectively characterize different land use types. We converted the parcel vector data to a parcel image and used a standard supervised classification approach in remote sensing to classify land use from the image. This approach was adopted for two reasons. First, image classification as a major topic in remote sensing has been studied by many researchers, and a standard and rigorous procedure for image classification and result assessment has been well established. Second, there have been many advanced image classification algorithms developed. It would be of advantage to experiment with these algorithms and compare their performance in the land use classification based on the GIS attributes of parcels.

The derived parcel image has 12 bands from the 12 parcel attributes. The image pixel values correspond to parcel attribute values at the same geographic locations. After considering choosing a fine spatial resolution that can represent small parcels,

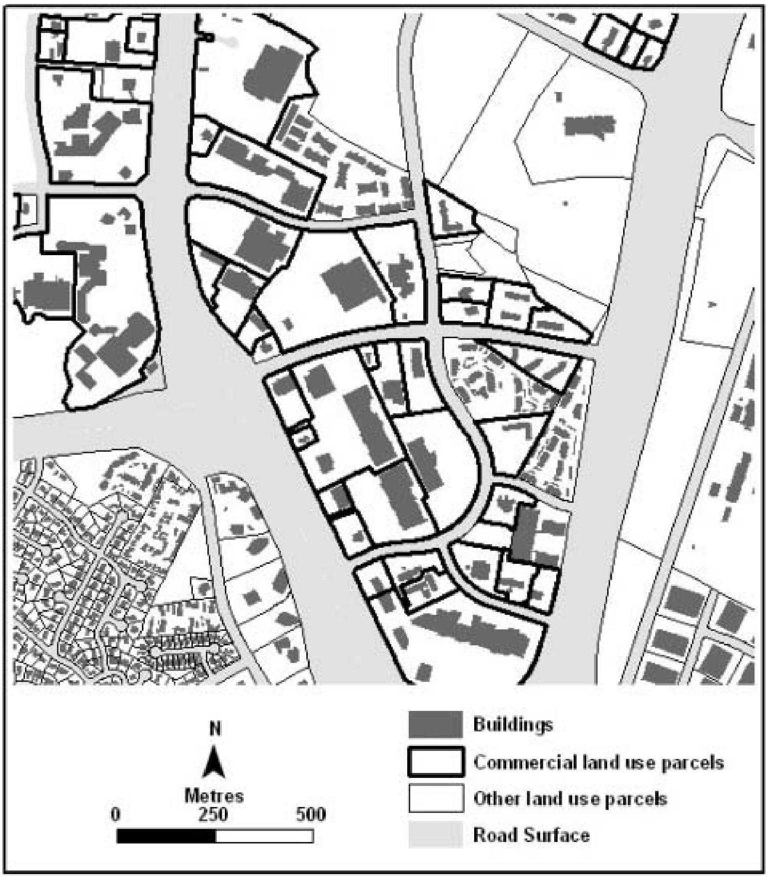

Figure 6. Examples of commercial land use parcels.

we subjectively decided a spatial resolution of 4.9 m (16 ft) for the parcel image. Since only the centroids of parcels were used as the sample points for accuracy assessment, the choices of spatial resolution would not have any effect on the classification results.

Since we are interested in how many numbers of parcels, instead of how many parcel areas, are correctly classified from each land use, using parcel attributes as discriminant criteria, we decided the number of training sample points for different land use proportional to their respective parcel numbers instead of parcel areas. Specifically, we first referenced parcels with the ground truth land use so that each parcel has information regarding its ground truth land use. Then we randomly selected approximately half of the parcels from each land use as training parcels and the other half as test parcels for accuracy assessment. In total, 16 506 training parcels and 16 519 test parcels were determined (table 3). The centroids of training parcels were used as the training points to classify the parcel image.

5. Classification results and assessment

We experimented with a variety of classification algorithms, including minimum distance, parallelepiped, spectral angular mapper, mahalanobis distance, maximum

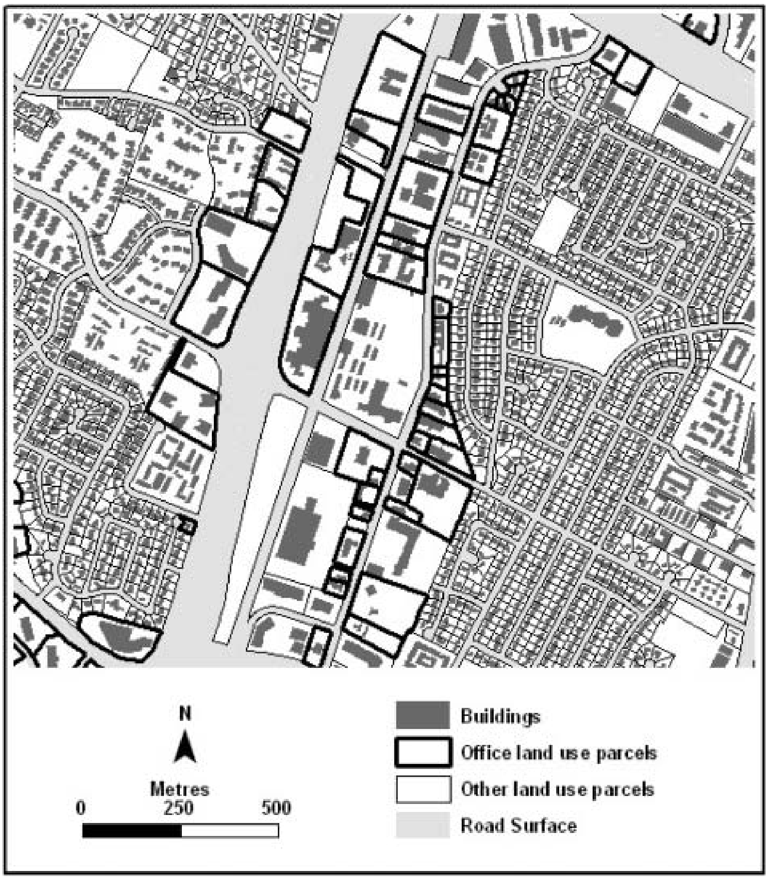

Figure 7. Examples of office land use parcels.

likelihood, binary encoding, neural network and decision tree. We used the 16 519 test parcel centroids to perform accuracy assessment. The best classification result was obtained when using a decision tree classifier with an overall accuracy of 93.53% and a Kappa coefficient of 0.7023 (tables 4 and 5 and figure 14). After observing the histograms of each parcel attribute for different land use (figures 12 and 13), we understand that some of the parcel attributes are not normally distributed. Since a decision tree classifier does not require assumptions regarding the statistical properties of the input data and is capable of handling both numeric and categorical inputs (McCauley and Goetz 2004), it is reasonable that the classifier produced the best result. Similarly, other non-statistical algorithms, such as parallelepiped and neural network, generated better results (with Kappa coefficients of 0.23 and 0.48 respectively) than algorithms that require statistical assumptions, such as Mahalanobis distance and maximum likelihood (with Kappa coefficients of 0.21 and 0.16 respectively).

A decision tree classifier groups data into hierarchical structures through a process of recursive partitioning of predictor variables into smaller, more homogeneous groups. The partitioning process is based on predetermined decision tree rules. We used the QUEST algorithm (Loh 2005) to automatically generate decision tree rules. QUEST stands for quick, unbiased and efficient statistical tree. It produces binary

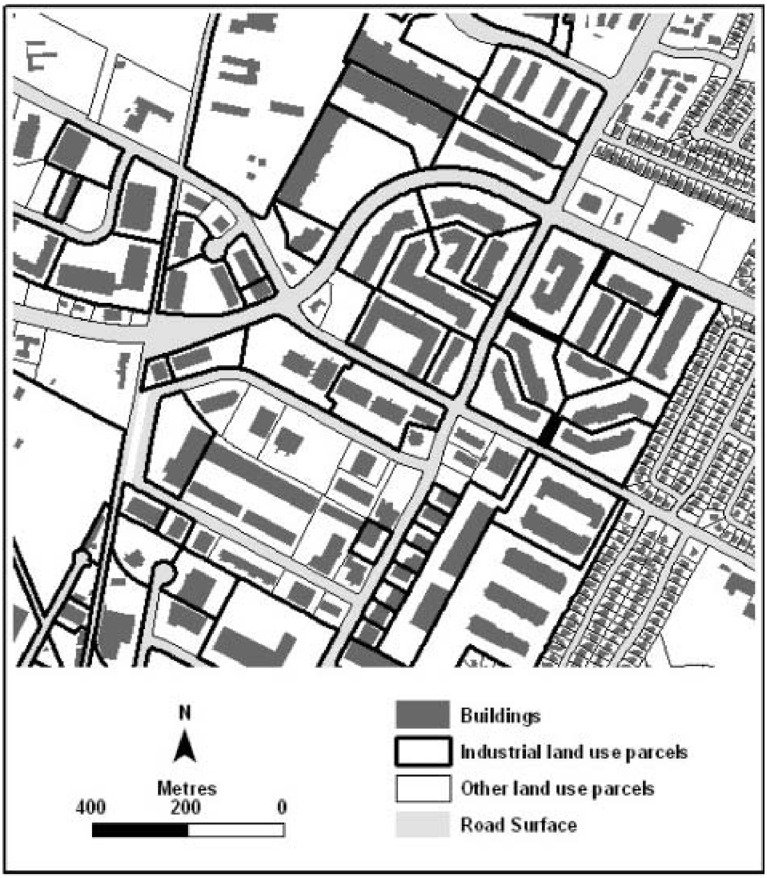

Figure 8. Examples of industrial land use parcels.

decision rules by examining all possible binary splits of the data along each predictor variable to select the split that most reduces node impurities (Loh and Shih 1997).

The overall accuracy is high, showing that approximately 93% of test parcels are correctly classified. On the other hand, the Kappa coefficient is relatively low compared with the overall accuracy. The considerable difference indicates that classification accuracies for different land use types are very uneven. Specifically, single family land use, which has high producer's and user's accuracies (both above 98%), has a large number of parcels (approximately 88% of the total), while other land use types, which have relatively low producer's and user's accuracies (mostly between 60% and 80%), only have small percentages of parcels. The large number and the high classification accuracies of single-family parcels have a great influence on the overall accuracy. In contrast, the Kappa coefficient is balanced by the low accuracies of other land use types and is, therefore, relatively low.

The reason that single family land use has good classification results is probably because of its relatively uniform parcels with distinctive characteristics (small, rectangular and with a single low-rise building). Next to single family land use, undeveloped land use has the best classification results with both producer's and user's accuracies above 75%. The land use of undeveloped, open space and transportation are similar in that most of their parcels do not have buildings.

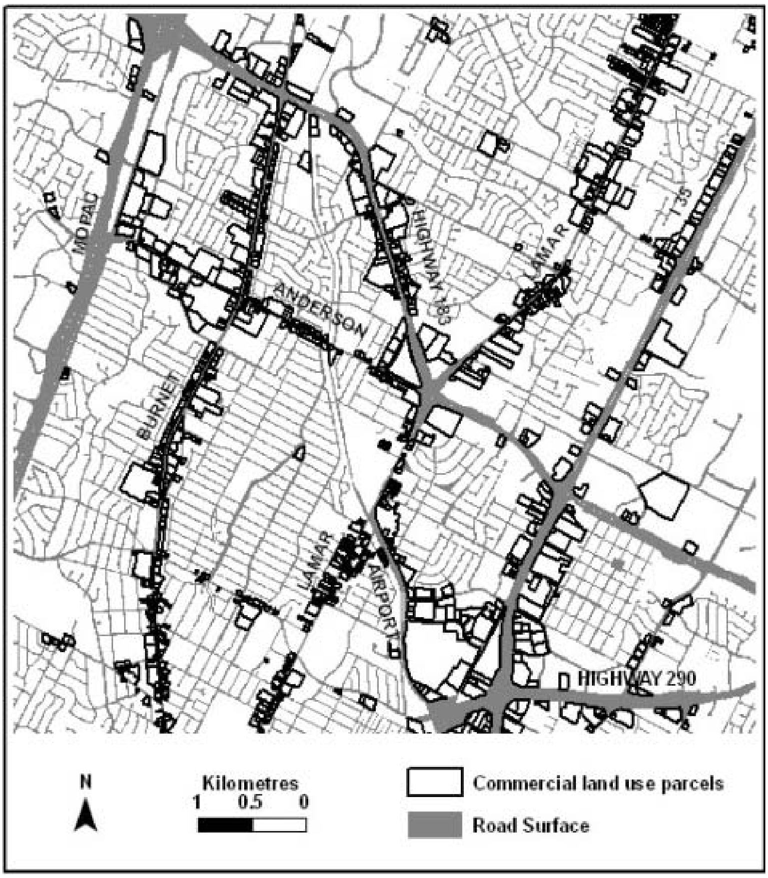

Figure 9. Examples of commercial land use parcels and their relationships to major roads.

Nevertheless, undeveloped land use has a relatively high producer's accuracy (87.63) compared with those of open space (53.89) and transportation (30.16). Undeveloped land use also has more parcels (372 in total) than open space (167) and transportation (63). In nature, undeveloped land use is heterogeneous in parcel attributes. Therefore, open space land use and transportation land use have high percentages (38.32% and 57.14% respectively) of parcels misclassified into undeveloped land use. Nevertheless, undeveloped land use has a large number of unique parcels that are small in size and contain land under construction, which explains its high classification accuracies.

Transportation land use has the lowest classification accuracies (30.16% and 50.00% for producer's and user's accuracies respectively) among the three non-building types of land use, which is probably owing to its heterogeneous nature of parcel attributes combined with its small number of test parcels (63, which is the fewest). For example, the impervious cover percentage of transportation land use parcels is usually low for the railroad facilities while high for the parking facilities, and the parcel sizes are usually small for a parking lot and large for a bus transfer centre.

There are six land use types that generally have one or more buildings within parcels, including single-family, multi-family, commercial, office, industrial and

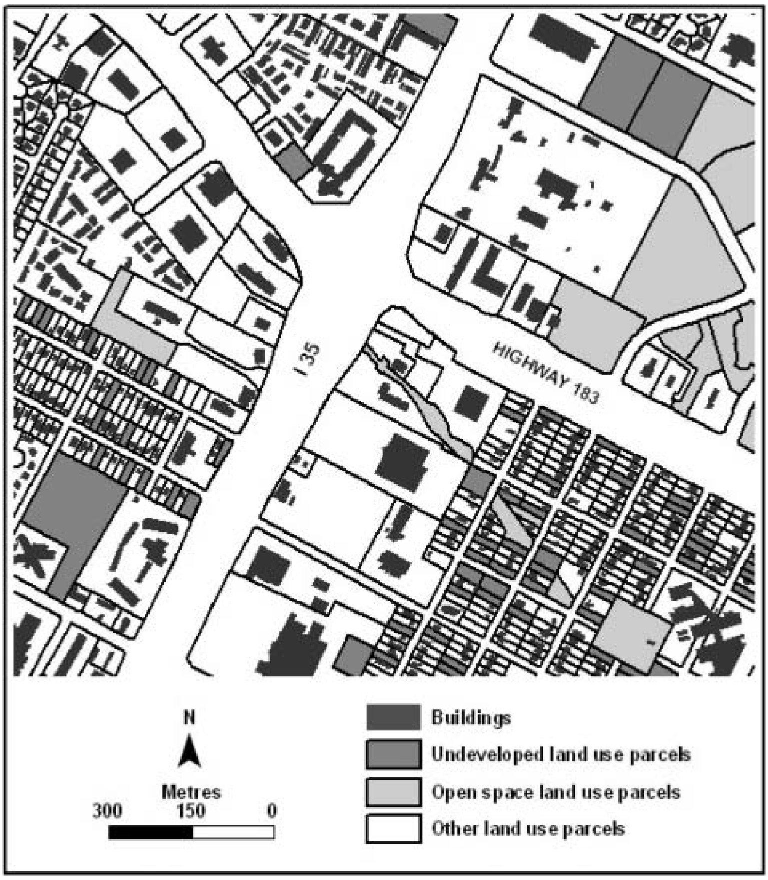

Figure 10. Examples of open space and undeveloped land use parcels.

civic. Among them, the types of office, industrial, and civic have the lowest producer's accuracies (8.3%, 20% and 5.13% respectively). Most of the errors come from misclassification into commercial land use, which is probably owing to the heterogeneous nature of commercial land use parcels that have attributes similar to the three. Specifically, commercial land use parcels have a wide range of parcel sizes and building numbers, from a small retail store parcel with a single low-rise building to a large shopping centre parcel with a number of high-rise buildings.

Besides single-family land use, multi-family land use is the only one that has both producer's and user's accuracies above 50% (66.20% and 56.63% respectively) among the six building types of land use. Compared with other building types of land use (except single-family land use), multi-family land use is relatively homogeneous. Apartments and condos consist of approximately 79% of multi-family land use parcels, which usually have numerous buildings within parcels and are thus distinctive from other land use types.

In summary, the classification results show that the six building types of land use can be differentiated from the three non-building types of land use. Among the six building types of land use, office, industrial, and civic have lower classification accuracies, mainly owing to their heterogeneous nature. On the other hand, single family land use can be separated well from others because of its distinctive parcel

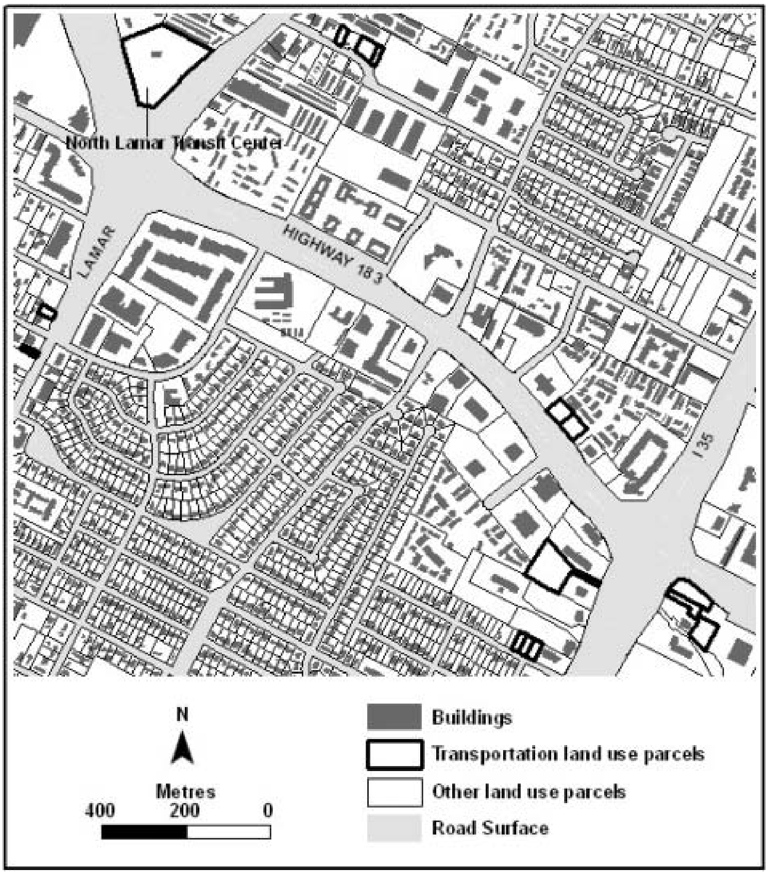

Figure 11. Examples of transportation land use parcels.

attributes. Among the three non-building types of land use, the classification accuracy for undeveloped land use is satisfactory because it has a large number of unique parcels that belong to land under construction.

To investigate further which parcel attributes help land use classification better, we calculated signature separabilities for training samples based on individual, as well as combinations of, parcel attributes (table 6). Signature separability measure was based on Bhattacharrya Distance, also called Jeffries-Matusita Distance. The measure is a real value between '0' and '2', where '0' indicates complete overlap between the signatures of two classes and '2' indicates a complete separation between the two classes. As we can see from table 6, the average separability based on all 12 parcel attributes has a satisfactory measure of 1.88. The eight building-relevant attributes (Nos 3 to 10) contribute to the overall separability with a combined separability measure of 1.72, while the parcel's two geometric attributes (Nos 1 and 2) contribute with a combined measure of 0.28. Within those eight building-relevant attributes, the three related to building areas (Nos 4 to 6) are most important (with a combined measure of 1.13), next the two related to building heights (Nos 7 and 8) (with a combined measure of 0.52), then the two related to building shape compactness (Nos 9 and 10) (with a combined measure of 0.36), and lastly the one related to the number of buildings (No. 3) (with a measure of 0.29). Furthermore, within the three

Table 2. General statistics of 12 parcel attributes (see table 6) by nine land use types (see table 1 for abbreviations).

No.	Sf	Mf	Com	Off	Ind	Civ	Open	Tran	Und
1	M=98	M=1976	M=743	M=1181	M=2176	M=3574	M=2052	M=313	M=276
	S=81	S=2411	S=1561	S=2672	S=8309	S=8663	S=4205	S=428	S=991
2	M=1203	M=1303	M=1236	M=1242	M=1266	M=1233	M=1568	M=1293	M=1233
	S=65	S=293	S=148	S=148	S=182	S=151	S=503	S=190	S=79
3	M=1.03	M=9.79	M=1.5	M=1.42	M=1.97	M=3.69	M=0.06	M=0.02	M=0.01
	S=0.21	S=11.53	S=1.55	S=1.04	S=2.48	S=7.18	S=0.30	S=0.12	S=0.07
4	M=23	M=106	M=144	M=180	M=401	M=331	M=1.0	M=1.3	M=0.2
	S=7	S=101	S=302	S=339	S=775	S=436	S=5.2	S=12.1	S=0.6
5	M=22	M=83	M=125	M=165	M=342	M=200	M=0.9	M=1.3	M=0.2
	S=7	S=89	S=227	S=312	S=384	S=241	S=4.6	S=12.1	S=0.6
6	M=25.81	M=30.01	M=25.41	M=23.43	M=31.11	M=18.33	M=0.11	M=0.15	M=0.03
	S=8.04	S=8.84	S=13.16	S=10.63	S=13.94	S=10.94	S=1.37	S=1.60	S=0.43
7	M=16.41	M=28.48	M=19.86	M=26.49	M=22.09	M=29.38	M=0.87	M=0.32	M=0.07
	S=4.17	S=10.03	S=10.58	S=16.99	S=6.32	S=17.87	S=3.87	S=2.60	S=0.90
8	M=16.38	M=24.59	M=18.96	M=25.09	M=21.00	M=24.28	M=0.84	M=0.32	M=0.07
	S=4.14	S=7.58	S=9.20	S=14.72	S=5.99	S=12.82	S=3.75	S=2.60	S=0.90
9	M=1256	M=1801	M=1398	M=1387	M=1426	M=1777	M=64	M=19	M=9
	S=93	S=510	S=297	S=278	S=329	S=596	S=278	S=150	S=131
10	M=1254	M=1609	M=1360	M=1358	M=1361	M=1490	M=64	M=19	M=9
	S=92	S=420	S=244	S=253	S=217	S=288	S=276	S=150	S=131
11	M=45	M=37	M=31	M=33	M=38	M=34	M=38	M=36	M=43
	S=6	S=9	S=11	S=11	S=9	S=8	S=10	S=12	S=9
12	M=42	M=54	M=63	M=59	M=68	M=52	M=36	M=56	M=34
	S=10	S=9	S=12	S=13	S=14	S=12	S=15	S=15	S=14

Note: M=mean, S=standard deviation.

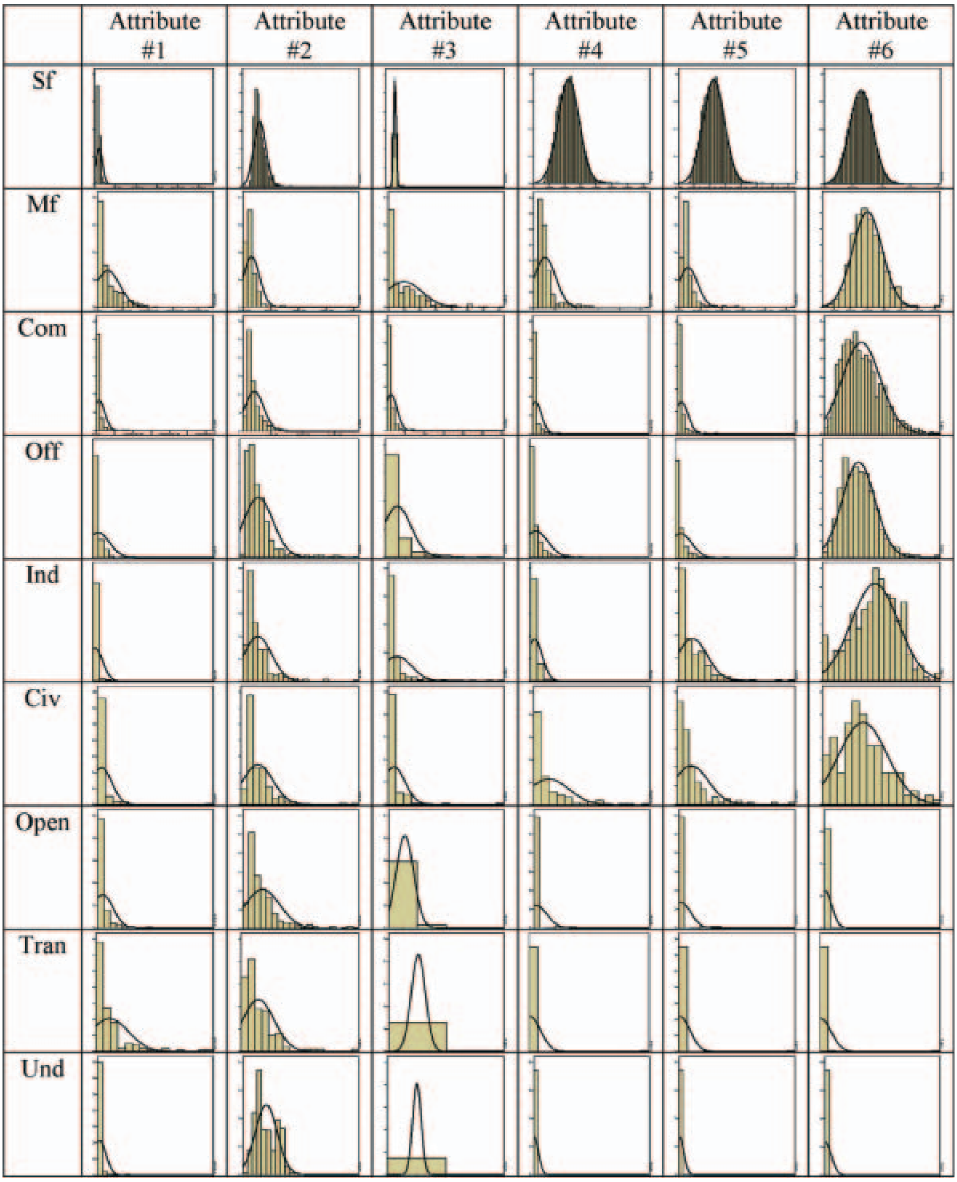

Figure 12. Histograms of parcel attributes Nos 1 to 6 (see table 6) by land use types (see table 1 for land use abbreviations).

building-area-related attributes (Nos 4 to 6), the standard deviation of the building's area (No. 5) and the maximum building's area (No. 4) are more important, with separability measures of 0.67 and 0.61 respectively.

The parcel attribute of the highest category of street within 50 m (No.11) does not provide much separability (only 0.04) between land use types as we expected. From figure 2, it is observed that although most commercial land use and some multi-family land use parcels are related to this attribute, there is still great variability of this attribute for other land use types, such as office and industrial.

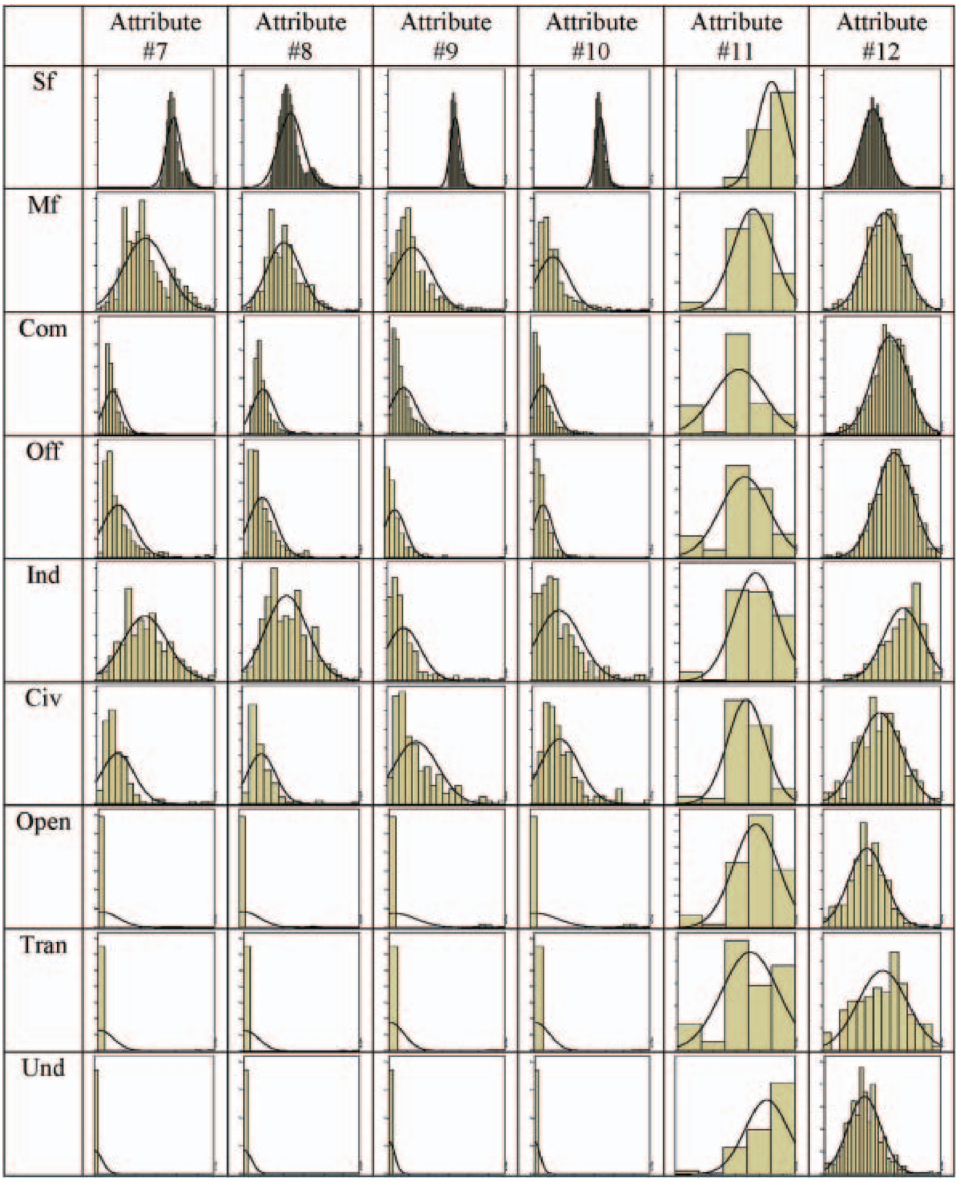

Figure 13. Histograms of parcel attributes Nos 7 to 12 (see table 6) by land use types (see table 1 for land use abbreviations).

The parcel attribute of the impervious cover percentage (No. 12) also has relatively low separability measure (0.09). The reason is probably that the impervious cover is classified based on the spectral reflectance of remote sensing images, and many impervious cover surfaces covered under tree canopies are erroneously classified into the pervious cover class. Hence, we can conclude that this vegetation-relevant parameter does not provide good separability between parcels of different land use types. In order to test how well the real impervious cover percentage parameter relates to land use types, we need more accurate methods for the delineation of the impervious cover surface.

Table 3. Number of training and test parcels for each land use type.

Land use type	No. of training parcels	No. of test parcels	Total No. of parcels
Single family	14 636	14 616	29 252
Multi-family	192	213	405
Commercial	647	639	1286
Office	232	241	473
Industrial	142	130	272
Civic	72	78	150
Open space	145	167	312
Transportation	69	63	132
Undeveloped	371	372	743
All land use types	16 506	16 519	33 025

Table 4. Confusion matrix (see table 1 for class abbreviations).

	Ground truth (pixels)									
Class	Sf	Mf	Com	Off	Ind	Civ	Open	Tran	Und	Total
Sf	14 340	16	89	69	13	12	9	0	2	14 550
Mf	34	141	27	14	12	20	1	0	0	249
Com	240	50	484	126	53	36	1	0	0	990
Off	0	1	23	20	24	4	0	0	0	72
Ind	2	5	15	12	26	1	0	0	0	61
Civ	0	0	1	0	1	4	0	0	0	6
Open	0	0	0	0	0	0	90	8	29	127
Tran	0	0	0	0	1	1	2	19	15	38
Und	0	0	0	0	0	0	64	36	326	426
Total	14 616	213	639	241	130	78	167	63	372	16 519

Table 5. Error and accuracy assessment (see table 1 for class abbreviations).

Class	CE (%)	OE (%)	CE (pixels)	OE (pixels)	PA (%)	UA (%)	PA (pixels)	UA (pixels)
Sf	1.44	1.89	210/14550	276/14616	98.11	98.56	14340/14616	14340/14550
Mf	43.37	33.80	108/249	72/213	66.20	56.63	141/213	141/249
Com	51.11	24.26	506/990	155/639	75.74	48.89	484/639	484/990
Off	72.22	91.70	52/72	221/241	8.30	27.78	20/241	20/72
Ind	57.38	80.00	35/61	104/130	20.00	42.62	26/130	26/61
Civ	33.33	94.87	2/6	74/78	5.13	66.67	4/78	4/6
Open	29.13	46.11	37/127	77/167	53.89	70.87	90/167	90/127
Tran	50.00	69.84	19/38	44/63	30.16	50.00	19/63	19/38
Und	23.47	12.37	100/426	46/372	87.63	76.53	326/372	326/426

Note: CE=commission error, OE=omission error, PA=producer's accuracy, US=user's accuracy.

6. Discussions

One limitation of per-field land use classification is the need for pre-determined field boundaries. This study relies on existing vector data of tax parcel boundaries. The advantage is that tax parcels always contain homogeneous land use and are usually up-to-date. The disadvantage is that tax parcels do not cover the entire land surface, specifically, the street surface. Also, some rural

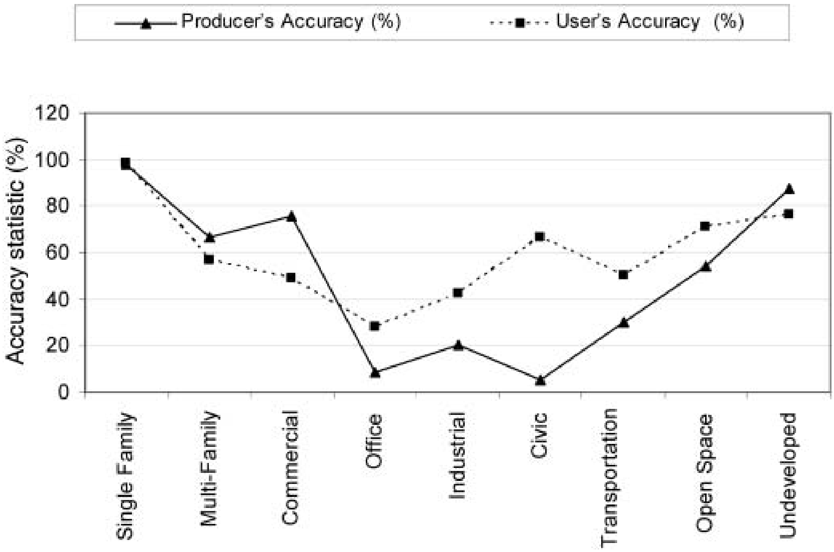

Figure 14. Graph of producer's accuracy and user's accuracy for different land use.

areas in the US and other countries may not have complete tax parcel database. Therefore, the major application areas of parcel-based per-field urban land use classification are the US cities, where tax parcel databases are usually available and up-to-date.

The current study relies on ancillary data to obtain parcel attributes, which can be used for land use classification. Within the four ancillary datasets used, building footprints vector polygon data are the most important but most likely to be unavailable for many urban areas. In the past, no effective ways for automatic extraction of residential buildings exist. Researchers generally had to visually identify and manually digitize dwelling units from high-spatial-resolution aerial photographs. With the advance of very high spatial resolution satellite images, such as IKONOS and QuickBird, and the improvement of feature extraction techniques, automatic extraction of dwelling units from satellite images has become possible (Haverkamp 2004). Another prospect for automatic building extraction is the advancement of three-dimensional (3D) object extraction

Table 6. Signature separabilities based on individual and combinations of parcel attributes.

No.	Parcel attribute	Signature separability			
1	Parcel size	0.24	0.24	0.28	1.88
2	Parcel shape compactness	0.04	0.04		
3	Number of buildings	0.29	0.29	1.72	
4	Maximum building's area	0.61	1.13		
5	Standard deviation of the building's area	0.67			
6	Total building-area percentage	0.11			
7	Maximum building's height	0.23	0.52		
8	Standard deviation of the building's height	0.16			
9	Maximum building shape compactness	0.12	0.36		
10	Standard deviation of building shape compactness	0.11			
11	The highest category of street within 50 m	0.04	0.04	0.04	
12	Impervious cover percentage	0.09	0.09	0.09	

techniques from LIDAR data (Rottensteiner 2003; Rottensteiner *et al.* 2004). With these new remote sensing data and building extraction techniques, population estimation by dwelling unit counts is likely to become a viable approach.

Another potential problem of using ancillary data is the mismatch of timeframes between datasets. For example, the four ancillary datasets and the ground truth land use data used in this study are all one or two years older than the parcel boundaries data and, therefore, might be mismatching to them. Furthermore, even for datasets of the same year, they still may not match well with each other, as some may not be updated in a timely manner.

The present study relies on GIS vector data while making use of existing remote sensing algorithms for land use classification. Although GIS systems are generally more capable of processing vector data, current GIS provide limited capabilities for land use classification. In contrast, land use classification has a long history in remote sensing and many classification algorithms have been developed. Furthermore, the use of GIS data in land use classification has not been explored extensively. There is a need for more integrated systems that allow making use of both remote sensing and GIS data for land use classification.

A relatively new approach for urban land use classification is to utilize contexture information regarding the spatial arrangement between land cover classes or landscape objects for land use classification (e.g. Barnsley *et al.* 2003, Herold *et al.* 2003, Bian and Xie 2004). Future researches may incorporate contextual parameters into per-field urban land use classification, such as a contextual statistics related to the spatial arrangement of buildings, and a contextual statistics related to the similarity between studied parcels and neighbouring parcels. In addition, further researches may compare the GIS parameters used in this study with some common image spectral and texture parameters for urban land use classification, or incorporate the image parameters as classification criteria.

7. Conclusions

The current study presents a per-field approach for classifying urban land use. We classified nine types of urban land use based on 12 tax parcel attributes. The nine types of urban land use include single family, multi-family, commercial, office, industrial, civic, open space, transportation and undeveloped. The 12 tax parcel attributes include parcel size, parcel shape compactness, the number of buildings, the maximum building's area, the standard deviation of the building's area, the total building-area percentage, the maximum building's height, the standard deviation of the building's height, the maximum building shape compactness, the standard deviation of building shape compactness, the highest category of streets within 50 m, and the impervious cover percentage. The signature separability analysis based on different combinations of parcel attributes indicates that the eight building-relevant attributes are the major discriminant factors between land use types, and the two parcel-relevant attributes are secondary. Within those eight building-relevant attributes, the three related to building areas are most important, next the two related to building heights, the two related to building shape compactness and, lastly, the one related to the number of buildings.

The classification results have an overall accuracy of 93.53% and a Kappa coefficient of 0.7023. The results show that the six building types of land use can be separated from other three non-building types of land use. Within the six building types of land use, single-family and multi-family land use can be separated from other non-residential land use types, as well as from each other. Within the three non-building types of land use, undeveloped can be separated from the other two. Overall, this study shows that a per-field approach based on tax parcel boundaries and GIS attributes of parcels can be used to classify detailed urban land use. This classification methodology can also be applied to other areas, provided that the parcel boundaries data are available, although different geographic regions may have different parcel attributes that are important for land use classification.

Although the current study shows that GIS data and GIS attributes are important for urban land use classification, it is worth noting that many of the GIS data and attributes are originated from remote sensing data. For example, the building heights are derived from remote sensing LIDAR data; building shapes and building areas are manually delineated from high-resolution aerial photographs. Therefore, remote sensing data are still the fundamental source of data for urban land use classification.

Urban land use classification in remote sensing has its fundamental limitation in that some land use types have similar physical characteristics or they are heterogeneous in nature and cannot be separated from each other. Compared with manual digitizing, digital classification of urban land use still has great uncertainties and is, therefore, not adopted by urban planners. Nevertheless, with the advancement of new remote sensing data and techniques in urban analysis, it is expected that the advantages of digital image processing, e.g., efficiency and objectivity, will one day justify it for real-world use in detailed urban land use classification.

References

APLIN, P., ATKINSON, P. and CURRAN, P., 1999, Fine spatial resolution simulated satellite sensor imagery for land cover mapping in the United Kingdom. *Remote Sensing of Environment*, **68**, pp. 206–216.

APLIN, P. and ATKINSON, P., 2001, Sub-pixel land cover mapping for per-field classification. *International Journal of Remote Sensing*, **22**, pp. 2853–2858.

BARNSLEY, M.J., ALAN, M.S. and BARR, S.L., 2003, Determining urban land use through an analysis of the spatial composition of buildings identified in LIDAR and multispectral image data. In *Remotely Sensed Cities*, V. Mesev (Ed.), pp. 83–108 (London: Taylor & Francis).

BERBEROGLU, S., LLOYD, C., ATKINSON, P. and CURRAN, P., 2000, The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. *Computers and Geosciences*, **26**, pp. 385–396.

BIAN, L. and XIE, Z., 2004, A spatial dependence approach to retrieving industrial complexes from digital images. *Professional Geographer*, **56**, pp. 381–393.

CITY OF AUSTIN, 2005a, City of Austin GIS Data Sets. Available online at: ftp://issweb.ci.austin.tx.us/pub/coa_gis.html (accessed 21 July 2005).

CITY OF AUSTIN, 2005b, 2003 Land Use Inventory Overview and Methodology. Available online at: http://www.ci.austin.tx.us/landuse/survey2003.htm (accessed 21 July 2005).

CITY OF AUSTIN, 2005c, Year 2003 Land Use Metadata. Available online at: ftp://issweb.ci.austin.tx.us/pub/GIS-Data/Regional/landuse/landuse2003.htm (accessed 21 July 2005).

CITY OF AUSTIN, 2005d, Site Development Regulations, Austin City Code. Available online at: http://www.amlegal.com/austin%5Ftx/ (accessed 21 July 2005).

CITY OF AUSTIN, 2005e, City of Austin Demographics. Available online at: http://www.ci.austin.tx.us/census (accessed 21 July 2005).

DEAN, A. and SMITH, G., 2003, An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. *International Journal of Remote Sensing*, **24**, pp. 2905–2920.

De Wit, A. and Clevers, J., 2004, Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing*, **25**, pp. 4091–4112.

Erol, H. and Akdeniz, F., 1996, A multispectral classification algorithm for classifying parcels in an agricultural region. *International Journal of Remote Sensing*, **17**, pp. 3357–3371.

Erol, H. and Akdeniz, F., 2005, A per-field classification method based on mixture distribution models and an application to Landsat Thematic Mapper data. *International Journal of Remote Sensing*, **26**, pp. 1229–1244.

Fuller, R., Smith, G., Sanderson, J., Hill, R. and Thomson, A., 2002, The UK Land Cover Map 2000: Construction of a parcel-based vector map from satellite images. *The Cartographic Journal*, **39**, pp. 15–25.

Geneletti, D. and Gorte, B., 2003, A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *International Journal of Remote Sensing*, **24**, pp. 1273–1286.

Haverkamp, D., 2004, Automatic building extraction from IKONOS imagery. *Proceedings of ASPRS 2004 Conference*, Denver, Colorado, 23–28 May, 2004. Available online at: http://www.spaceimaging.com/techpapers/default.htm (accessed 21 July 2005).

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

Hill, R., Smith, G., Fuller, R. and Veitch, N., 2002, Landscape modelling using integrated airborne multi-spectral and laser scanning data. *International Journal of Remote Sensing*, **23**, pp. 2327–2334.

Janssen, L. and Molenaar, M., 1995, Terrain objects, their dynamics and their monitoring by the integration of GIS and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, pp. 749–758.

Johnsson, K., 1994, Segment-based land-use classification from spot satellite data. *Photogrammetric Engineering and Remote Sensing*, **60**, pp. 47–53.

Lloyd, C., Berberoglu, S., Curran, P. and Atkinson, P., 2004, A comparison of texture measures for the per-field classification of Mediterranean land cover. *International Journal of Remote Sensing*, **25**, pp. 3943–3965.

Lobo, A., Chic, O. and Casterad, A., 1996, Classification of Mediterranean crops with multisensor data: per-pixel versus per-object statistics and image segmentation. *International Journal of Remote Sensing*, **17**, pp. 2385–2400.

Loh, W., 2005, QUEST Classification Tree. Available online at: http://www.stat.wisc.edu/~loh/quest.html (accessed 21 July 2005).

Loh, W. and Shih, Y., 1997, Split selection methods for classification trees. *Statistica Sinica*, **7**, pp. 815–840.

McCauley, S. and Goetz, S., 2004, Mapping residential density patterns using multi-temporal Landsat data and a decision-tree classifier. *International Journal of Remote Sensing*, **25**, pp. 1077–1094.

Pedley, M. and Curran, P., 1991, Per-field classification: an example using SPOT HRV imagery. *International Journal of Remote Sensing*, **12**, pp. 2181–2192.

Rottensteiner, F., 2003, Automatic generation of high-quality building models from Lidar data. *IEEE Computer Graphics and Applications*, **23**, pp. 42–50.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K. and Lovell, B., 2004, Building detection by Dempster-Shafer fusion of lidar data and multispectral aerial imagery. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, **2**, 23–26 August 2004, Cambridge, UK, pp. 339–342 (Washington: IEEE Computer Society).

Schalkoff, R.J., 1989, *Digital Image Processing and Computer Vision* (New York: Wiley).

Smith, G. and Fuller, R., 2001, An integrated approach to land cover classification: an example in the Island of Jersey. *International Journal of Remote Sensing*, **22**, pp. 3123–3142.

US Census Bureau, 2000, TIGER/Line Files Technical Documentation. Available online at: http://www.census.gov/geo/www/tiger/rd_2ktiger/tgrrd2k.pdf (accessed 21 July 2005).

Weiler, R. and Stow, D., 1991, Spatial analysis of land cover patterns and corresponding remotely-sensed image brightness. *International Journal of Remote Sensing*, **12**, pp. 2237–2257.

Zhan, Q., Molenaar, M. and Gorte, B., 2000, Urban land use classes with fuzzy membership and classification based on integration of remote sensing and GIS. *International Archives of Photogrammetry and Remote Sensing*, **33**, pp. 1751–1759.