

Using Geometrical, Textural, and Contextual Information of Land Parcels for Classification of Detailed Urban Land Use

Shuo-Sheng Wu,* Xiaomin Qiu,* E. Lynn Usery,[†] and Le Wang[‡]

*Department of Geography, Geology, and Planning, Missouri State University

[†]Center of Excellence for Geospatial Information Science, U.S. Geological Survey

[‡]Department of Geography, State University of New York at Buffalo

Detailed urban land use data are important to government officials, researchers, and businesspeople for a variety of purposes. This article presents an approach to classifying detailed urban land use based on geometrical, textural, and contextual information of land parcels. An area of 6 by 14 km in Austin, Texas, with land parcel boundaries delineated by the Travis Central Appraisal District of Travis County, Texas, is tested for the approach. We derive fifty parcel attributes from relevant geographic information system (GIS) and remote sensing data and use them to discriminate among nine urban land uses: single family, multifamily, commercial, office, industrial, civic, open space, transportation, and undeveloped. Half of the 33,025 parcels in the study area are used as training data for land use classification and the other half are used as testing data for accuracy assessment. The best result with a decision tree classification algorithm has an overall accuracy of 96 percent and a kappa coefficient of 0.78, and two naive, baseline models based on the majority rule and the spatial autocorrelation rule have overall accuracy of 89 percent and 79 percent, respectively. The algorithm is relatively good at classifying single-family, multifamily, commercial, open space, and undeveloped land uses and relatively poor at classifying office, industrial, civic, and transportation land uses. The most important attributes for land use classification are the geometrical attributes, particularly those related to building areas. Next are the contextual attributes, particularly those relevant to the spatial relationship between buildings, then the textural attributes, particularly the semivariance texture statistic from 0.61-m resolution images. *Key Words:* contextual classification, field-based, land use classification, per field, textural classification.

详细的城市土地利用数据对于政府官员，研究人员和商界人士的多种用途是很重要的。本文介绍了一种基于土地分块小区的几何特征，质地特征和背景信息的详细城市土地利用的分类方法。本研究的试验区选取了德克萨斯州奥斯汀市一块6乘14公里的区域，其区划边界是Travis县的Travis中心评估区。本研究使用了相关的地理信息系统（GIS）和遥感数据，并利用它们划分了九种不同的城市土地用途：单个家庭，多个家庭，商业，办公，工业，民用，开放空间，交通和未开发区。试验区包括了总共33025个土地分块小区，其中的一半被用作分类训练区，另一半被用作分类结果的准确性评估。研究表明，决策树分类算法获得了最高的分类精度，总体精度达到了96%，卡帕相关系数为0.78。另外两种基于主体原则和空间自相关原则的原始基线算法模型分别获得了89%和79%的总体精度。决策树分类算法在分类单个家庭，多个家庭，商业，开放空间，和未开发区的土地用途类别上取得了相对较好的结果，但是在办公，工业，民用，交通的土地用途类别上效果不佳。在土地利用分类的算法中，最重要的属性是土地的几何特征，特别是在那些有建筑物的地区。其次是土地的背景特征，特别是那些有关建筑物之间的空间关系。最后是土地的质地特征，特指那些在0.61米分辨率图像上获取的土地质地的统计半方差。*关键词：*背景分类，基于领域，土地利用分类，每个领域，质地分类。

Para una variedad de propósitos, los datos detallados sobre uso del suelo urbano son importantes para agentes gubernamentales, investigadores y hombres de negocios, Este artículo presenta un enfoque para clasificar en detalle los usos del suelo urbano, a partir de información geométrica, textural y contextual de las parcelas. Este

enfoque se puso a prueba en un área de 6 X 14 km, en Austin, Texas, con los linderos de las parcelas delineados por el Distrito Central de Avalúos Travis, del condado Travis. Con datos relevantes generados por un sistema de información geográfica (SIG) y por teledetección, derivamos cincuenta atributos de las parcelas que se utilizaron para discriminar entre nueve usos del suelo urbano: familiar, multifamiliar, comercial, oficinas, industria, cívico, espacio abierto, transporte y no desarrollado. La mitad de las 33.025 parcelas del área de estudio fungió como espacio de datos para entrenamiento en la clasificación del uso del suelo, y la otra mitad como campo de datos para prueba para efectos de la exactitud en la evaluación. El mejor resultado logrado con un algoritmo clasificatorio por árbol de decisiones tiene una exactitud general del 96 por ciento y un coeficiente kappa de 0.78, y dos modelos de línea de base basados el uno en la regla de la mayoría y el otro en la regla de la autocorrelación espacial, tienen exactitudes totales de 89 y 79 por ciento, respectivamente. El algoritmo es relativamente bueno en lo que concierne a la clasificación de los usos del suelo unifamiliar, multifamiliar, comercial, espacio abierto y usos no desarrollados, y relativamente deficiente en cuanto a la clasificación de usos del suelo para oficinas, industria, cívico y transporte. Los atributos más importantes para la clasificación del uso del suelo son los atributos geométricos, en particular aquellos relacionados con áreas de edificios. Luego vienen los atributos contextuales, particularmente los relevantes a las relaciones espaciales entre los edificios; después los atributos texturales, en particular la estadística de semivarianza de textura de imágenes de 0.61-m de resolución. *Palabras clave: clasificación contextual; basada en campo; clasificación de usos del suelo; por campo; clasificación textural.*

Land use data show where and how land is used. In urban areas, information on detailed land use, such as single family, multifamily, industrial, and commercial, is useful to urban planners and researchers for a variety of purposes, including population estimation and forecast, corridor and transit planning, neighborhood planning and zoning, watershed and floodplain modeling, hazard and pollution analysis, and budgeting and service planning for water and wastewater, energy, and fire (Donnay and Unwin 2001; City of Austin 2007e). Currently, the update of urban land use data in municipal governments relies mainly on aerial photo interpretation and field survey, sometimes referenced with appraisal records, development records, census data, or other relevant data (Donnay and Unwin 2001; City of Austin 2007f). For large municipalities, the process of updating land use data is laborious and time consuming. Although zoning ordinances as city regulations designate permitted uses of land within city limits, a zoning system, designed for long-term planning purposes, is schematically different from the present land use situation. For example, in the city of Austin, Texas, one land use class may fall into many zoning categories, and one zoning category can contain several permissible land use classes. Moreover, due to grandfathering, there might be a mismatch between the zoning-allowed land use and the actual land use on the ground.¹ Land use classification is a primary research topic in remote sensing. The classification of detailed urban land use classes from remotely sensed images, particularly, has been a challenge for researchers and practitioners. The main reason is that an urban land use classification scheme is developed based on socioeconomic functionality instead of biophysical characteristics that

are closely related to the spectral reflectance detected by remote sensing images (Zhan, Molenaar, and Gorte 2000). An urban land use category is commonly composed of various land cover types, each with different spectral properties. Therefore, many urban land use classes are heterogeneous in spectral and textural characteristics and it is difficult to separate them by classification algorithms based on these characteristics alone. The utilization of high spatial and spectral resolution data and advanced data processing techniques in recent years has improved, to varied extents, urban land cover mapping (Hodgson et al. 2003; Benediktsson, Palma-son, and Sveinsson 2005; Unsalan and Boyer 2005) and urban land use classification (Zhang and Wang 2003; Wu, Xu, and Wang 2006). Nevertheless, the improvement is still considerably limited by the heterogeneous nature of urban land use and land cover classes as well as the conventional image classification approach.

This study presents a per field approach for urban land use classification that is based on geometrical, textural, and contextual attributes of predetermined field boundaries. The approach is designed to overcome the heterogeneous nature of urban land use classes in land use classification. Using a case study in Austin, Texas, we test fifty parcel attributes to classify nine types of urban land use. This study is a substantial improvement of a previous per field classification study of the same area by Wu, Silván-Cárdenas, and Wang (2007), in which twelve parcel attributes, including ten geometrical, one textural, and one contextual, were tested. By incorporating additional relevant textural and contextual attributes (for a total of fifty parcel attributes), this study improves the overall accuracy from 93.5 percent to 95.6 percent and the kappa coefficient from

70.2 percent to 78.1 percent. To manage the numerous parcel attributes as discriminating criteria in an image classification procedure and to efficiently test different classification algorithms, this study adopts an alternative approach by generating an artificial fishnet image from the parcel data with a relatively small size to use for classification. Additionally, compared to the study by Wu, Silván-Cárdenas, and Wang (2007), this study explicitly compares the land use classification capability between relevant geometrical, textural, and contextual attributes of land parcels. This study also investigates whether housing and demographic statistics from the U.S. Census data, which are widely available to the public, can help to improve classification accuracy.

In the following sections, we first review general categories of image classification methods in remote sensing. Then, in detail, we review per field classification approaches, which are relevant to our land use classification methodology. We then introduce the semivariance statistic to be used for calculating parcel attributes. After that, we present the baseline model of classification and accuracy statistics to be used for assessing the classification results. Finally, we describe our analysis procedures and the outcome.

Image Classification in Remote Sensing

The earth's surface contains a variety of spectrally different materials, such as water, soil, rocks, grass, trees, concrete, asphalt, plastic, and metal. Remote sensing images record the electromagnetic radiation of the earth's surface in digital values, which provides spectral and spatial information for land use and land cover classification. Spectral information is conveyed by the digital value of image pixels and the unique light spectrum, and spatial information comes from the variation of pixel values in an area. Spatial information involves a variety of image characteristics, such as color, tones, pattern, shape, size, and texture; among those, texture is the most commonly used measure for image classification (Schalkoff 1989). Qualitatively, texture provides people with an impression of coarseness or smoothness to facilitate photo interpretation. Quantitatively, texture can be defined in various mathematical functions to represent the local variability of pixel values for assisting image classification.

In remote sensing, conventional pixel-based classification assigns each pixel to one of the candidate classes based on its pixel value from the spectral reflectance of the earth's surface. However, using spectral information

alone may not be sufficient for classifying spectrally heterogeneous but spatially similar land uses; therefore, classification based on textural information is employed to complement spectral classification. In textural classification, new textural bands are created; each pixel is assigned a new digital value to represent the local texture (e.g., within a pixel window), based on specific mathematical transformations. Researchers have developed a wide range of textural measures for image classification, including first-order statistics (such as means, variance, and standard deviation), second-order statistics (such as those based on a gray-level co-occurrence matrix), texture spectrum, fractals, and semivariance (Jensen 2005); yet no single approach that is both efficient and effective has been widely adopted (Jensen 2005).

To improve conventional spectral classification, researchers also have used higher order contextual information to distinguish between different land use and land cover classes (de Jong and van der Meer 2004). For example, contextual information regarding the spatial arrangement between land covers or land objects can be incorporated as discriminating criteria between land use classes. Studies by Barnsley, Alan, and Barr (2003), Zhang and Wang (2003), and Wu, Xu, and Wang (2006) apply various forms of convolution kernels and quantification techniques to measure the spatial arrangement between urban land cover (e.g., trees, grass, buildings, and roads) for classifying urban land use classes. Although this approach has achieved promising results in classifying urban land use classes that contain land covers arranged in regular, detectable spatial patterns, the use of predefined convolution kernels with specified quantification techniques places a limitation on its applicability and reliability. Moreover, this classification approach is likely to have difficulty with urban areas that contain small land use patches or land use with spatial patterns that are not very distinct.

Conventional land use or land cover classification in remote sensing assigns classes on a per pixel basis, which could be based on spectral, textural, or contextual properties of images. In contrast to this per pixel classification, per field classification determines land use and land cover by predetermined, meaningful field boundaries, with the premise that each field belongs to a single, homogeneous class (Pedley and Curran 1991; Aplin, Atkinson, and Curran 1999; Geneletti and Gorte 2003). Per field classification may overcome some weaknesses of per pixel classification. For example, per pixel classification commonly produces a pixelly result, which may look noisy and require a postprocessing filter to improve the outlook as well as

the classification accuracy. Furthermore, classification by pixels often has difficulties classifying spectrally and texturally heterogeneous classes and delineating class boundaries. Using high-spatial-resolution data (e.g., IKONOS images) or high-spectral-resolution data (e.g., Hyperion images) in per pixel classification may improve the classification result but is still limited by the weaknesses of per pixel classification already discussed.

In addition to overcoming the problems with per pixel classification, per field classification has the advantage of allowing the incorporation of some potentially useful field attributes, such as the size, shape, and perimeter of the field, as the classification criteria, because fields have geographical meaning on the landscape, whereas pixels are artifacts of the data format. Past studies have reported improved classification results by employing a per field approach or by mixing a per field approach with a per pixel approach for land use and land cover classification, such as the studies by Dean and Smith (2003), Erol and Akdeniz (2005), and Platt and Rapoza (2008). The next section reviews past studies using per field classification methods.

Per Field Classification

To classify land use and land cover based on the spatial unit of fields instead of pixels, researchers need to first determine field boundaries that partition the interested area into homogeneous regions. Many per field studies utilize existing polygon features in digital data or hard-copy maps as field boundaries (Zhan, Molenaar, and Gorte 2000; Aplin and Atkinson 2001; Wu, Silván-Cárdenas, and Wang 2007). Some per field studies use digital image segmentation techniques, such as edge detection and region growing, to help determine or directly determine field boundaries. Edge detection techniques may derive linear features (e.g., roads, railways, and rivers) as potential field boundaries, and region growing techniques may generate areas of homogeneous spectral, spatial, or contextual statistics as potential fields. Examples include Hill et al. (2002), Geneletti and Gorte (2003), and Platt and Rapoza (2008). Other than relying on image segmentation techniques, some studies delineate field boundaries by visual interpretation and manually digitizing identifiable fields from digital images or maps (Berberoglu et al. 2000; Herold, Liu, and Clarke 2003; Lloyd et al. 2004). Of these three approaches for determining field boundaries, utilizing existing ancillary vector data is preferred if the data are available and have a satisfactory degree of accuracy and precision. Many

existing vector data come from field surveys or photo interpretation and, therefore, provide meaningful field boundaries to meet users' needs. In contrast to utilizing existing ancillary vector data to derive field boundaries, manually digitizing fields through visual interpretation is labor intensive and time consuming, and image segmentation techniques often cannot produce desirable fields that pertain to meaningful land cover or land use classes (De Wit and Clevers 2004).

After field boundaries are determined, researchers may utilize field boundaries in two ways. One way is to use in a preclassification stage for deriving field attributes (e.g., image spectral and textural statistics) that can be used as discriminant criteria for classifying fields (Erol and Akdeniz 1996; Lobo, Chic, and Cast-erad 1996; Wu, Silván-Cárdenas, and Wang 2007). The other use is in a postclassification stage. A typical approach in this category is to assign the major class in a field to all pixels within the field after an initial per pixel classification (Janssen and Molenaar 1995; Aplin, Atkinson, and Curran 1999; Aplin and Atkinson 2001). Moreover, some studies first utilize field boundaries in a preclassification stage to derive field attributes and classify fields and, then, in a postclassification stage, obtain additional field attributes to modify field classes for improving the final classification accuracy (Smith and Fuller 2001; Fuller et al. 2002; Hill et al. 2002). Berberoglu et al. (2000) particularly compared the two approaches of utilizing field boundary data with the same class-discriminating criteria and found that utilizing field boundaries in a preclassification stage provides better classification results.

In addition to image statistics within fields, the geometrical and contextual attributes of fields can also be used for land use and land cover classification. For example, Zhan, Molenaar, and Gorte (2000) classified fields based on proportion of built-up area, green space, and water surface within the field. De Wit and Clevers (2004) used field size and shape as criteria to reassign field class in a postprocessing stage. Platt and Rapoza (2008) incorporated the border percentage of fields to residential class as a parameter to classify nonresidential and other land use and land cover classes. Liu and Herold (2008) calculated spatial metrics of land covers and semivariograms of Normalized Difference Vegetation Index (NDVI) values for predefined land use zones to classify seven detailed urban land use classes.

This study adopts a per field approach for classification of nine detailed urban land uses based on digital land parcel boundaries that are originally from the tax parcel database of the Travis Central Appraisal

District (TCAD) in Texas. Our per field classification approach is different from common per field studies in three ways. First, most per field studies classified spectrally homogeneous land cover types, a few classified residential and nonresidential urban land use, but very few per field studies classified detailed urban land use classes. Importantly, the goal of this study is to classify nine detailed urban land use classes.

Second, compared to conventional per field classification approaches, this study focuses on the classification of urban land use from land parcels instead of remote sensing images. In other words, the parcel data are the primary data and the image data are the ancillary data used to derive some parcel attributes. This strategy is conceptually different from conventional per field classification studies and has implications for classification methodology and the results. Specifically, in addition to using image parameters for classification, our per field classification incorporates a variety of relevant field attributes derived from ancillary geographic information system (GIS) data for classification. The discriminating power between land use classes is, therefore, increased and the classification accuracy is also improved.

Third, this study uses a variety of textural and contextual attributes of fields derived from ancillary GIS and image data to classify detailed urban land use classes. Many of the field attributes have not been tested in past per field studies, such as the neighborhood building relational statistic, the similarity index to neighboring parcels, and a number of census housing and demographic statistics. These field attributes are considered based on theoretical reasons, and some of them have proven to be effective criteria for discriminating among detailed urban land use classes in our study.

Semivariance Textural Statistic

Past studies have shown the effectiveness of textural statistics for characterizing or classifying spectrally heterogeneous land use classes (Gong, Marceau, and Howarth 1992; Karathanassi, Iossifidis, and Rokos 2000; Zhang et al. 2003). Texture statistics can indicate the degree of spectral variation or the level of landscape heterogeneity of an area that is related to the local land use. For example, a land use class with small houses and minimal distance between houses, which indicates higher degree of landscape heterogeneity, usually has higher values of image texture statistics, compared with those

land use classes having larger houses and greater distance between houses. In this study, the texture statistic of semivariance is incorporated into per field classification. Semivariance is calculated as half the average of the squared difference between paired pixel values separated by a specific number of pixels (referred to as the lag; Burrough and McDonnell 1998). The mathematical equation of semivariance, $g(h)$, separated by lag, h , can be expressed as:

$$g(h) = \frac{1}{2N_h} \sum_{i=1}^N (z_i - z_{i+h})^2 \quad (1)$$

where N_h is the number of paired pixels separated by lag h , and z_i and z_{i+h} are the values of the pair of pixels separated by lag h . A graph of semivariance against the lag is called the *variogram*, which is usually used to indicate the extent of spatial autocorrelation across space. Semivariance is used in this study as a texture statistic to measure the degree of landscape heterogeneity. A past study by Wu, Xu, and Wang (2006) has indicated the effectiveness of applying semivariance to quantify the spatial relationship between buildings for classifying detailed urban land use classes. Most important, the semivariance is used because it can provide multiple lag scales for discriminating between land use classes.

A Baseline Model of Classification

A baseline model of classification is a null or naive model used as a standard to evaluate other classification algorithms (Pontius et al. 2007). Because there are no agreed-on criteria of what is considered satisfactory, a reasonable minimum criterion would be that the agreement between the validation data and the prediction from a scientist's model should be better than the agreement between the validation data and the prediction from a null model (Pontius, Huffaker, and Denman 2004). The prediction from a null model is the naive prediction that one would make if one were not to create any analytical model. A baseline model is directly linked to the research question and the study site and is, therefore, appropriate to be used as a benchmark for comparison. This study presents two naive models as baseline models. The first is based on the training data and a naive assumption of spatial dependence, which assumes that every parcel is in the same land use category as the nearest training parcel. The second naive model is based on the training data and a naive assumption

of majority rule. The training data show that approximately 89 percent of the total parcels are single-family land use parcels. If a naive person is given this information and asked to predict the other testing parcels, a naive strategy would be to predict all parcels as single-family land use parcels, which would yield an overall accuracy of approximately 89 percent.

Measures for Accuracy Assessment

Many methods of assessing the accuracy of image classification have been proposed in the remote sensing literature (Kalkhan, Reich, and Czaplewski 1995; Koukoulas and Blackburn 2001). Many widely used accuracy measures can be derived from a confusion matrix (Lark 1995; Stehman 1997; Foody 2002), which is a cross-tabulation of the classified class (represented in each row) against the ground class (represented in each column) for the number of cases. For example, the overall accuracy (percentage correct) is calculated by summing the numbers of correctly classified cases, which are the diagonal elements in the confusion matrix, and dividing by the total number of cases in the study area. For individual classes, the accuracy is derived by dividing the number of cases correctly classified to the class by the total number of cases of that class, which can be based on either the ground class or the classified class. If the class-level accuracy is based on the ground class, the matrix's column marginal is used as the divider and the measure is termed the *producer's accuracy*. If the class-level accuracy is based on the classified class, the matrix's row marginal is used as the divider and the measure is termed the *user's accuracy*. In other words, the producer's accuracy calculates the percentage of the ground class that is correctly classified, whereas the user's accuracy calculates the percentage of the classified class that is correct. They are summary accuracy statistics for individual classes viewed from different perspectives, which can provide a more holistic view of classification accuracy.

The measures of the percentages of total cases correctly classified have often been criticized because some cases may have been allocated to the correct class purely by chance. Consequently, the kappa coefficient has been proposed to accommodate the effects of chance agreement (Congalton 1991). Specifically, the kappa coefficient measures the proportion of agreement that occurs beyond what would be expected by random spatial allocation of the classes, given the errors in the estimates of the proportions of the classes (Pontius

2000). Many modelers, remote sensing specialists, and statisticians endorse the kappa coefficient as a standard measure of classification accuracy because of its many attractive features (Hudson and Ramm 1987; Stehman 1996; Smits, Dellepiane, and Schowengerdt 1999). In addition to making compensation for chance agreement, the kappa coefficient allows a variance to be calculated for testing the statistical significance of differences in comparing different classification results (Foody 2002). Some researchers further suggest normalizing the confusion matrix to aid this comparison, which makes each row and column of the matrix sum to one (Congalton 1991; Smits, Dellepiane, and Schowengerdt 1999).

Nevertheless, many researchers are concerned about the flaws of the kappa coefficient (Foody 1992; Ma and Redmond 1995; Stehman 1997). One major problem is that the chance agreement may be effectively overestimated in the kappa coefficient, resulting in an underestimation of classification accuracy. Specifically, due to purposeful, convenient, or haphazard procedures of selecting reference locations, as a nonprobability-based measure, the kappa coefficient is an inappropriate basis for accuracy assessment (Stehman and Czaplewski 1998; Foody 2002). From another point of view, the goal of image classification is to obtain similar marginal distributions for the confusion matrix. Therefore, the standard method to compute the expected proportion-correct classification by chance is not appropriate for accuracy assessment.

Pontius (2000) argued that the kappa coefficient fails to reward the classification for accurately estimating the quantity of each class as shown in the marginal totals of the confusion matrix. Instead, he presented an approach that specifies quantity error and location error of classification. The quantity error is due to the less-than-perfect ability of the classification model to specify quantity of classes, and the location error is due to the less-than-perfect ability of the classification model to specify location of classes (Pontius et al. 2007). The quantity error can be derived by subtracting the expected percentage correct of a model that allocates the specified quantity of classes (as in the classified map) accurately in space from the expected percentage correct of a model that specifies both the location and the quantity of classes accurately (which is 1). The location error can be derived by subtracting the observed percentage correct (overall accuracy) of the model from the expected percentage correct of a model that accurately assigns locations for the specified quantity of classes. Pontius (2000) argued that a spatially explicit

classification model should be judged on its ability to accurately specify both quantity and location of classes. Knowing the quantity error helps us understand the constraint of the model regarding the percentage correct that could be potentially improved. Knowing the location error helps us interpret the spatial aspect of the error that could potentially be removed given the specified quantity of classes (Pontius et al. 2007). Separating the overall classification error into quantity error and location error provides insight for scientists to decide whether to dedicate energy to improve a model's ability to specify quantity or location (Pontius, Huffaker, and Denman 2004). This study compares different classification models by the overall accuracy, location error, and quantity error, and the popular yet controversial kappa coefficient accuracy statistic is not adopted. Nevertheless, for the best classification model that has the highest overall accuracy, in addition to presenting a confusion matrix and the producer's and user's accuracies for individual classes, we also reported the kappa coefficient to facilitate the comparison with other land use classification studies that have reported a kappa coefficient.

Methods

We applied a per field approach for urban land use classification based on the TCAD land parcel boundaries that have been digitized by the City of Austin Neighborhood Planning and Zoning Department (NPZD). The main advantage of using the digital TCAD land parcel data is that individual parcels always contain the same type of land use by definition. Additionally, the parcel boundaries have a high degree of spatial precision and are updated annually (City of Austin 2007g). Because the land parcels do not cover city street surface areas, this study focuses on classifying land use from the parcel areas, and street surfaces are not considered.

After determining field boundaries for our per field classification, we derived fifty parcel attributes (e.g., parcel shape, building height, proximity to major roads, similarity to neighboring parcels), from relevant GIS and remote sensing data, which have been derived independently of the land use variable we are trying to acquire. Parcel attributes are then used as discriminating criteria among nine urban land use classes to classify the parcels. The hypothesis is that these parcel attributes are related to land use classes. The goal is, therefore, to investigate how well these parcel attributes, collectively

as well as individually, can be used to distinguish among different land use classes. Standard image classification procedures in remote sensing (i.e., training, classification, and accuracy assessment) are adopted to perform the analysis.

Study Area

We selected an area of approximately 6 by 14 km in the north central part of Austin, Texas, to test our per field land use classification (Figure 1). The city's three main thoroughfares, I-35, MoPac, and Highway 183, all run through this area. With commercial land use areas close to major roads, a variety of residential and nonresidential land uses are between the roads, providing a suitable environment for our analysis. Austin is the capital of the state of Texas and the county seat of Travis County. Situated in central Texas, it is the fourth-largest city in Texas and the sixteenth-largest in the United States. Occupying a land area of 704 km², Austin had a 2000 Census population of 656,562. As of the 2007 U.S. Census estimate, Austin had a population of 743,074. The city is the core cultural and economic center of the Austin–Round Rock metropolitan area with a population of 1.4 million. The eastern part of the city is relatively flat, whereas the western part and western suburbs consist of scenic rolling hills on the edge of the Texas Hill Country (Texas State Historical Association 2006; City of Austin 2007c).

Data Source

We obtained the TCAD land parcel boundaries data, four ancillary data sets, and the ground land use data from the NPZD, either directly downloaded from its File Transfer Protocol server (City of Austin 2007d) or acquired through personal contact. The four ancillary data sets are for deriving parcel attributes and include building data, elevation data, street data, and image data.

The TCAD land parcel boundaries data are in vector polygon format and are current to 2006. Nevertheless, to match the time frame of our ground land use data, which are current to 2003, we obtained the parcel boundaries data from 2003 for our analysis. The NPZD has used parcel information from the TCAD tax parcel database to update the city's land use data since 1994 (City of Austin 2007g). The NPZD also generated a digital file of land parcel boundaries based on the TCAD tax parcel boundaries, without attaching the original TCAD parcel information. The TCAD tax

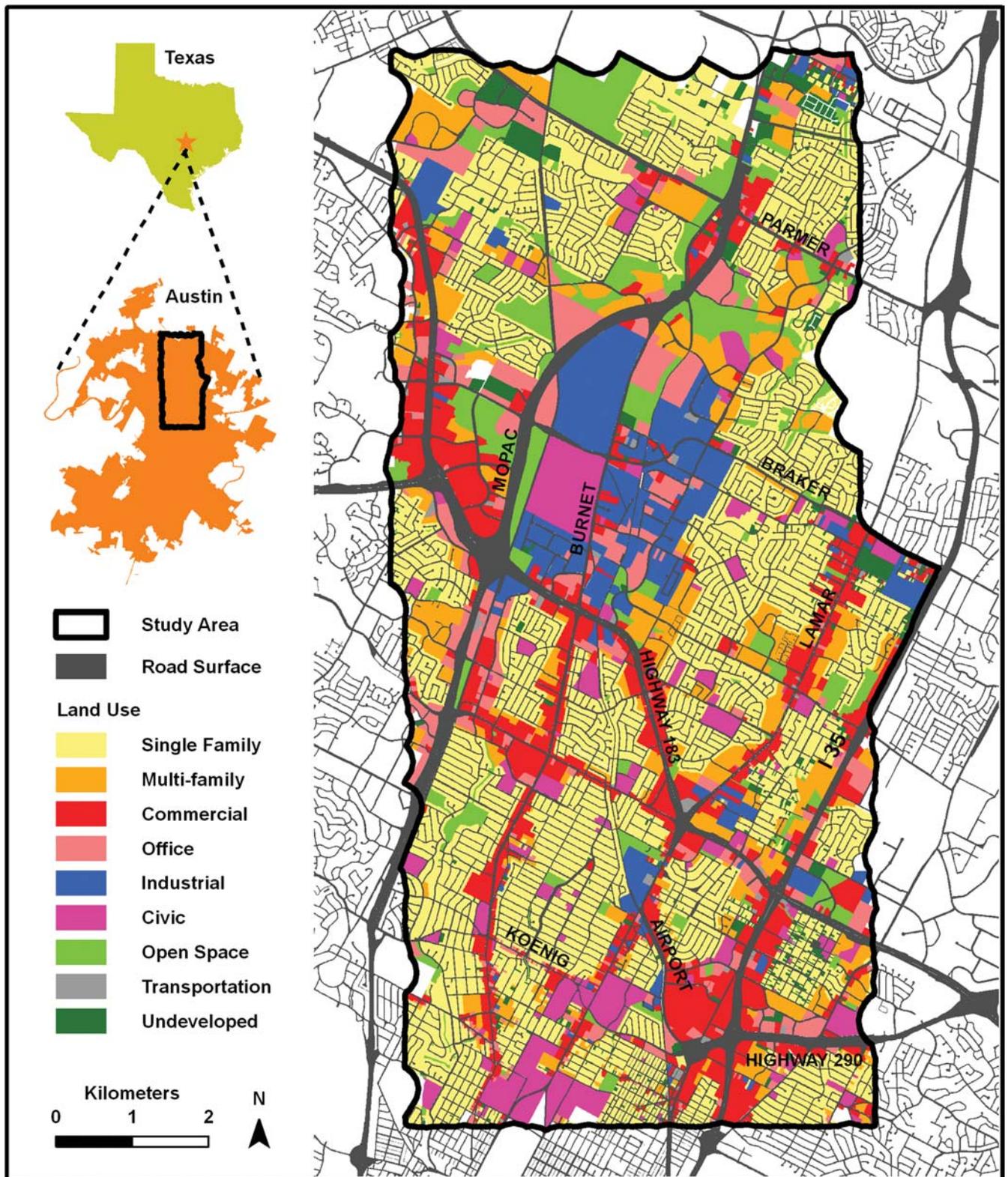


Figure 1. Study area in Austin, Texas, and the ground land use.

parcel database provides information regarding the use of the property based on structures on the property, although the structure usage categories do not necessarily correspond to land use classes defined by the NPZD. Furthermore, because TCAD does not identify the use of tax-exempt properties (e.g., some governmental properties), many government officials and researchers cannot use the TCAD tax parcel database for land use data (City of Austin 2007a, 2007g).

The building and elevation data are both current to 2003. The building data are building footprints in vector polygon format and contain information on the average altitude for individual building roofs. The elevation data are in 0.61-m (2-ft.) contour lines measuring the elevation for the ground surface. The building and elevation data are both generated by Analytical Surveys Incorporated (ASI), which is contracted with the city. ASI first digitized building footprints from aerial photographs and then estimated each individual building's roof altitude from a light detection and ranging (LIDAR) derived surface model, which is also used to generate the ground surface contour lines (Dolph Scott, Senior GIS Analyst, City of Austin, personal conversation, 5 July 2005). To estimate building height for land use classification, we first interpolated a digital elevation model (DEM) from the ground surface contour lines. We then calculated the mean ground elevation for individual building footprints based on the DEM. Finally, building height was computed by subtracting the mean ground elevation from the building roof altitude.

The street data are in vector line format and are current to 2004. They are originally from the U.S. Census 2000 Topologically Integrated Geographic Encoding

and Referencing (TIGER) street data, and the NPZD has updated the data regularly. The street data contain the Census Feature Class Codes (CFCCs), which represent street categories and indicate the level or size of streets: The highest level of streets has a CFCC starting with the number one and includes primary highways; the second level of streets has a CFCC starting with the number two and includes primary roads; the third level of streets has a CFCC starting with the number three and includes secondary roads; the fourth level of streets has a CFCC starting with the number four and includes local, neighborhood, and rural roads (U.S. Census Bureau 2000).

The image data are three-band (green, red, and near-infrared) color infrared (CIR) digital orthophotos with 0.61-m (2-ft.) spatial resolution. The source aerial photographs were taken by ASI during 2003.

The ground land use data are in vector polygon format for 2003. They are generated and updated by the NPZD based on a number of sources, including historical land use data, the TCAD tax parcel database, the city-owned parcels database, building footprint data, natural preserves GIS data, and aerial photographs (City of Austin 2007a). The NPZD divides the city's land use into sixteen general classes, which are further divided into thirty-seven detailed subclasses (City of Austin 2007b). We merged the sixteen land use classes into nine land use classes as our land use classification scheme (Table 1). The merged land use classes are either rare in the study area (including land use of mobile homes, large-lot single family, mining, and utilities) or are not applicable to the tax parcel areas (including land use of streets, water, and unknown).

Table 1. Land use classification scheme of nine classes, corresponding subclasses of land use, and total numbers of parcels for each class in the study area

| Land use class | Descriptions of subclasses of land use | No. of parcels |
|----------------|---|----------------|
| Single family | Mobile homes, large-lot single family, single-family detached, and two-family attached, duplex | 29,252 |
| Multifamily | Three- or fourplex, apartment or condominium, dormitories, and retirement housing | 405 |
| Commercial | 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 | 1,286 |
| Office | Administrative offices, financial services (banks), medical offices, research and development | 473 |
| Industrial | Manufacturing, warehousing, equipment sales and service, recycling and scrap, animal handling, mining facilities | 272 |
| Civic | Semi-institutional housing, hospitals, government services, educational facilities, meeting and assembly facilities, cemeteries, day care facilities | 150 |
| Open space | Parks, recreational facilities, golf courses, preserves and protected areas, water drainage areas and detention ponds | 312 |
| Transportation | Railroad facilities, transportation terminal, aviation facilities, parking facilities, utilities facilities | 132 |
| Undeveloped | Vacant land and land under construction | 743 |

Calculation of Parcel Attributes

We derived fifty parcel attributes for land use classification by overlaying the parcel data with the ancillary data in a GIS. The fifty parcel attributes include ten geometrical, eighteen textural, and twenty-two contextual attributes of parcels. The ten geometrical attributes are related to the size, shape, and height of parcels or buildings: parcel size; parcel shape compactness; the number of buildings; the maximum, standard deviation, and percentage of building area; the maximum and standard deviation of building height; and the maximum and standard deviation of building shape compactness (Table 2, 1–10). The eighteen textural attributes are image-based statistics calculated from the digital orthophotos, including fifteen semivariance texture statistics from lag one to lag fifteen, the average and standard deviation of NDVI, and the impervious cover percentage within parcels (Table 2, 11–28). The twenty-two contextual attributes include the highest CFCC category of nearby streets, fifteen neighborhood building relational statistics from lag one to lag fifteen, a similarity index to neighboring parcels, and five local housing and demographic data from the 2000 U.S. Census (Table 2, 29–50). The five local census statistics are the number of parcels in located block, population density of located block, housing unit occupancy rate of located block, average units in structure of located block group, and median family income of located block group.

These parcel attributes are incorporated for land use classification based on individual theoretical reasons (Figure 2). For example, single-family land use parcels are usually small and rectangular, and they mostly contain only one building. Therefore, the attributes of parcel size, parcel shape compactness, and number of buildings are used to distinguish single-family land use parcels from the others. Compared with civic land use parcels, multifamily land use parcels have buildings that are relatively small and uniform in size, and the parcels generally have higher building-area percentages. Therefore, the attributes of the maximum building area, the standard deviation of building area, and the building-area percentage can be used to separate civic land use parcels from multifamily land use parcels (Table 2, 4–6). Industrial land use parcels usually have higher building-area percentages and impervious cover percentages than other land use parcels (Table 2, 6 and 28). Commercial land use parcels are usually located close to major roads and have a smaller street-category statistic than other types of land use parcels (Table 2, 29). Although

office land use parcels also tend to locate close to major roads, buildings in commercial land use parcels are generally lower in height (Table 2, 7) and, therefore, the attribute of the maximum building height can be used to separate office land use parcels from commercial land use parcels. Furthermore, in contrast to the aforementioned land use parcels that usually contain buildings, open space, undeveloped, and transportation land use parcels usually have no or few buildings (Table 2, 3). Among these three nonbuilding types of land use parcels, transportation land use parcels commonly have less vegetation and higher impervious cover percentages than the other two types of land use parcels (Table 2, 26 and 28).

The shape compactness measure used in this study was calculated by dividing the polygon area with the square of polygon perimeter (Schalkoff 1989). The more curved shape has a smaller compactness measure.

We computed the semivariance at multiple lags as image texture statistics for land use classification because we expect that using texture statistics at multiple lag scales will provide more discriminating power among land use classes. We calculated the semivariance texture statistics based on the infrared band of the CIR digital orthophotos because the infrared band is more sensitive to the spatial arrangement between vegetation and man-made structures and we expect that to be relevant to land use classes.

NDVI measures vegetation abundance by pixels and the value is always between -1 and $+1$. A negative NDVI indicates an absence of vegetation, and a high NDVI indicates areas covered by vegetation. We calculated the average and standard deviation of NDVI within parcels for land use classification with the hypothesis that they relate to land use.

We incorporated the impervious cover percentage as a criterion for land use classification because the city has site development regulations regarding the maximum impervious cover allowed for most zoning districts (City of Austin 2007i); we expect this statistic will help to characterize land use. Impervious covers are areas of covered spaces, paved areas, walkways, and driveways (City of Austin 2007h). To classify impervious covers from the digital orthophotos, we first classified nonvegetated land based on areas with NDVI lower than 0.3. We then partitioned the nonvegetated land into thirty-two classes using the unsupervised algorithm Iterative Self-Organizing Data Analysis Technique (ISODATA). Finally, we manually selected the impervious cover classes by visual interpretation of the digital orthophotos.

Table 2. Average over all parcels by land use class for fifty attributes

| Attribute No. | Sf | Mf | Com | Off | Ind | Civ | Open | Tran | Und |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 9 | 184 | 69 | 110 | 157 | 332 | 191 | 291 | 26 |
| 2 | 556 | 511 | 538 | 534 | 520 | 544 | 389 | 500 | 531 |
| 3 | 103 | 979 | 150 | 142 | 197 | 369 | 6 | 2 | 1 |
| 4 | 210 | 985 | 1,341 | 1,672 | 3,729 | 3,074 | 9 | 12 | 2 |
| 5 | 1 | 164 | 187 | 141 | 349 | 763 | 0 | 0 | 0 |
| 6 | 26 | 30 | 25 | 23 | 31 | 18 | 0 | 0 | 0 |
| 7 | 500 | 868 | 605 | 808 | 673 | 896 | 27 | 10 | 2 |
| 8 | 13 | 16 | 14 | 16 | 15 | 15 | 1 | 0 | 0 |
| 9 | 52 | 45 | 49 | 48 | 49 | 48 | 3 | 1 | 0 |
| 10 | 51 | 37 | 47 | 47 | 47 | 43 | 3 | 1 | 0 |
| 11 | 1,678 | 1,624 | 1,701 | 1,674 | 1,417 | 1,218 | 996 | 1,640 | 1,117 |
| 12 | 2,838 | 2,712 | 2,799 | 2,803 | 2,381 | 2,081 | 1,651 | 2,605 | 1,895 |
| 13 | 3,513 | 3,373 | 3,441 | 3,498 | 2,927 | 2,626 | 2,033 | 3,155 | 2,363 |
| 14 | 3,970 | 3,844 | 3,899 | 4,008 | 3,293 | 3,020 | 2,275 | 3,567 | 2,675 |
| 15 | 4,273 | 4,173 | 4,227 | 4,377 | 3,545 | 3,302 | 2,431 | 3,869 | 2,878 |
| 16 | 4,484 | 4,429 | 4,490 | 4,674 | 3,742 | 3,530 | 2,558 | 4,140 | 3,028 |
| 17 | 4,623 | 4,627 | 4,698 | 4,902 | 3,899 | 3,705 | 2,662 | 4,395 | 3,142 |
| 18 | 4,716 | 4,782 | 4,859 | 5,079 | 4,022 | 3,845 | 2,719 | 4,547 | 3,231 |
| 19 | 4,778 | 4,904 | 4,995 | 5,237 | 4,121 | 3,964 | 2,734 | 4,704 | 3,308 |
| 20 | 4,806 | 4,997 | 5,115 | 5,359 | 4,202 | 4,072 | 2,777 | 4,798 | 3,367 |
| 21 | 4,796 | 5,078 | 5,204 | 5,452 | 4,280 | 4,176 | 2,834 | 4,877 | 3,385 |
| 22 | 4,767 | 5,143 | 5,271 | 5,513 | 4,339 | 4,269 | 2,860 | 4,858 | 3,375 |
| 23 | 4,707 | 5,192 | 5,325 | 5,593 | 4,388 | 4,357 | 2,885 | 4,804 | 3,387 |
| 24 | 4,649 | 5,232 | 5,363 | 5,654 | 4,435 | 4,439 | 2,925 | 4,682 | 3,420 |
| 25 | 4,598 | 5,265 | 5,421 | 5,686 | 4,478 | 4,514 | 2,927 | 4,642 | 3,410 |
| 26 | 383 | 335 | 323 | 336 | 327 | 345 | 381 | 331 | 382 |
| 27 | 110 | 94 | 72 | 87 | 67 | 77 | 84 | 67 | 85 |
| 28 | 39 | 55 | 66 | 61 | 71 | 53 | 32 | 57 | 30 |
| 29 | 393 | 346 | 282 | 300 | 351 | 332 | 349 | 314 | 375 |
| 30 | 716 | 585 | 410 | 380 | 312 | 330 | 11 | 2 | 1 |
| 31 | 1,462 | 1,197 | 847 | 783 | 646 | 676 | 23 | 5 | 3 |
| 32 | 2,041 | 1,673 | 1,192 | 1,098 | 910 | 946 | 32 | 7 | 4 |
| 33 | 2,653 | 2,181 | 1,564 | 1,437 | 1,199 | 1,235 | 41 | 9 | 5 |
| 34 | 3,272 | 2,698 | 1,945 | 1,785 | 1,500 | 1,530 | 50 | 11 | 7 |
| 35 | 3,731 | 3,084 | 2,234 | 2,046 | 1,727 | 1,752 | 57 | 12 | 8 |
| 36 | 4,307 | 3,588 | 2,614 | 2,388 | 2,036 | 2,040 | 66 | 15 | 9 |
| 37 | 4,713 | 3,951 | 2,889 | 2,633 | 2,256 | 2,248 | 73 | 16 | 10 |
| 38 | 5,159 | 4,366 | 3,206 | 2,916 | 2,519 | 2,484 | 80 | 18 | 11 |
| 39 | 5,597 | 4,787 | 3,530 | 3,202 | 2,791 | 2,723 | 86 | 21 | 12 |
| 40 | 5,956 | 5,138 | 3,803 | 3,443 | 3,021 | 2,922 | 92 | 22 | 13 |
| 41 | 6,318 | 5,506 | 4,091 | 3,696 | 3,273 | 3,130 | 97 | 24 | 14 |
| 42 | 6,625 | 5,823 | 4,339 | 3,914 | 3,492 | 3,309 | 102 | 26 | 14 |
| 43 | 6,935 | 6,153 | 4,598 | 4,143 | 3,728 | 3,497 | 106 | 28 | 15 |
| 44 | 7,203 | 6,440 | 4,824 | 4,343 | 3,937 | 3,660 | 110 | 29 | 16 |
| 45 | 15 | 176 | 437 | 484 | 541 | 842 | 173 | 128 | 66 |
| 46 | 43 | 55 | 36 | 37 | 43 | 30 | 52 | 36 | 74 |
| 47 | 254 | 316 | 189 | 173 | 82 | 181 | 181 | 198 | 180 |
| 48 | 29 | 28 | 26 | 18 | 28 | 32 | 35 | 24 | 49 |
| 49 | 16 | 38 | 32 | 41 | 64 | 28 | 31 | 32 | 23 |
| 50 | 56 | 44 | 45 | 52 | 51 | 51 | 59 | 44 | 50 |

Notes: Sf = single family; Mf = multifamily; Com = commercial; Off = office; Ind = industrial; Civ = civic; Open = open space; Tran = transportation; Und = undeveloped. 1 = parcel size (100 * square meters); 2 = parcel shape compactness (1/10,000); 3 = number of buildings (1/100); 4 = maximum building's area (square meters); 5 = standard deviation of the building's area (square meters); 6 = building-area percentage (%); 7 = maximum building's height (meters/100); 8 = standard deviation of the building's height (meters/100); 9 = maximum building shape compactness (1/1,000); 10 = standard deviation of building shape compactness (1/1,000); 11 to 25 = image texture statistics at lag one to lag fifteen, respectively; 26 = average Normalized Difference Vegetation Index (NDVI; in unsigned sixteen bit); 27 = standard deviation of NDVI (in unsigned sixteen bit); 28 = impervious cover percentage (%); 29 = the highest Census Feature Class Code category of nearby streets (1/100); 30 to 44 = neighborhood building relational statistics at lag one to lag fifteen, respectively; 45 = similarity index to neighboring parcels (square meters/10); 46 = number of parcels in located block; 47 = population density of located block (100 * persons per square kilometer); 48 = housing unit occupancy rate of located block (%); 49 = average units in structure of located block group (1/10); 50 = median family income of located block group (*1,000).

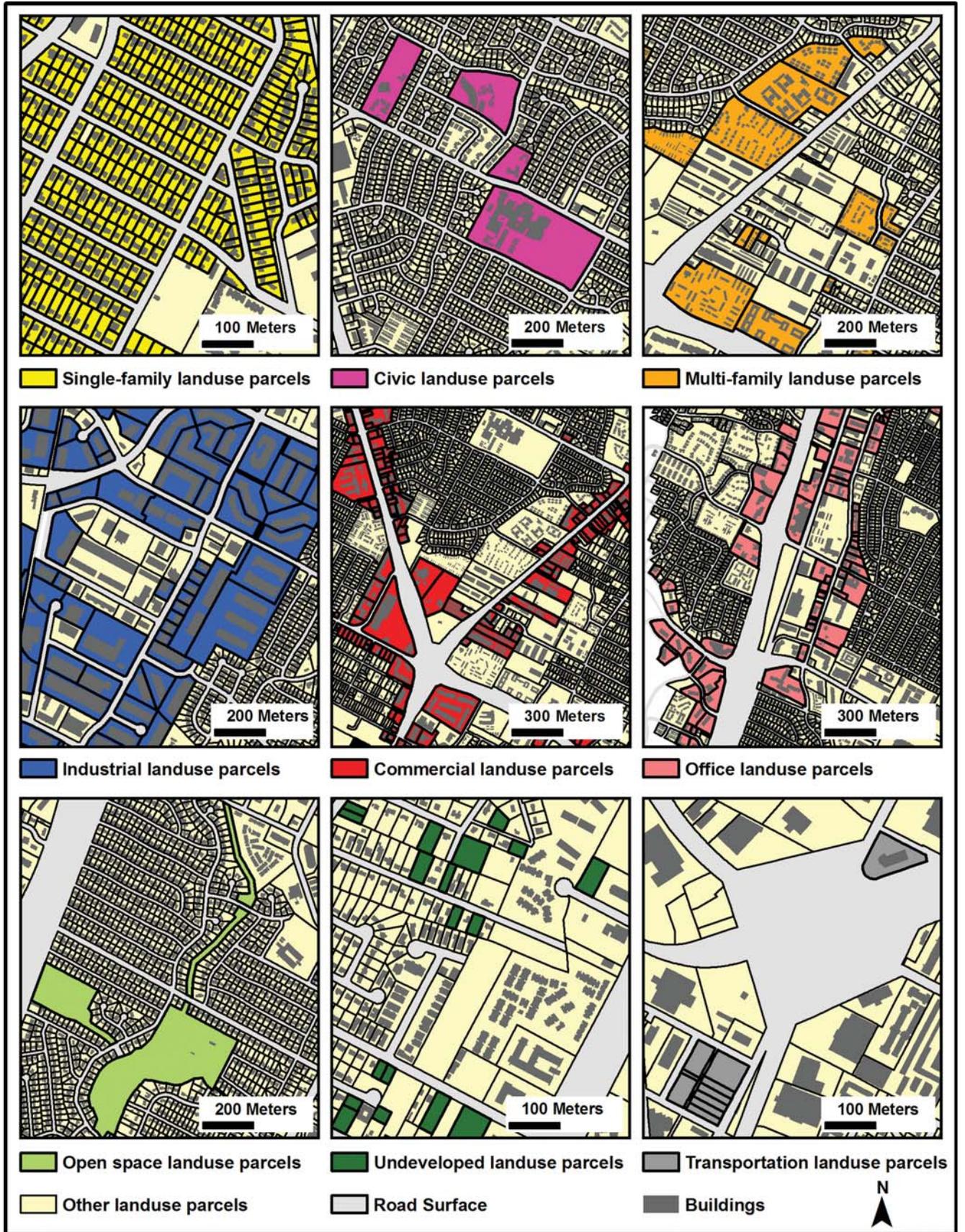


Figure 2. Examples of single-family, civic, multifamily, industrial, commercial, office, open space, undeveloped, and transportation land use parcels.

Downloaded By: [University at Buffalo, the State University of New York (SUNY)] At: 21:08 15 September 2009

The neighborhood building relational statistics at lag one to lag fifteen are the semivariances calculated within a 30-m buffer for individual parcels from a binary building image in which building pixels are designated a value of one and nonbuilding pixels are given a value of zero. The statistics are used to indicate the complexity of building spatial patterns in the neighborhood and have been shown to relate to land use classes (Wu, Xu, and Wang 2006). The more frequently building and nonbuilding areas are interspersed in a neighborhood, the higher the statistics tend to be. For example, single-family and multifamily land use neighborhoods generally have relatively high building relational statistics due to their more heterogeneous spatial patterns of buildings (Table 2, 30–44). Considering the relative size between parcels and buildings, we calculated the statistics based on a 30-m neighborhood, except for parcels without buildings.

We incorporated a similarity index to neighboring parcels for land use classification because parcels of the same land use, particularly single-family land use, tend to cluster together. The similarity index compares parcel attributes with those of neighboring parcels. A study by Wu, Silván-Cárdenas, and Wang (2007) has shown that the parcel attribute of the standard deviation of the building's area provides the most separability among land use classes among the twelve parcel attributes used. Therefore, we computed the similarity index as the average difference of this parcel attribute with those of contiguous parcels.

To investigate whether relevant housing and demographic statistics from the U.S. Census data can be useful for detailed urban land use classification, we reviewed available census statistics at the census block and block group levels. Based on relevant census statistics at the finest scale available, we decided to incorporate five parcel attributes for land use classification: number of parcels in located block, population density of located block, housing unit occupancy rate of located block, average units in structure of located block group, and median family income of located block group.

Census blocks in urbanized neighborhoods are usually small and contain numerous parcels. In this study, we tested the hypothesis that the parcel attribute of the number of parcels in the located block (Table 2, 46) is useful for land use classification.

Residential neighborhoods usually have higher population densities than nonresidential neighborhoods, and multifamily neighborhoods usually have higher population densities than single-family neighborhoods. Therefore, we incorporated the parcel attribute of

population density of located block (Table 2, 47) for land use classification.

Single-family housing units usually have a higher housing unit occupancy rate than multifamily housing units, and multifamily structures generally have more units than single-family structures. Moreover, single-family residents generally have higher median family income than multifamily residents do. For these reasons, the parcel attributes of housing unit occupancy rate of located block, average units in structure of located block group, and median family income of located block group (Table 2, 48–50) were tested for land use classification.

It is worth noting that census blocks, as well as block groups, are generally much larger than parcels. For example, 800 of the 1,398 census blocks in the study area contain a number of parcels between twelve and thirty-two (238 blocks contain between 1 and 11 parcels; 360 blocks contain between 33 and 330 parcels). As a result, census statistics at the block or block group level may not be well applied to the parcel level. In other words, the land use of a local parcel might not be relevant to the census statistics of the local block or block group.

Land Use Classification Based on Parcel Attributes

After obtaining the fifty parcel attributes, we randomly selected half of the parcels from each land use as training parcels to train classification algorithms and the other half as testing parcels to assess classification accuracy. Single-family land use made up a great majority (approximately 88.6 percent) of the total number of parcels (Table 1), which is mainly because of its small parcel size and large coverage (approximately 40 percent) of the study area (Figure 1). In fact, single-family land use parcels make up approximately 90 percent of the total land use parcels in Austin and cover approximately 50 percent of the area.

Because we are interested in how many parcels, instead of how many land areas, are correctly classified based on the fifty parcel attributes, we generated an artificial, fifty-band, 1-m resolution, fishnet image with the number of pixels the same as the number of parcels and the pixel values the same as the parcel attributes to use for classification.² This approach allows us to efficiently test different classification algorithms on a relatively small, artificial image.

We tested a variety of existing classification algorithms, including parallelepiped, minimum distance, Mahalanobis distance, maximum likelihood, spectral

Table 3. Overall accuracy, quantity error, and location error for different classification algorithms

| Classification algorithm | Overall accuracy (%) | Quantity error (%) | Location error (%) |
|--------------------------------------|----------------------|--------------------|--------------------|
| Decision tree | 95.6 | 1.9 | 2.5 |
| Neural network | 88.6 | 11.4 | 0.0 |
| Majority rule-based naive model | 88.6 | 11.4 | 0.0 |
| Spatial dependence-based naive model | 79.4 | 0.2 | 20.4 |
| Mahalanobis distance | 78.8 | 11.7 | 9.5 |
| Binary encoding | 76.9 | 17.9 | 5.2 |
| Parallelepiped | 74.1 | 21.2 | 4.7 |
| Minimum distance | 71.6 | 23.1 | 5.3 |
| Spectral angular mapper | 66.9 | 27.6 | 5.5 |
| Maximum likelihood | 2.8 | 96.9 | 0.3 |

angular mapper, binary encoding, neural network, and decision tree. In addition, we applied two baseline models as the benchmark of comparison as described earlier.

Accuracy Assessment of Classification Results

Comparing different classification algorithms, the decision tree algorithm has the highest overall accuracy of 95.6 percent; its location error and quantity error are also relatively low (Figure 3 and Table 3). A decision tree algorithm groups data into hierarchical structures through recursive partitioning of predictor variables into smaller, more homogeneous groups. According to McCauley and Goetz (2004), the advantage of a decision tree algorithm is that it is capable of handling both numeric and categorical inputs and it does not require assumptions regarding the statistical properties of the input data. Because many of the parcel attributes are not normally distributed (some were illustrated in Wu, Silván-Cárdenas, and Wang 2007), it is reasonable that a decision tree algorithm is effective for land use classification.

The partitioning process in a decision tree algorithm is based on predetermined decision rules. We generated decision rules using the Quick, Unbiased, and Efficient Statistical Tree (QUEST) algorithm (Loh 2005), which examines all possible binary splits of the data along each predictor variable to select the split that most reduces node impurities (Loh and Shih 1997). We then imported the generated decision rules into the decision

tree tool in ENVI software (ENVI 2007) to classify the parcel image.

Table 3 shows that the decision tree classification algorithm has a location error of 2.5 percent and a quantity error of 1.9 percent. It indicates that if the algorithm can improve specification of quantity, then the overall accuracy could potentially increase by 1.9 percentage points. Also, given the specified quantity of classes in the classified map, if the algorithm can improve specification of location, the overall accuracy could potentially increase by 2.5 percentage points.

The majority rule-based naive model has a certain amount of quantity error but no location error because the model assigns all parcels as single-family land use based on its naive assumption. Given that all parcels are assigned as single-family land use, there is no room to improve the specification of class location. In contrast, the spatial dependence-based naive model has relatively small quantity error (0.2 percent) and large location error (20.4 percent). The small quantity error is because the model assigns approximately the same quantity of classes as in the training parcels to the testing parcels based on its naive assumption; because our sampling scheme is based on the 50–50 split strategy, the model resulted in small error in specifying class quantity for testing parcels. Furthermore, the large location error of the spatial dependence-based naive model indicates that the spatial dependence of land use classes is relatively weak at the parcel level.

The neural network classification algorithm (as well as the majority rule-based naive model) produces the second highest overall accuracy. Interestingly, the neural network algorithm has exactly the same classification results as from the majority rule-based naive model that classified all parcels as single-family land use. A neural network algorithm is a machine-learning computer program that simulates the trial-and-error learning process of human brains to generate optimal network weights between variables and produce a model for a given data set (NeuroDimension 2007). Past studies have reported the effectiveness of a neural network algorithm for land use and land cover classification (Jensen, Qiu, and Patterson 2001; Seto and Liu 2003). As with the decision tree algorithm, the neural network algorithm can work with categorical data and does not make statistical assumptions. Nevertheless, a neural network is a black box regarding the internal decision process and, therefore, may be less preferred than a decision tree algorithm that is based on clear, specific decision rules.

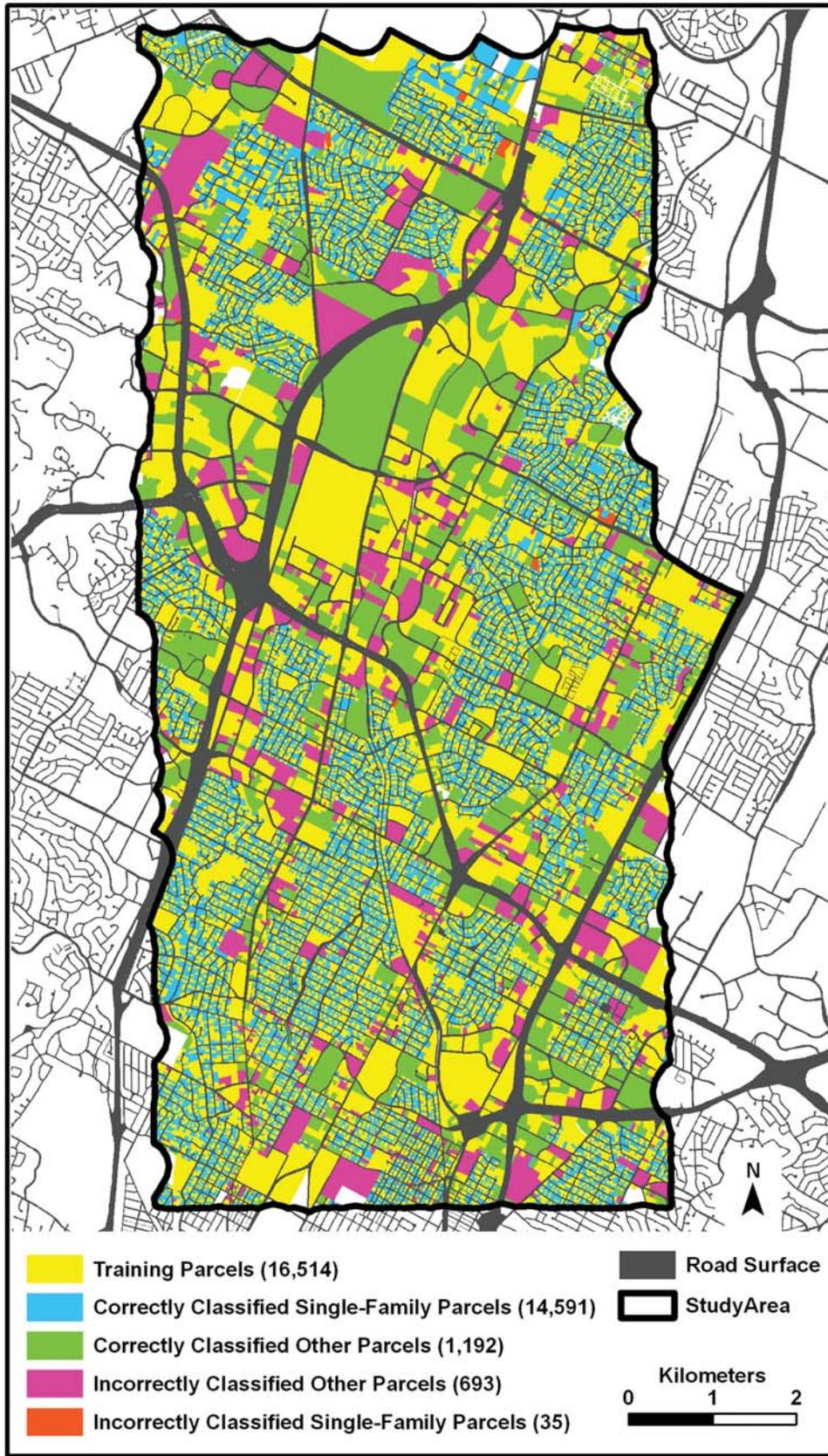


Figure 3. Classification results from a decision tree classification algorithm. The number of parcels is indicated in parentheses.

Table 4. Confusion matrix by the number of parcels from the decision tree classification algorithm

| Classified class | Ground class | | | | | | | | | |
|-----------------------|--------------|-----|-----|-----|-----|-----|------|------|-----|--------|
| | Sf | Mf | Com | Off | Ind | Civ | Open | Tran | Und | Total |
| Single family (Sf) | 14,591 | 21 | 120 | 70 | 17 | 23 | 2 | 0 | 1 | 14,845 |
| Multifamily (Mf) | 4 | 134 | 19 | 10 | 4 | 7 | 0 | 0 | 0 | 178 |
| Commercial (Com) | 28 | 37 | 456 | 86 | 52 | 26 | 3 | 1 | 0 | 689 |
| Office (Off) | 2 | 4 | 15 | 46 | 2 | 3 | 0 | 0 | 0 | 72 |
| Industrial (Ind) | 0 | 2 | 32 | 23 | 58 | 1 | 0 | 0 | 0 | 116 |
| Civic (Civ) | 0 | 4 | 1 | 1 | 3 | 15 | 1 | 0 | 0 | 25 |
| Open space (Open) | 1 | 0 | 0 | 0 | 0 | 0 | 106 | 10 | 18 | 135 |
| Transportation (Tran) | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 28 | 3 | 34 |
| Undeveloped (Und) | 0 | 0 | 0 | 0 | 0 | 0 | 41 | 27 | 349 | 417 |
| Total | 14,626 | 202 | 643 | 236 | 136 | 75 | 156 | 66 | 371 | 16,511 |

Observing the confusion matrix (Table 4) and the producer's and user's accuracy table (Figure 4) for the decision tree classification algorithm, we see that the single-family land use class has the highest producer's and user's accuracies among all land use classes, which can be explained by its characteristic and uniform parcels that are small, rectangular, and contain only a single low-rise building (Figure 2). Next to single-family land use, undeveloped land use has the best classification results, with both producer's and user's accuracies above 80 percent. A large amount of undeveloped land use parcels are distinguished by their small size, similar to single-family land use parcels but without buildings (Figure 2).

The confusion matrix shows that undeveloped land use parcels have a high degree of confusion with open space and transportation land use parcels, as most of these parcels do not contain buildings. Particularly, transportation land use has a large proportion (twenty-seven of sixty-six) of its testing parcels misclassified into undeveloped land use, which causes its relatively low

producer's accuracy (42 percent). Transportation land use contains a number of land use subclasses (Table 1) with dissimilar parcel attributes and is, therefore, relatively difficult to classify. For example, the impervious cover percentage can be quite low for railroad facilities but relatively high for parking facilities.³ Land parcel size can be quite small for a parking lot but relatively large for a bus transfer center (Figure 2).

In contrast to undeveloped, open space, and transportation land uses that usually do not have buildings within land parcels, the land uses of single family, multifamily, commercial, office, industrial, and civic generally have one or more buildings within land parcels. The confusion matrix shows that there is little confusion between the building type of land use and the nonbuilding type of land use. Among these six building types of land use, office, industrial, and civic land uses have relatively low producer's accuracy. The confusion matrix shows that the low producer's accuracies are mainly due to confusion with the commercial land use class. Commercial land use has the most subclasses of land use classes (Table 1), and each land use subclass can have very different parcel attributes. For example, a commercial land use parcel containing numerous high-rise shopping centers is considerably large, whereas a commercial land use parcel containing a single low-rise retail store is relatively small. This heterogeneous nature of commercial land use causes its parcel attributes to be similar to those of other land use and results in subsequent confusion in classification. As for why other land uses tend to be misclassified into commercial land use instead of the opposite, it is likely related to individual land use characteristics and the nature of a decision tree classification algorithm rather than the relatively large number of commercial land use parcels compared to the three easily misclassified land uses.

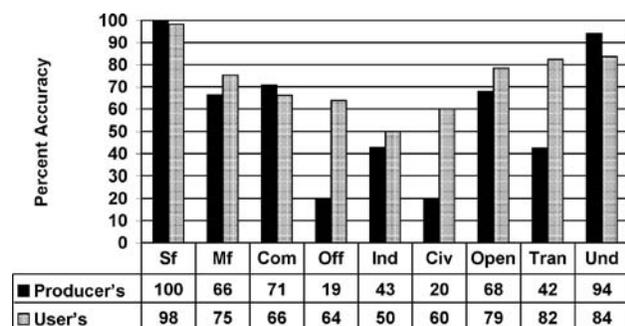


Figure 4. Producer's and user's accuracies by land use classes from the decision tree classification algorithm. See Table 4 for land use abbreviations.

Of the six building types of land use, multifamily land use has similarly high producer's and user's accuracies compared to commercial land use. However, in contrast to commercial land use, multifamily land use has the fewest subclasses of the six building types of land use. The high classification accuracy of multifamily land use can be explained by the fact that approximately 79 percent of parcels within multifamily land use contain apartments or condominiums, which have relatively uniform building characteristics.

Comparison Among Parcel Attributes

To compare the ability of different parcel attributes to separate land use classes, we conducted the signature separability analysis (Wu, Silván-Cárdenas, and Wang 2007). A signature separability statistic indicates how well two land use classes can be distinguished based on specific parcel attributes. We used the signature separability statistic of Bhattacharya (or Jeffries–Mastusuta) distance (PCI Geomatics 2007) in this study. The measure is a real value between zero and two, where zero indicates complete overlap between two classes and two indicates a complete separation between the classes. We calculated the average Bhattacharya distance between paired classes based on individual and combinations of parcel attributes. The derived Bhattacharya distance is further rescaled to zero to one to provide a separability measure that is easy to interpret (Table 5). The results show that the separability measure based on all fifty parcel attributes is 0.98, which is very close to a complete separability measure of one. The ten geometrical attributes of parcels (1–10) have a signature separability measure (of 0.87) higher than that of the twenty-two contextual attributes (0.78, 29–50), which further has a higher separability measure than that of the eighteen image-based textural attributes (0.66, 11–14).

Among the ten geometrical attributes, the eight building-relevant attributes (3–10, with a separability measure of 0.86) are more important than the other two parcel geometry-based attributes (1 and 2, with a separability measure of 0.43). Within the eight building-relevant attributes, the three building area-related attributes (4–6, with a separability measure of 0.81) are the most important, next are the two building height-related attributes (7 and 8, with a separability measure of 0.66), the two building shape-related attributes (9 and 10, with a separability measure of 0.59), and finally the building-number attribute (3, with a separability measure of 0.57).

Individually the fifteen image texture statistics (11–25) provide relatively low separability measures (between 0.05 and 0.08), yet collectively they provide a relatively high separability measure (of 0.62), which indicates that semivariance texture statistics at different lags have relatively distinctive capabilities for discriminating among land use classes. In contrast, the fifteen neighborhood building relational statistics (30–44) have relatively overlapping capabilities for discriminating between land use classes, because the separability measures for individual statistics (between 0.49 and 0.50) are not very different from their collective separability measure (of 0.73).

The parcel attribute of the impervious cover percentage (28) has a separability measure (of 0.20) lower than we expect, probably because many impervious cover surfaces are covered under tree canopies and cannot be correctly classified from aerial photographs. Advanced methods for impervious cover classification may be needed to accurately calculate this parcel attribute.

The parcel attribute of the highest CFCC category of nearby streets (29) has a relatively low separability measure (0.11). Although many commercial land use and multifamily land use parcels are close to major roads (Figure 1), other land use classes do not necessarily relate to this attribute.

The parcel attribute of the similarity index to neighboring parcels does not have a high separability measure (of 0.31) as expected. Overall, parcels of the same classes do not have similar similarity indexes or have similarity indexes that are not distinctive from those of other classes.

The five census statistics-based parcel attributes all have relatively low separability measures, indicating that the census statistics at the block or block group levels cannot be effectively applied to the parcel level for urban land use classification. More analyses need to explore if these census housing and demographic analyses are really related to urban land use by using land use data at a spatial scale that is comparable to (or larger than) the scales of census block or block groups.

Discussion

When compared with the study by Wu, Silván-Cárdenas, and Wang (2007), this study improves the overall accuracy by two percentage points and the kappa coefficient by eight percentage points. In addition, the producer's accuracies for individual land use classes all improved except for the commercial class (Figure 5).

Table 5. The signature separability based on individual and combinations of parcel attributes

| No. | Parcel attribute | | Class separability | | | |
|-----|--|------|--------------------|------|------|------|
| 1 | Parcel size | 0.38 | 0.43 | 0.43 | 0.87 | 0.98 |
| 2 | Parcel shape compactness | 0.08 | | | | |
| 3 | Number of buildings | 0.57 | 0.57 | 0.86 | | |
| 4 | Maximum building's area | 0.57 | 0.81 | | | |
| 5 | Standard deviation of the building's area | 0.66 | | | | |
| 6 | Building-area percentage | 0.48 | | | | |
| 7 | Maximum building's height | 0.50 | 0.66 | | | |
| 8 | Standard deviation of the building's height | 0.53 | | | | |
| 9 | Maximum building shape compactness | 0.49 | 0.59 | | | |
| 10 | Standard deviation of building shape compactness | 0.48 | | | | |
| 11 | Image texture statistic at lag one | 0.08 | 0.62 | 0.66 | 0.66 | |
| 12 | Image texture statistic at lag two | 0.07 | | | | |
| 13 | Image texture statistic at lag three | 0.07 | | | | |
| 14 | Image texture statistic at lag four | 0.07 | | | | |
| 15 | Image texture statistic at lag five | 0.07 | | | | |
| 16 | Image texture statistic at lag six | 0.07 | | | | |
| 17 | Image texture statistic at lag seven | 0.06 | | | | |
| 18 | Image texture statistic at lag eight | 0.06 | | | | |
| 19 | Image texture statistic at lag nine | 0.06 | | | | |
| 20 | Image texture statistic at lag ten | 0.06 | | | | |
| 21 | Image texture statistic at lag eleven | 0.06 | | | | |
| 22 | Image texture statistic at lag twelve | 0.05 | | | | |
| 23 | Image texture statistic at lag thirteen | 0.05 | | | | |
| 24 | Image texture statistic at lag fourteen | 0.05 | | | | |
| 25 | Image texture statistic at lag fifteen | 0.05 | | | | |
| 26 | Average NDVI | 0.19 | 0.29 | | | |
| 27 | Standard deviation of NDVI | 0.13 | | | | |
| 28 | Impervious cover percentage | 0.20 | 0.20 | | | |
| 29 | Highest CFCC category of nearby streets | 0.11 | 0.11 | 0.78 | 0.78 | |
| 30 | Neighborhood building relational statistic at lag one | 0.49 | 0.73 | | | |
| 31 | Neighborhood building relational statistic at lag two | 0.49 | | | | |
| 32 | Neighborhood building relational statistic at lag three | 0.50 | | | | |
| 33 | Neighborhood building relational statistic at lag four | 0.50 | | | | |
| 34 | Neighborhood building relational statistic at lag five | 0.50 | | | | |
| 35 | Neighborhood building relational statistic at lag six | 0.50 | | | | |
| 36 | Neighborhood building relational statistic at lag seven | 0.50 | | | | |
| 37 | Neighborhood building relational statistic at lag eight | 0.51 | | | | |
| 38 | Neighborhood building relational statistic at lag nine | 0.51 | | | | |
| 39 | Neighborhood building relational statistic at lag ten | 0.51 | | | | |
| 40 | Neighborhood building relational statistic at lag eleven | 0.51 | | | | |
| 41 | Neighborhood building relational statistic at lag twelve | 0.51 | | | | |
| 42 | Neighborhood building relational statistic at lag thirteen | 0.50 | | | | |
| 43 | Neighborhood building relational statistic at lag fourteen | 0.50 | | | | |
| 44 | Neighborhood building relational statistic at lag fifteen | 0.50 | | | | |
| 45 | Similarity index to neighboring parcels | 0.31 | 0.31 | | | |
| 46 | Number of parcels in located block | 0.04 | 0.29 | | | |
| 47 | Population density of located block | 0.09 | | | | |
| 48 | Housing unit occupancy rate of located block | 0.02 | | | | |
| 49 | Average units in structure of located block group | 0.10 | | | | |
| 50 | Median family income of located block group | 0.03 | | | | |

Notes: NDVI = Normalized Difference Vegetation Index; CFCC = Census Feature Class Code.

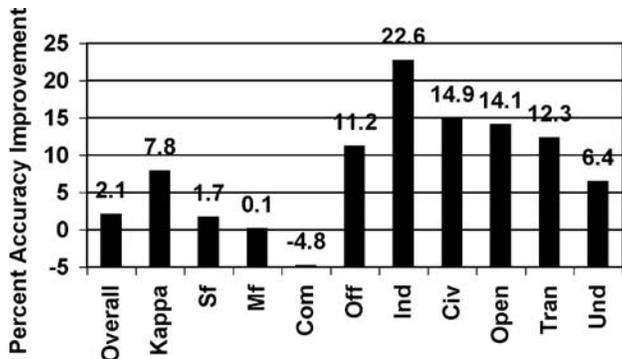


Figure 5. Improvements of overall accuracy, kappa coefficient, and producer's accuracies of individual classes compared to the study by Wu, Silván-Cárdenas, and Wang (2007). See Table 4 for land use abbreviations.

Producer's accuracies for six of the nine land use classes improved more than five percentage points. Although the accuracy improvements were achieved as the result of increasing the classification attributes from twelve to fifty, it is worth noting that many parcel attributes are in fact the same properties presented in different statistical forms (e.g., average NDVI and standard deviation of NDVI) or the same types of statistics calculated with different parameters (e.g., the image texture statistics in fifteen different lags) based on theoretical reasons of individual attributes. From this point of view, there are only seventeen types of parcel attributes used for land use classification. Nevertheless, incorporating the same types of attributes in different statistical forms or parameters is necessary to achieve a certain level of classification accuracy because the signature separability analysis indicated that all fifty parcel attributes play a role in classifying land use (Table 5).

In reviewing past per field classification, we noted that the field boundaries can be utilized in a preclassification stage, a postclassification stage, or both. This study uses field boundaries in a preclassification stage by first deriving GIS attributes of land parcels and then classifying land use based on the parcel attributes. Future research might use field boundaries in a postclassification stage to improve classification accuracy by, for example, modifying preclassified land use classes based on (additional) parcel attributes. Future studies may also apply this per field classification approach to other geographic regions and test whether our findings regarding the relative importance of different parcel attributes for land use classification and the relative classification accuracies between different land use classes are consistent.

This study relies on different ancillary data sets to obtain parcel attributes for land use classification. A potential problem of spatial analysis of different data sets is the mismatch of spatial precision and time frame between data sets, and that will contribute to the final classification error. For example, the road data and the parcel data do not match precisely with the image data and the building data in geographical locations, and the road data are one year later than other data sets. Even data sets from the same year still might not match well with each other if they are updated at different times of the year.

This study utilizes the digital boundary data of land parcels from the TCAD for our per field land use classification. Land parcel boundary data for property tax purposes are mostly available for urban areas. Therefore, the major application area of the proposed land use classification method is urban areas.

Many parcel attributes used in this study are derived from existing GIS data that are produced through the extensive human labor of manual digitizing and field work, such as building footprints and road networks. With the advancement of high-resolution remote sensing data and data processing techniques, automatic extraction of some of the GIS data and parcel attributes are now possible. For example, building height can be derived by analyzing the multiple return signals from LIDAR data (Rottensteiner 2003; Luo and Gavrilova 2006). Building footprints and road networks may be derived from IKONOS images using advanced feature extraction techniques (Fraser, Baltsavias, and Gruen 2002; Kim et al. 2004).

The proposed classification approach may not be applicable to fast-growing urban areas in developing countries (where up-to-date land use maps are in eminent need) because the required GIS data for land use classification are not readily available or too expensive to collect. As for major urban areas in developed countries (where land use data are available and accurate for most parcels), producing a land use map using the proposed method may not be desirable to city officials if it is more feasible to produce an updated land use map based on existing sources (such as the latest updated land use map, appraisal records, and development records) in conjunction with a certain amount of field check (as in the case of Austin, Texas). On the other hand, communities in metropolitan areas may be willing to spend the money needed to acquire the GIS data to produce land use maps for fast-growing areas considering the advantages of the proposed method as a relatively robust, accurate, and efficient approach for updating land use

data. In the long run, we expect the proposed land use classification methods will be of practical use for local government agencies when the required GIS data become relatively accessible and more affordable.

Conclusions

This study presents a per field approach for detailed urban land use classification based on land parcel attributes derived from relevant GIS and remote sensing data. Six building types of land use (single family, multifamily, commercial, office, industrial, and civic) and three nonbuilding types of land use (open space, transportation, and undeveloped) were classified based on fifty parcel attributes, which include ten geometrical attributes, eighteen image-based textural attributes, and twenty-two contextual attributes. The signature separability analysis shows that the most important attributes for land use classification are the eight building-relevant geometrical attributes, and of those eight attributes the three related to building areas are the most important.

Land use classification with a decision tree classification algorithm has the highest overall accuracy of 95.6 percent and kappa coefficient of 0.78 percent, which also has relatively low location error and quantity error of 2.5 percent and 1.9 percent, respectively. Two naive, baseline models based on the majority rule and the spatial autocorrelation rule have overall accuracy of 88.6 percent and 79.4 percent, respectively.

The confusion matrix from the decision tree classification algorithm shows that there is not much classification confusion between building-type land use and non-building-type land use. Of the six building types of land use, single-family, multifamily, and commercial land uses have relatively high classification accuracy. Of the three nonbuilding types of land use, open space and undeveloped land uses have relatively high classification accuracy. Of all the land use classes, single-family land use can be most effectively distinguished from other land use classes due to its characteristic and uniform parcels. After single-family land use, undeveloped land use has the best classification results, with both producer's and user's accuracies above 83 percent. On the other hand, the decision tree classification algorithm found it particularly challenging to classify office and civic land uses accurately.

Digital image classification in remote sensing provides a systematic and efficient way to generate a land use and land cover map. Urban land use classification has a fundamental limitation in that some land

use classes are heterogeneous in nature or have similar physical characteristics to others. As a result, misclassification and confusion among classes are unavoidable. Although generally a land cover map with 85 percent overall accuracy is considered valid and acceptable to apply toward other applications, the accuracy requirement of an urban land use map used by local government is commonly higher. Therefore, digital classification of urban land use has low classification accuracy unacceptable to land use practitioners. This study of urban land use classification has achieved promising results. As classification accuracy becomes better with the advancement of feasible remote sensing and GIS data and techniques, digital classification of urban land use is expected to be a viable approach for real-world applications in the future.

Acknowledgments

While this article was being written, Shuo-Sheng Wu was supported by a post-doctoral fellowship sponsored by the United States Geological Survey (USGS) Center of Excellence for Geospatial Information Science (CEGIS) through University Consortium for Geographic Information Science (UCGIS). The authors would like to thank the three anonymous reviewers and the editor, Mei-Po Kwan, who provided constructive comments that greatly improved this article from its original form.

Notes

1. Grandfathering is the practice of exempting current rights holders from a new regulation or legal qualification.
2. For computational convenience, all parcel attributes are rescaled to a value between 0 and 65,535 (unsigned sixteen bit) before transforming to pixel values.
3. The standard deviation of this parcel attribute for transportation land use is indeed the highest among all land use classes. See Wu, Silván-Cárdenas, and Wang (2007) for details.

References

- Aplin, P., and P. Atkinson. 2001. Sub-pixel land cover mapping for per-field classification. *International Journal of Remote Sensing* 22 (14): 2853–58.
- Aplin, P., P. Atkinson, and P. Curran. 1999. Fine spatial resolution simulated satellite sensor imagery for land cover

- mapping in the United Kingdom. *Remote Sensing of Environment* 68 (3): 206–16.
- Barnsley, M. J., M. S. Alan, and S. L. Barr. 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*, ed. V. Mesev, 83–108. London: Taylor & Francis.
- Benediktsson, J. A., J. A. Palmason, and J. R. Sveinsson. 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing* 43 (3): 480–91.
- Berberoglu, S., C. D. Lloyd, P. M. Atkinson, and P. J. Curran. 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. *Computers & Geosciences* 26 (4): 385–96.
- Burrough, P. A., and R. McDonnell, 1998. *Principles of geographical information systems*. New York: Oxford University Press.
- City of Austin. 2007a. 2003 *Land use inventory overview and methodology*. <http://www.ci.austin.tx.us/landuse/survey2003.htm> (last accessed 30 August 2007).
- . 2007b. 2003 *Land use metadata*. ftp://coageoid01.ci.austin.tx.us/GIS-Data/Regional/landuse_2003.htm (last accessed 30 August 2007).
- . 2007c. *Austin, Texas: Basic facts*. <http://www.ci.austin.tx.us/citymgr/basicfac.htm> (last accessed 30 August 2007).
- . 2007d. *City of Austin GIS data sets*. ftp://coageoid01.ci.austin.tx.us/GIS-Data/Regional/coa_gis.html (last accessed 30 August 2007).
- . 2007e. *City of Austin—Land use GIS inventory and analysis*. <http://www.ci.austin.tx.us/landuse> (last accessed 30 August 2007).
- . 2007f. *Land use survey methodology*. <http://www.ci.austin.tx.us/landuse/survey.htm> (last accessed 30 August 2007).
- . 2007g. *Travis Central Appraisal District (TCAD) parcel GIS project*. <http://www.ci.austin.tx.us/landuse/tcadgis.htm> (last accessed 30 August 2007).
- . 2007h. §25-1-23 *Impervious cover measurement, Austin City Code*. [http://www.amlegal.com/austin_nxt/gateway.dll/Texas/austin/title25landdevelopment/chapter25-1generalrequirementsandprocedu?f=templates\\$fn=altmain-nf.htm\\$3.0#JD_25-1-23](http://www.amlegal.com/austin_nxt/gateway.dll/Texas/austin/title25landdevelopment/chapter25-1generalrequirementsandprocedu?f=templates$fn=altmain-nf.htm$3.0#JD_25-1-23) (last accessed 30 August 2007).
- . 2007i. §25-2-492 *Site development regulations, Austin City Code*. [http://www.amlegal.com/austin_nxt/gateway.dll/Texas/austin/title25landdevelopment/chapter25-2zoning?f=templates\\$fn=altmain-nf.htm\\$3.0#JD_25-2-492](http://www.amlegal.com/austin_nxt/gateway.dll/Texas/austin/title25landdevelopment/chapter25-2zoning?f=templates$fn=altmain-nf.htm$3.0#JD_25-2-492) (last accessed 30 August 2007).
- Congalton, R. G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37 (1): 35–46.
- Dean, A. M., and G. M. Smith. 2003. An evaluation of per-parcel land cover mapping using maximum likelihood class probabilities. *International Journal of Remote Sensing* 24 (14): 2905–20.
- de Jong, S. M., and F. D. van der Meer, eds. 2004. *Remote sensing image analysis: Including the spatial domain*. Amsterdam: Kluwer Academic.
- De Wit, A. J. W., and J. G. P. W. Clevers. 2004. Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing* 25 (20): 4091–112.
- Donnay, J. P., and D. Unwin. 2001. Modeling geographical distributions in urban areas. In *Remote sensing and urban analysis*, ed. J. P. Donnay, M. J. Barnsley, and P. A. Longley, 205–24. New York: Taylor & Francis.
- ENVI, Version 4.2. Boulder, CO: ITT Visual Information Solutions.
- ENVI. 2007. *ENVI tutorial: Decision tree classification*. https://www.itvis.com/portals/0/tutorials/envi/06_Decision_Tree_Class.zip (last accessed 30 August 2007).
- Erol, H., and F. Akdeniz. 1996. A multispectral classification algorithm for classifying parcels in an agricultural region. *International Journal of Remote Sensing* 17 (17): 3357–71.
- . 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 (6): 1229–44.
- Footy, G. M. 1992. On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering and Remote Sensing* 58 (10): 1459–60.
- . 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80 (1): 185–201.
- Fraser, C. S., E. Baltsavias, and A. Gruen. 2002. Processing of Ikonos imagery for submetre 3D positioning and building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 56 (3): 177–94.
- Fuller, R. M., G. M. Smith, J. M. Sanderson, R. A. Hill, and A. G. Thomson. 2002. The UK Land Cover Map 2000: Construction of a parcel-based vector map from satellite images. *Cartographic Journal* 39 (1): 15–25.
- Geneletti, D., and B. G. H. Gorte. 2003. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *International Journal of Remote Sensing* 24 (6): 1273–86.
- Gong, P., D. J. Marceau, and P. J. Howarth. 1992. A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment* 40 (2): 137–51.
- Herold, M., X. Liu, and K. C. Clarke. 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering and Remote Sensing* 69 (9): 991–1001.
- Hill, R. A., G. M. Smith, R. M. Fuller, and N. Veitch. 2002. Landscape modelling using integrated airborne multi-spectral and laser scanning data. *International Journal of Remote Sensing* 23 (11): 2327–34.
- Hodgson, M. E., J. R. Jensen, J. A. Tullis, K. D. Riordan, and C. M. Archer. 2003. Synergistic use of lidar and color aerial photography for mapping urban parcel imperviousness. *Photogrammetric Engineering and Remote Sensing* 69 (9): 973–80.
- Hudson, W. D., and C. W. Ramm. 1987. Correct formulation of the kappa coefficient of agreement. *Photogrammetric Engineering and Remote Sensing* 53 (4): 421–22.
- Janssen, L. L. F., and M. Molenaar. 1995. Terrain objects, their dynamics and their monitoring by the integration of

- GIS and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing* 33 (3): 749–58.
- Jensen, J. R. 2005. *Introductory digital image processing: A remote sensing perspective*. Upper Saddle River, NJ: Prentice Hall.
- Jensen, J. R., F. Qiu, and K. Patterson. 2001. A neural network image interpretation system to extract rural and urban land use and land cover information from remote sensor data. *Geocarto International* 16 (1): 19–28.
- Kalkhan, M. A., R. M. Reich, and R. L. Czaplewski. 1995. Statistical properties of five indices in assessing the accuracy of remotely sensed data using simple random sampling. *Proceedings of the ACSM/ASPRS Annual Convention and Exposition 2*:246–57.
- Karathanassi, V., C. Iossifidis, and D. Rokos. 2000. A texture-based classification method for classifying built areas according to their density. *International Journal of Remote Sensing* 21 (9): 1807–23.
- Kim, T. J., S. R. Park, M. G. Kim, S. Jeong, and K. O. Kim. 2004. Tracking road centerlines from high resolution remote sensing images by least squares correlation matching. *Photogrammetric Engineering and Remote Sensing* 70 (12): 1417–22.
- Koukoulas, S., and G. A. Blackburn. 2001. Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments. *Photogrammetric Engineering and Remote Sensing* 67 (4): 499–510.
- Lark, R. M. 1995. Components of accuracy of maps with special reference to discriminant analysis on remote sensor data. *International Journal of Remote Sensing* 16 (8): 1461–80.
- Liu, X., and M. Herold. 2008. Of patterns and processes: Spatial metrics and geo-statistics in urban analysis. In *Integration of GIS and remote sensing*, ed. V. Mesev, 93–115. New York: Wiley.
- Lloyd, C. D., S. Berberoglu, P. J. Curran, and P. M. Atkinson. 2004. A comparison of texture measures for the per-field classification of Mediterranean land cover. *International Journal of Remote Sensing* 25 (19): 3943–65.
- Lobo, A., O. Chic, and A. Casterad. 1996. Classification of Mediterranean crops with multisensor data: Per-pixel versus per-object statistics and image segmentation. *International Journal of Remote Sensing* 17 (12): 2385–400.
- Loh, W. 2005. *QUEST classification tree (version 1.9.2)*. <http://www.stat.wisc.edu/~loh/quest.html> (last accessed 30 August 2007).
- Loh, W., and Y. Shih. 1997. Split selection methods for classification trees. *Statistica Sinica* 7 (4): 815–40.
- Luo, Y., and M. L. Gavrilova. 2006. 3D building reconstruction from LIDAR data. *Lecture Notes in Computer Science* 3980:431–39.
- Ma, Z., and R. L. Redmond. 1995. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing* 61 (4): 435–39.
- Mccauley, S., and S. Goetz. 2004. Mapping residential density patterns using multi-temporal Landsat data and a decision-tree classifier. *International Journal of Remote Sensing* 25 (6): 1077–94.
- NeuroDimension. 2007. *What is a neural network?* <http://www.neurosolutions.com/products/ns/whatisNN.html> (last accessed 30 August 2007).
- PCI Geomatics. 2007. *Signature separability*. <http://www.pci-geomatics.com/cgi-bin/pci/CLWORKS/Signature+Separability> (last accessed 30 August 2007).
- Pedley, M., and P. Curran. 1991. Per-field classification: An example using SPOT HRV imagery. *International Journal of Remote Sensing* 12 (11): 2181–92.
- Platt, R. V., and L. Rapoza. 2008. An evaluation of an object-oriented paradigm for land use/land cover classification. *The Professional Geographer* 60 (1): 87–100.
- Pontius, R. G., Jr. 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing* 66 (8): 1011–16.
- Pontius, R. G., Jr., D. Huffaker, and K. Denman. 2004. Useful techniques of validation for spatially explicit land-change models. *Ecological Modeling* 179 (4): 445–61.
- Pontius, R. G., Jr., R. Walker, R. Yao-Kumah, E. Arima, S. Aldrich, M. Caldas, and D. Vergara. 2007. Accuracy assessment for a simulation model of Amazonian deforestation. *Annals of the Association of American Geographers* 97 (4): 677–95.
- Rottensteiner, F. 2003. Automatic generation of high-quality building models from Lidar data. *IEEE Computer Graphics and Applications* 23 (6): 42–50.
- Schalkoff, R. J. 1989. *Digital image processing and computer vision*. New York: Wiley.
- Seto, K. C., and W. Liu. 2003. Comparing ARTMAP neural network with the maximum-likelihood classifier for detecting urban change. *Photogrammetric Engineering and Remote Sensing* 69 (9): 981–90.
- Smith, G. M., and R. M. Fuller. 2001. An integrated approach to land cover classification: An example in the Island of Jersey. *International Journal of Remote Sensing* 22 (16): 3123–42.
- Smits, P. C., S. G. Dellepiane, and R. A. Schowengerdt. 1999. Quality assessment of image classification algorithms for land-cover mapping: A review and proposal for a cost-based approach. *International Journal of Remote Sensing* 20 (8): 1461–86.
- Stehman, S. V. 1996. Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing* 62 (4): 401–07.
- . 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment* 62 (1): 77–89.
- Stehman, S. V., and R. L. Czaplewski. 1998. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment* 64 (3): 331–44.
- Texas State Historical Association. 2006. *Handbook of Texas online: Austin, TX*. <http://www.tshaonline.org/handbook/online/articles/AA/hda3.html/> (last accessed 30 August 2007).
- Unsalan, C., and K. L. Boyer. 2005. Classifying land development in high-resolution satellite imagery using hybrid structural-multispectral features. *IEEE Transactions on Geoscience and Remote Sensing* 42 (12): 2840–50.
- U.S. Census Bureau. 2000. *Redistricting Census 2000 TIGER/Line@files technical documentation*. http://www.census.gov/geo/www/tiger/rd_2ktiger/tgrrd2k.pdf (last accessed 30 August 2007).

- Wu, S., J. L. Silván-Cárdenas, and L. Wang. 2007. Per-field urban land use classification based on tax parcel boundaries. *International Journal of Remote Sensing* 28 (12): 2777–800.
- Wu, S., B. Xu, and L. Wang. 2006. Urban land use classification using variogram-based analysis with aerial photographs. *Photogrammetric Engineering and Remote Sensing* 72 (7): 813–22.
- Zhan, Q., M. Molenaar, and B. Gorte. 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 (77/4): 1751–59.
- Zhang, Q., and J. Wang. 2003. A rule-based urban land use inferring method for fine-resolution multispectral imagery. *Canadian Journal of Remote Sensing* 29 (1): 1–13.
- Zhang, Q., J. Wang, P. Gong, and P. Shi. 2003. Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. *International Journal of Remote Sensing* 24 (21): 4137–60.

Correspondence: Department of Geography, Geology, and Planning, Missouri State University, 901 S. National Avenue, Springfield, MO 65897, e-mail: swu@missouristate.edu (Wu); qiu@missouristate.edu (Qiu); Center of Excellence for Geospatial Information Science, U.S. Geological Survey, 1400 Independence Road, Rolla, MO 65401, e-mail: usery@usgs.gov (Usery); Department of Geography, The State University of New York at Buffalo, 105 Wilkeson Quad, Buffalo, NY 14261, e-mail: lewang@buffalo.edu (Wang).