RES 105611

INT. J. REMOTE SENSING, 2004, VOL. 25, NO. 00, 1–16

Taylor & Francis
Taylor & Francis Group

Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case

D. CHEN*†, D. A. STOW† and P. GONG‡

†Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA
‡Department of Environmental Science, Policy & Management, University of California at Berkeley, Berkeley, CA 94720, USA

(*Received 7 August 2002; in final form 25 June 2003*)

Abstract. The purpose of this paper is to evaluate spatial resolution effects on image classification. Classification maps were generated with a maximum likelihood (ML) classifier applied to three multi-spectral bands and variance texture images. A total of eight urban land use/cover classes were obtained at six spatial resolution levels based on a series of aggregated Colour Infrared Digital Orthophoto Quarter Quadrangle (DOQQ) subsets in urban and rural fringe areas of the San Diego metropolitan area. The classification results were compared using overall and individual classification accuracies.

Classification accuracies were shown to be influenced by image spatial resolution, window size used in texture extraction and differences in spatial structure within and between categories. The more heterogeneous are the land use/cover units and the more fragmented are the landscapes, the finer the resolution required. Texture was more effective for improving the classification accuracy of land use classes at finer resolution levels. For spectrally homogeneous classes, a small window is preferable. But for spectrally heterogeneous classes, a large window size is required.

1. Introduction

One of the fundamental characteristics of a remotely sensed image is its spatial resolution, or the characteristic size on the ground associated with the radiance measurement of a pixel. For an optical remote sensing system, the spatial resolution is the smallest distinguishable spatial unit (ground resolution element) recorded in an image (e.g. 30 m per pixel for Thematic Mapper (TM) imagery or 10 m per pixel for SPOT panchromatic imagery). Often the ground sampling distance (pixel size) in an image after image re-sampling is used to represent the spatial resolution, but it can be different from the spatial resolution of the sensor that records the image. For this paper, spatial resolution is used in both its strict and its broad meaning of ground sampling distance of an image.

The basic information contained in a remotely sensed image is strongly dependent on spatial resolution (Woodcock and Strahler 1987). Spatial resolution

*Present address: Department of Geography, Queen's University, Kingston, Ontario, K7L 3N6, Canada; e-mail: chendm@post.queensu.ca

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160310001618464

reveals varying patterns and distributions of objects. Improper choice of different spatial resolution can lead to misleading interpretation, e.g. in a Landsat Multi-Spectral Scanner image, the urban residential environment is sensed as a relatively homogeneous entity. However, when observed at finer resolution, the residential area is mostly made of individual houses, roads and plants. Furthermore, the spatial resolution of an image may not correspond to any particular ground feature within a scene. With the development of remote sensing techniques, the spatial resolution of various satellite sensors now ranges from 0.5 m to 25 000 m. Furthermore, high resolution airborne data acquisition technology has developed rapidly in recent years. As an increasing number of higher resolution datasets become available, the selection of the appropriate spatial resolution becomes more complex. In other words, the factor of spatial resolution plays an increasingly important role in the employment of remotely sensed imagery (Quattrochi and Goodchild 1997, Cihlar 2000).

Appropriate spatial resolution is a function of the type of environment, the kind of information desired and the techniques used to extract information. The scale (or spatial resolution) of available image data, analysis methods, environmental situations and primary questions about these environments should all be considered when determining the appropriate spatial resolution for analysis. Many researchers have explored potential problems in developing stable parameters from remotely sensed imagery at different spatial resolutions due to autocorrelation and spatial dependency (Campbell 1981, Cushnie 1987, Johnson and Howarth 1987, Gong and Howarth 1990a, Lam and Quattrochi 1992, Marceau *et al*. 1994a, b, Arbia *et al*. 1996, Hlavka and Livingston 1997, Hodgson 1998, Bian and Butler 1999, Chen 1999, Treiz and Howarth 2000, Chen and Stow 2002). 1

The spatial resolution of an image substantially affects every stage of image classification. Markham and Townshend (1981) found that image classification accuracy is affected by two factors. The first factor is the influence of boundary pixels on classification results. The second factor which influences classification accuracy is that finer spatial resolution increases the spectral-radiometric variation of land cover types. Based on a forest environment, Marceau *et al*. (1994a, b) evaluated the impact of measurement scale and spatial aggregation on the information context and classification accuracy. Their results revealed that the impact of resolution change is greater than the change of aggregation level. It is commonly agreed that at a particular classification level, some classes are better classified at fine spatial resolution while others require coarser resolution due to the complexity of environments. Hence, the need exists for a clear assessment of the distorting effects of spatial resolution on landscapes of varying degrees of heterogeneity.

For high spatial resolution images, land use/cover classes tend to be represented by spatial units of heterogeneous spectral reflectance characteristics. Previous studies have shown that a decrease in land use/cover classification accuracy is likely to occur as the spatial resolution of the data is improved, but other sensor characteristics are kept unchanged (Townshend and Justice 1981, Toll 1984, 1985, Latty *et al*. 1985, Martin *et al*. 1988, Gong and Howarth 1990b, Marceau *et al*. 1994a, Treitz and Howarth 2000). This is especially true for urban environments, due to the particularly heterogeneous nature of most urban land use/cover types (Kontoes *et al*. 2000). Most of the previous studies (as cited above) report results based on simulated images or on studies of natural environments such as forest and agriculture, using images with spatial resolutions coarser than 10 m. Relatively few cases are reported on the spatial resolution effect on urban land use/cover classification, especially at finer resolution levels (< 10 m).

The objective of this paper is to determine how statistical properties of training site data and classification accuracy change with varying spatial resolution. The effect of different window sizes used to extract texture features is also examined at different resolutions. One metre resolution USGS Colour Infrared Digital Orthophoto Quarter Quadrangle (DOQQ) data were aggregated to generate a series of coarser image resolutions of 4 m, 8 m, 12 m, 16 m, 20 m and 24 m. Each image is classified into eight land use/cover types using a supervised maximum likelihood classification method. The study site, classification scheme and aggregation method are described in the following section. The classification accuracy at each resolution is reported.

2. Research design

2.1. *Dataset*

The urban and rural fringe areas of the Del Mar quadrangle within San Diego County, California were selected as the study area. This area is one of the primary locations of urban expansion in San Diego County in recent years, with large tracts of undeveloped land being converted into built land uses. These land use dynamics make it difficult for planners and land managers to obtain or maintain up-to-date land cover/use information (Coulter *et al.* 1999). Also, there is a wide variety of ancillary data available for this study area including land use, land use plan, and parcel boundary data layers.

USGS CIR Digital DOQQ data acquired in September, 1996 are the major image data source for this study. A mosaic of CIR DOQQ data for San Diego County had been radiometrically balanced and georeferenced to the State Plane coordinate system. The image data have a spatial resolution of 1 m with three spectral bands (green, red and near-infrared (NIR)).

Two image subsets, one encompassing the Sorrento Valley and Mesa industrial park and the other covering a predominantly suburban residential area immediately east of the city of Del Mar, were selected to examine the effectiveness of the proposed methods. Figures 1 and 2 show two NIR band subsets at 4 m resolution. Each subset contains several major land use/cover types so that a variety of land use/cover types are included.

Figures 1 and 2 illustrate that the spatial structure of these two study sites is exhibited in a variety of forms. The Sorrento Valley study site is dominated by single-residential, industrial/commercial, and vacant/or cleared areas, while the Del Mar study site is covered mainly by single- and multi-residential and undeveloped areas. Compared to the Sorrento Valley area, the landscape of the Del Mar site is more fragmented with a greater number of land use classes mixed together. The land use/cover classes in these two study sites cover most land use/cover categories of San Diego urban–suburban areas.

2.2. *Aggregation method*

To evaluate the effect of spatial resolution, the original 1 m resolution images were aggregated progressively into six different levels (4 m, 8 m, 12 m, 16 m, 20 m and 24 m) of squared blocks of variable size. Edge effects were reduced by considering a mirror effect in the bordering pixels of the image. Since the nearest neighbour and cubic convolution algorithms of interpolating data induce

Remote Sensing *(gamma)* RES33912.3d 19/12/03 15:36:17 Rev 7.51n/W (Jan 20 2003)
The Charlesworth Group, Huddersfield 01484 517077

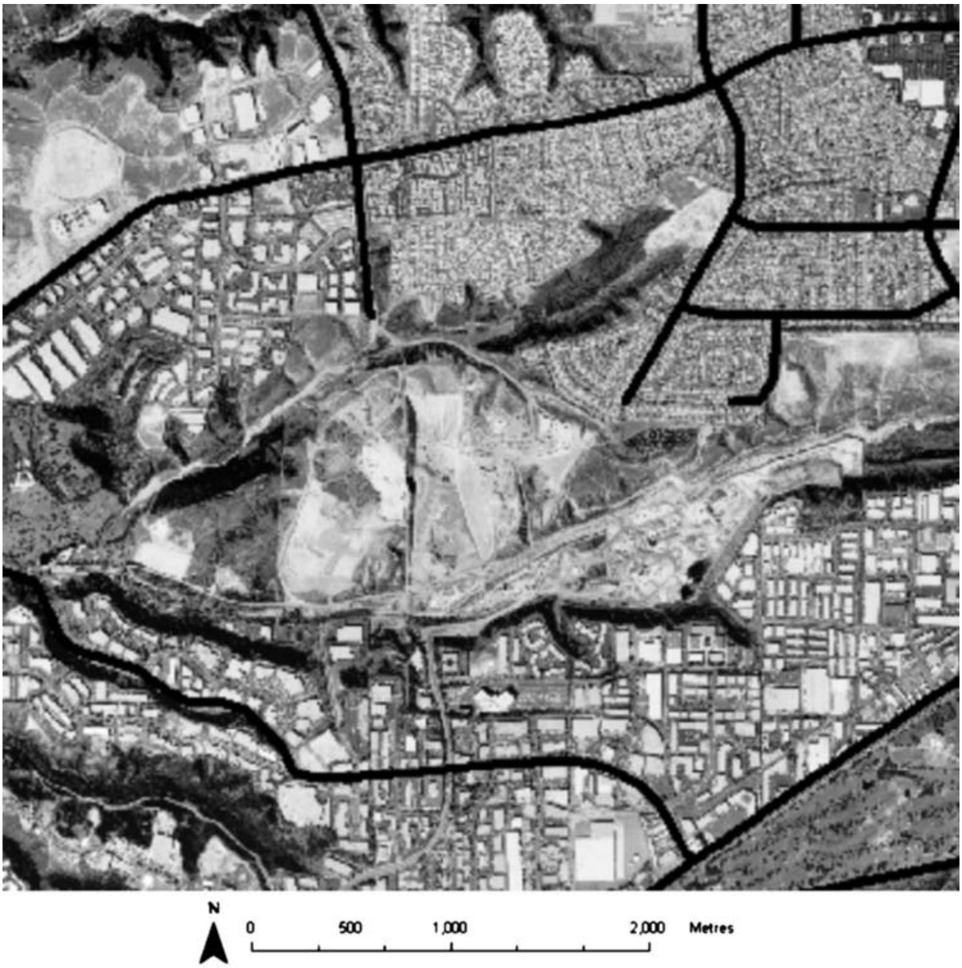


Figure 1. The near-infrared (NIR) band of a subset of a USGS colour infrared DOQQ for an area of Sorrento Valley, CA. The black lines mask out the major roads.

sharpening or smoothing effects that influence the resulting analysis (Marceau *et al.* 1994a, Bian and Butler 1999), a methodology similar to that of Woodcock and Strahler (1987) was used, by applying a $n \times n$ pixel averaging window across each image. This averaging routine was regarded as an efficient and simple way to represent the physical aggregation process of a sensor's IFOV by Marceau *et al.* (1994b). 2

2.3. *Classification scheme*

The land use/cover classification scheme used in this study is listed in table 1. Similar to the USGS scheme (Anderson *et al.* 1976), the classification scheme used in this research should be considered a land use/cover scheme. It was developed by modifying a classification scheme developed by the San Diego Association of Governments (SANDAG), a regional planning agency, and then emphasizing image pattern and spatial variability rather than economic and functional differences, in order for the remote sensing data to be most effectively used.

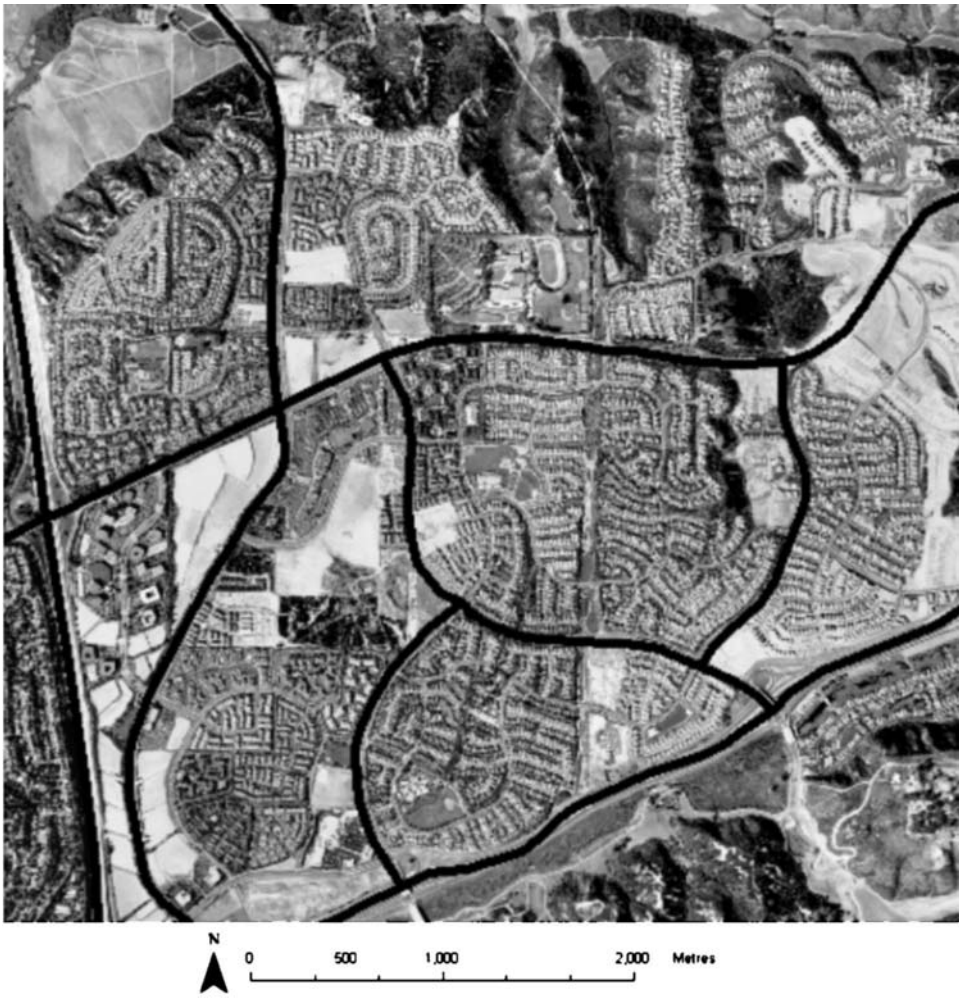


Figure 2. The near-infrared (NIR) band of a subset of a USGS colour infrared DOQQ for an area east of Del Mar, CA. The black lines mask out the major roads.

Transportation systems are linear features and may require different algorithms for identification (see Wang *et al*. 1992, Lakshmana 1996). Thus, all freeways, highways and major roads were masked out by applying a 40 m buffer zone using the SANDAG coverage of major roads (see http://www.sandag.cog.ca.us/ris/gis/basemap.html for definition). Minor roads are considered components of land use parcels and are not classified separately.

2.4. *Spatial feature extraction*

A variance texture measure was computed for each NIR spectral band at different resolution levels. The success of the texture method in land use/cover classification depends largely on the appropriate pixel window size being selected for spatial feature generation. Seven different window sizes (3×3, 5×5, 7×7, 9×9, 11×11, 13×13, 15×15) were tested. Transformed divergence (TD) (Swain and Davis 1978) was used to compute texture signature separability between class pairs.

Table 1. Land use classification scheme, definitions and the number of training pixels used in this study.

Land use/cover class	Characteristics	The number of training pixels at 4 m
Single-family residential area	Single-family detached housing units, on lots smaller than 1 acre (4047 m^2).	335
Multi-family residential area	Attached housing units, two or more units per structure – includes duplexes, townhouses, condominiums apartments.	371
Industrial/Commercial area	All industrial, offices, wholesale and shopping centres. Structures are usually large and cover the majority of the parcel with little vegetation.	680
Artificial (irrigated) grassland	Includes public or private golf courses, parks dominated by three types of land covers: well-maintained grass, normal grass and small brushes. The spectral reflectance of the well-maintained grass is very high in the infrared band.	256
High-density vegetation	Parks and natural vegetation areas with dense trees and bushes	238
Cleared for construction	Lands on which construction will be undertaken or is underway.	668
Undeveloped land	Includes natural shrub land, wetland and riparian vegetation.	306
Agriculture land	Fields for agriculture uses.	185

2.5. *Classification method*

The Gaussian maximum likelihood (GML) classifier was used in this study for three reasons. First, it is relatively convenient to implement. Secondly, the maximum likelihood decision rule is by far the most common supervised classification method and is widely used. Finally, the GML is robust and utilizes means, variances and covariances of training site statistics, where most other decision rules are based on simpler statistics.

Supervised training was adopted in this study. Groups of contiguous pixels were selected as training samples in the class signatures. The number of training samples for each class in the finest scale is shown in table 1. The detailed explanation and examples of each class can be found in Chen and Stow (2002). These training samples were distributed around the whole area as evenly as possible. Selection of the training samples was aided by field reconnaissance. The same set of training samples were used at all spatial resolutions to examine differences in signature statistics such as mean and standard deviation with variable spatial resolution.

Accuracy was assessed through a comparison of test pixels. Reference data were generated by manually interpreting aerial photographs and by ground-level observations. In order to use one set of reference data at several resolution levels, boundary or mixed pixels which could not be clearly identified as the same classes at both the 4 m and 8 m resolutions were deleted from the sample set. A total of 865 samples were identified for the Sorrento Valley study area, while 1002 samples were utilized for the Del Mar study site.

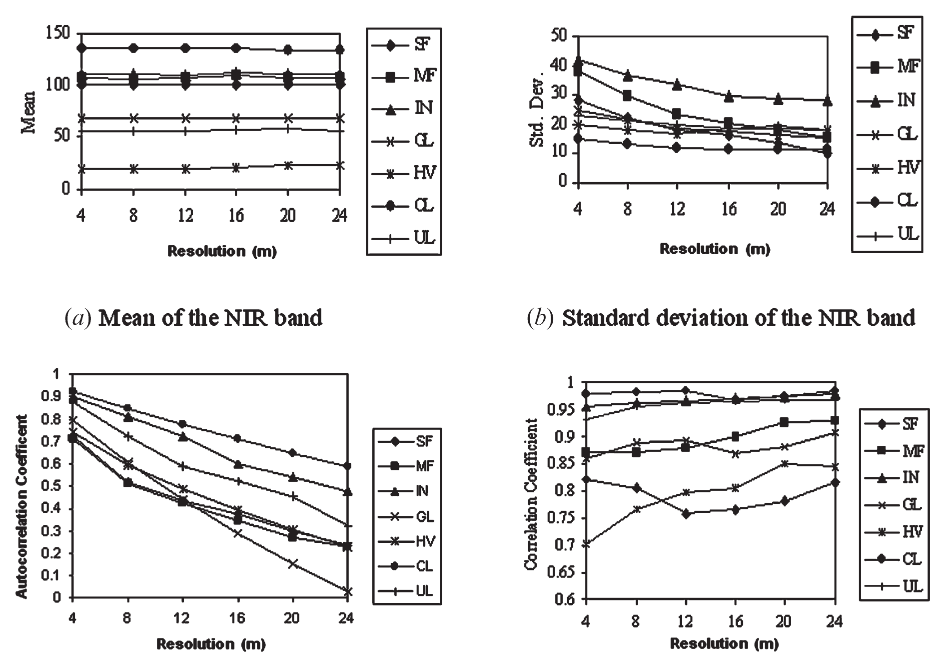
For each class at each spatial resolution, the overall and individual Kappa coefficient is calculated for each confusion matrix to evaluate the agreement between the classification results and the reference data. The Kappa coefficient takes all the elements in the confusion matrix into consideration, rather than just the diagonal elements as occurs with the calculation of the overall classification accuracy, and has been recommended as a suitable accuracy measure in thematic classification for representing the whole confusion matrix (Cohen 1965, Stehman 1992).

3. Results and discussion

3.1. *Resolution effects on the basic statistical measure of training data*

Figures 3(*a*)–(*c*) summarize the major results concerning spatial resolution effects on some basic statistical measures of training data. Figures 3(*a*) and (*b*) show the means and standard deviations of image digital number (DN) for each class at different resolutions. As expected, the means are relatively stable at each resolution when the averaging aggregation method is used. The standard deviations are damped by aggregation.

The autocorrelation coefficient shown in figure 3(*c*) measures the relationship between the difference of adjacent pixels. Adjacent pixels that have similar values are highly correlated spatially, while pixels with very different values are not autocorrelated spatially. The spatial autocorrelation between adjacent pixels is decreased for all classes as resolution becomes coarser. However, the rate of decrease of the autocorrelation coefficient is influenced by the spatial structure of material



(*a*) **Mean of the NIR band**

(*b*) **Standard deviation of the NIR band**

(*c*) **Autocorrelation coefficient of adjacent pixels in the NIR band**

(*d*) **Correlation coefficient between the NIR and the green band**

Figure 3. (*a*) Mean, (*b*) standard deviation, (*c*) autocorrelation, (*d*) correlation coefficient of training data against resolution for each class (SF, Single-family; MF, Multi-family; IN, Industry; GL, Irrigated grass land; HV, High-density vegetation; CL, Cleared land; UL, Undeveloped land). 4

Remote Sensing *(gamma)* RES33912.3d 19/12/03 15:36:28 Rev 7.51n/W (Jan 20 2003)
The Charlesworth Group, Huddersfield 01484 517077

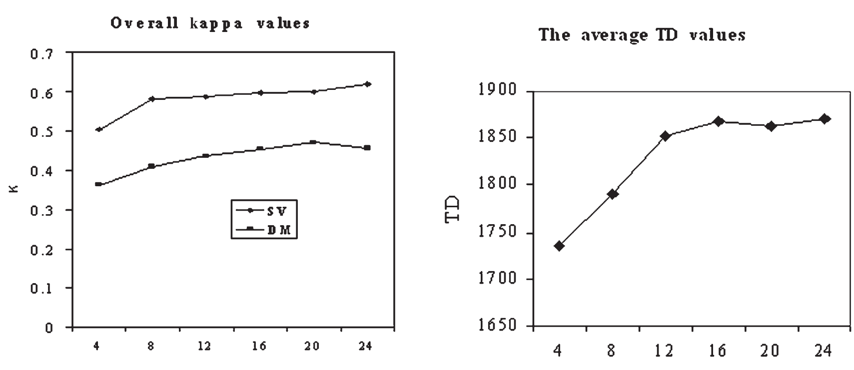
components associated with each class. The cleared and industrial land use types are composed of relatively homogeneous surfaces. Therefore, they have higher autocorrelation coefficients at all spatial resolutions compared to other classes. The rapid decrease in the case of irrigated grassland is due to the small vegetation patches within this class. The heterogeneous characteristics within single-family and multi-family residential lands contribute to their relatively lower autocorrelation coefficients. These results parallel the empirical and theoretical findings in the literature (see e.g. Arbia *et al*. 1996).

Figure 3(*d*) shows how resolution effects the correlation between the NIR and green bands. As the degree of similarity between pixels is induced with resolution becoming coarser, the correlation coefficient between two bands is supposed to increase. However, the situation considered here is a little different. The training pixel values at coarser resolution are influenced not only directly by the values of the same training pixels of finer resolution, but also by the values of their surrounding pixels during aggregation. As a consequence, the curves in figure 3(*d*) are not monotonically increasing as the resolution level increases.

3.2. *Resolution effects on separability and overall classification accuracy*

Figure 4 shows the overall Kappa and average TD values for classification products derived at five different resolutions, when only spectral bands are used for the two study areas. TD values obtained at all resolutions are below 1900, suggesting at least some signatures overlap when only spectral bands are considered. (The TD statistic saturates at a value of 2000, meaning complete separability.) Spatially aggregating the image to coarser spatial resolution resulted in an increase of TD when the initial resolution is finer than 16 m. The overall accuracy and Kappa values are similar for the two study areas. Going from 4 m to 20 m, the classification accuracy gradually improves. However, this improvement is greater at finer resolutions. When the spatial resolution is coarser than 12 m, differences in classification accuracy are less than 2% for both study areas.

At each resolution the classification accuracy in the Del Mar study area is always lower than that in the Sorrento Valley area. The highest accuracy is obtained for the 24 m resolution image of the Sorrento Valley study area and 20 m



(*a*) (*b*)

Figure 4. (*a*) Overall Kappa and (*b*) average transformed divergence (TD) values obtained at each resolution when only spectral bands are used. SV and DM represent Sorrento Valley and Del Mar study areas, respectively. 5

Del Mar set. This may be explained by the landscape difference between the two study areas. The different urban land use/cover classes in the Del Mar are intermixed, while the Sorrento Valley area contains more homogeneous land use parcels.

3.3. *Texture window size and resolution effect on overall classification accuracy*

For each resolution and all bands, different window sizes were used to calculate the local variance as a texture measure. The average TD value and overall Kappa value were obtained at each window size. Since variance features generated from different bands yield similar results in classification accuracy, only the NIR band variance was used.

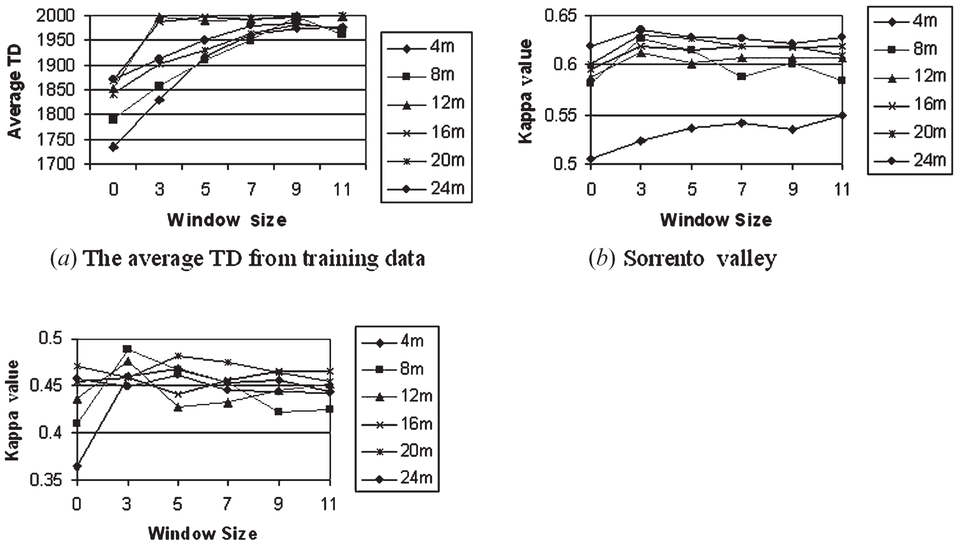
Table 2 presents the TD values and classification accuracy statistics when combined variance measures of texture generated from different window sizes were used for the 8 m resolution images. From table 2 it is apparent that combining more than one texture image generated from different window sizes did not result in significant improvement of classification accuracy compared with that of using only one texture image. Therefore, in the following sections only the results obtained using NIR band variance at a single window size are reported.

To examine the influence of texture window size on classification accuracy at different resolutions, the average TD values and overall Kappa values are plotted versus texture window size at each resolution for the two study areas in figure 5. Each curve in figure 5 starts with the average TD values or overall Kappa value obtained by using only spectral bands, which serves as a reference for comparison and is plotted as the label '0'.

Comparing the TD and classification accuracy values obtained by using only spectral bands with those when texture features are included, it can be seen that inclusion of texture substantially improves both TD values and classification accuracy. Figure 5(*a*) shows that the average separability among curves increases and reaches close to their maximum 2000 as the window size increases from 3×3 to 9×9 for all resolutions. However, when the window size reaches 11×11, the average TD value decreases for 8 m and 24 m resolution images. The average separability among these six curves can be grouped in two similar patterns. At 12 m and 16 m resolutions, the average separability among curves increases rapidly to almost 2000 (saturation value) for a 3×3 texture window, and then maintains high

Table 2. Transformed Divergence (TD) values and classification accuracy when combining variance features generated at different texture window sizes using an 8 m NIR image.

Window size	Average TD	Minimum TD	Overall Kappa	Overall accuracy (%)
None	1790	284	0.5522	64.01
3×3	1857	493	0.6272	69.13
5×5	1911	724	0.6147	67.98
7×7	1952	1346	0.5881	65.55
9×9	1997	1934	0.6012	66.71
11×11	1963	1484	0.5842	65.2
$(3\times3)+(5\times5)$	1906	889	0.6056	67.17
$(3\times3)+(7\times7)$	1959	1416	0.6083	67.4
$(3\times3)+(9\times9)$	1973	1433	0.6109	67.63
$(3\times3)+(11\times11)$	1969	1438	0.6245	68.79
$(5\times5)+(11\times11)$	1976	1557	0.5976	66.47
$(5\times5)+(9\times9)$	1978	1513	0.5924	66.01



(*a*) The average TD from training data

(*b*) Sorrento valley

(*c*) Del Mar

Figure 5. (*a*) Average transformed divergence (TD) and (*b*), (*c*) overall Kappa values obtained at different resolutions plotted against texture window size for two study areas.

separability at larger window sizes. At other resolutions, the average TD values increase gradually as the window size becomes larger.

As observed in previous aspects of this study, there is not an exact correspondence between peaks in the average TD curves and the greatest Kappa value curves by comparing figures 5(*a*) and (*b*). Figures 5(*b*) and (*c*) show that for most of the window sizes, classification accuracies are improved when compared with the results without texture images. The overall Kappa value at each resolution reached its maximum at different window sizes.

For the Sorrento Valley study area, the highest overall Kappa values were obtained at texture window sizes of 3×3 for all resolutions except 4 m, in which case the classification accuracy increased as window size increased from 3×3 to 7×7, and then had a small dip at a window size of 9×9. This indicates that at finer spatial resolutions the optimum window size used for extracting texture features should be relatively larger. In fact, at 16 m, 20 m, and 24 m resolution levels, the Kappa values obtained at different window sizes were similar to those obtained when only spectral bands were used. The significance test shows that classification accuracy improvements were significant at the 95% confidence level at window sizes of 7×7 and 11×11 for 4 m, 3×3 and 5×5 for 8 m, 3×3 for 12 m. This means that adding variance information for bands did not significantly improve classification results at relatively coarser resolutions such as 16 m, 20 m and 24 m. Although the window sizes tested in this research did not exceed 11×11, the use of even larger window sizes is not likely to result in significantly improved classification accuracy, considering the boundary effect of larger window sizes. Other studies also suggest that small window sizes yield higher classification accuracies (Gong and Howarth 1990b, 1992), though this varies from category to category.

Similar trends were found for the Del Mar study area. Adding texture improved the classification accuracy more at finer resolutions than at coarser resolutions. The Kappa values obtained with texture features at 16 m, 20 m, and 24 m were close to those without texture. The largest improvements with texture images were 0.011, 0.011 and 0.005 at 16 m, 20 m and 24 m, respectively. At 4 m the classification accuracy improvement was significant at the 95% confidence level for all window sizes. The highest overall Kappa value, 0.468, was achieved with a window size of 5×5. This leads to an increase in Kappa value of 0.106 as compared with the Kappa value of 0.363 obtained without texture. Three other significant improvements at the 95% confidence level occurred for a window size of 3×3 for 12 m and window sizes of 3×3 and 5×5 for 8 m.

Comparing figures 5(*b*) and (*c*) one can see some noticeably different results in accuracy, which reflect the landscape difference between the two study areas. First, the classification accuracy improvement when including texture was much higher for the Del Mar than for the Sorrento Valley study area at 4 m, 8 m and 12 m, but lower at 16 m, 20 m and 24 m. Secondly, the curves for different resolutions in the Del Mar area show more complicated trends than those in Sorrento Valley. For example, in figure 5(*b*) the overall Kappa value is consistently lowest at 4 m resolution when compared with those obtained with the same window size at other resolutions. But, as seen in figure 5(*c*), this is not the case except when no texture information in bands was used. Thirdly, at 4 m resolution the window size at which the highest overall Kappa value was achieved is 5×5 in Del Mar, which is smaller than that in the Sorrento Valley area. When no texture was used, the highest classification accuracy was achieved for 20 m in Del Mar and 24 m in Sorrento Valley. At a window size of 3×3, the highest Kappa value occurred for 8 m in Del Mar and 16 m in Sorrento Valley.

All of the above suggests that there is no single optimum window size or resolution level at which the best result will be achieved for all areas. The optimum resolution or window size used for extracting texture features strongly depends on the landscape pattern in the study area when the classification system is fixed. The more fragmented and mixed the landscape, the smaller the window sizes and the finer the resolution that should be chosen.

3.4. *Window size effects on individual category accuracy at different resolutions*

As a measure of classification accuracy for individual classes, conditional Kappa values of each class were obtained for different resolution images. The curves in figure 6 present the conditional Kappa values plotted against window size. The figure shows that different classes reach their maximum accuracies at different window sizes for different resolutions. Table 3 summarizes the the window size at which maximum Kappa values were obtained for each class at different resolutions. Spatially heterogeneous land-use classes, such as single-family, multi-family, and industry, appear to reach their maxima at relatively larger window sizes, while spatially homogeneous classes with only one dominant land-cover type, such as grass land, cleared land and high-density vegetation, reach their maxima at smaller window sizes.

Several trends can be observed in figure 6. For spectrally heterogeneous land use classes, such as single-family, multi-family and industry, individual Kappa values increased as window sizes increased, at all resolutions. However, this increasing trend is most evident for the single-family class, while it is least for industry. The increasing trend indicates that the window sizes used in this paper may not be big

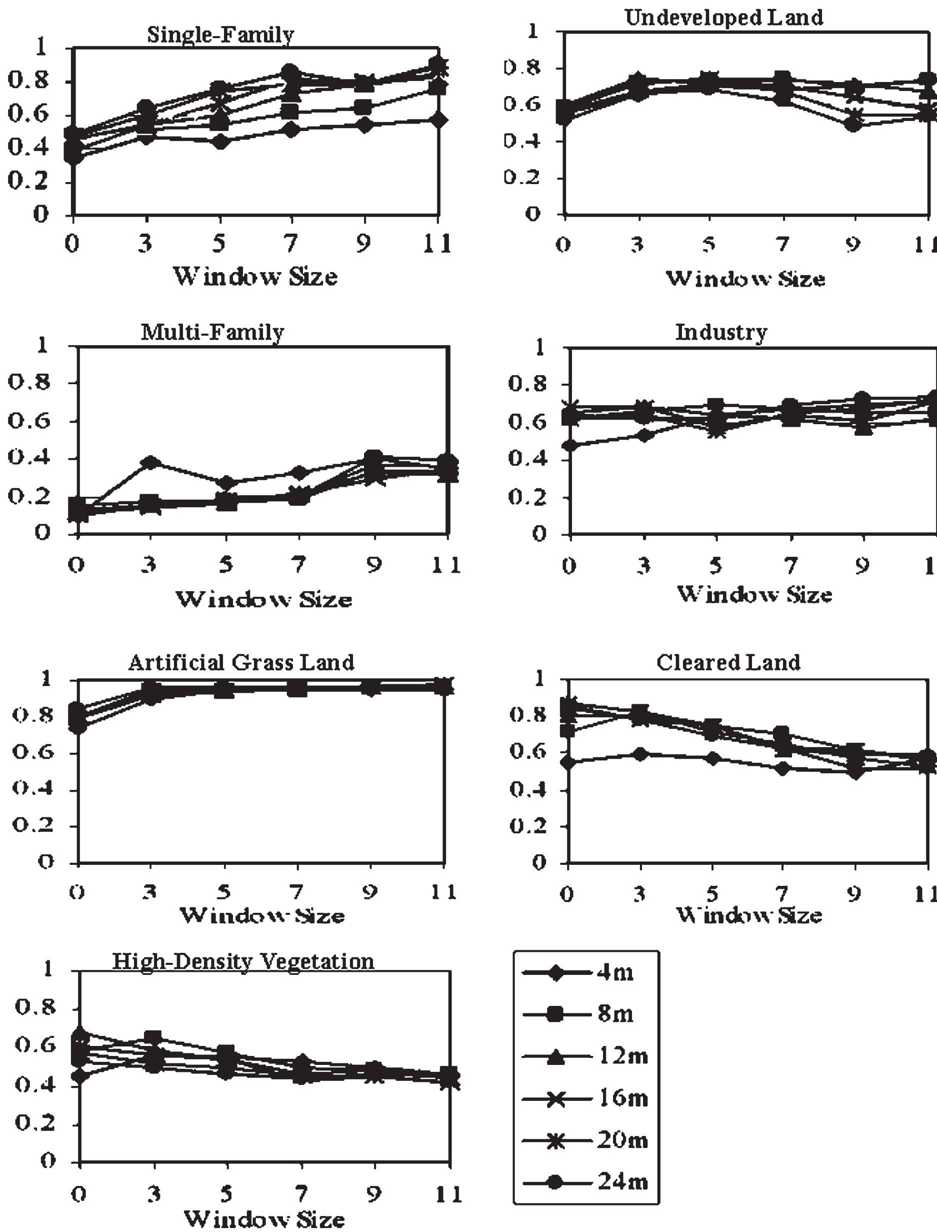


Figure 6. Conditional Kappa values for classification results in the Sorrento Valley study area against window size. Window size '0' indicates classification accuracy obtained without texture.

enough to obtain the highest individual classification accuracy, especially for industrial land. For irrigated grassland and undeveloped land, Kappa values increase as the window size increases from 0 to 5 and then remain stable at larger sizes. The exception is when the window size exceeds 7 and the individual Kappa values of undeveloped land decrease at resolutions of 16 m, 20 m and 24 m. The relatively stable Kappa values at large window sizes are due to the large integrated

Table 3. Window sizes at which maximum Kappa values were achieved for each class for different spatial resolutions.

Spatial resolution	4 m	8 m	12 m	16 m	20 m	24 m
Class						
Single-family	11	11	11	9	7	9
Multi-family	9	9	11	11	11	11
Industry	7	5	3	11	11	11
Artificial grass land	11	11	11	7	3	9
High-density vegetation	5	3	3	0	0	0
Cleared land	9	7	3	3	3	5
Undeveloped land	5	5	3	3	5	3

areas of these two classes. Adding a texture improved the individual Kappa values of two spectrally pure classes (cleared land and high-density vegetation) when the resolution is coarser than 12 m. The highest individual Kappa values were achieved at 12 m and 24 m for cleared land and high-density vegetation, respectively, when only spectral bands are used. The relatively lower accuracies of spectrally pure classes further supports the view that texture is more useful for improving classification accuracy of heterogeneous land use classes than relatively homogeneous land cover classes in the study area.

3.5. *Sensitivity of classification results at different resolutions*

In order to analyse the difference between classification results obtained at different resolutions, the classified images based only on spectral bands for different resolutions were compared and categorical difference images were generated. Pixels classified to the same class between two classification maps were assigned '0's, otherwise they were '1's. Figure 7 shows two examples of these binary maps. The percentage of pixels that were classified into different classes ('1's) between two

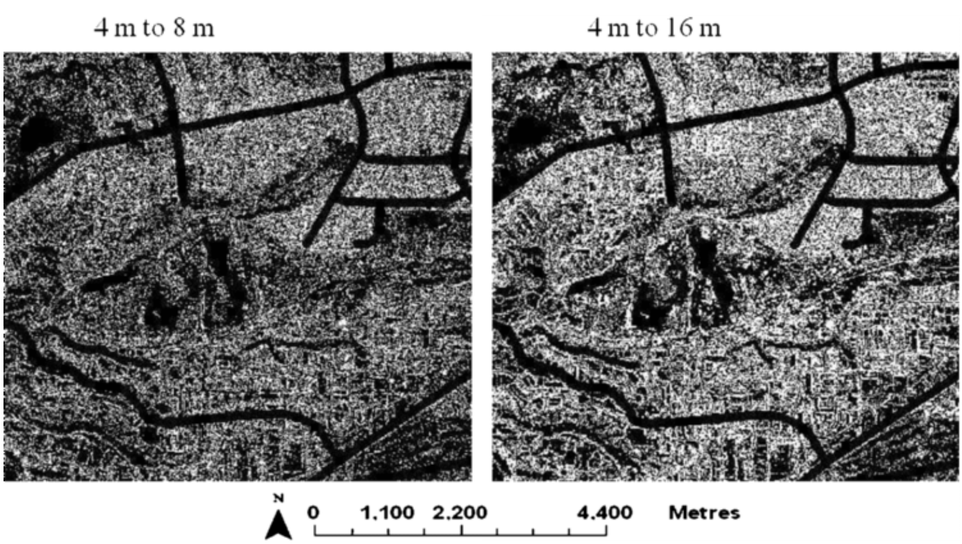


Figure 7. Binary maps showing the agreement between classification results at different resolutions. The white indicates that the pixel was assigned to different categories in the classification of the two different spatial resolution images being compared. 6

compared resolution levels was calculated and listed in the upper-right half (above the diagonal) of table 4. This is the overall disagreement between pairs of image classification output for different resolutions or windows. The differences between the overall accuracy are listed in the lower-left half (below the diagonal) of table 4.

In general, the overall disagreement between classified images at different resolutions is much higher than the difference between their overall accuracies. For example, there is only a 0.81% difference between the overall accuracies of classified images at 8 m and 16 m when only spectral bands were used. However, when the classes of the same pixels between their classified images were compared, as much as 36.6% of pixels were classified into different classes. The range of overall accuracy difference is 0.02–9.13% among the five resolutions, while the percentage of mismatch pixels ranges from 30.91–49.99%. The high percentage of per-pixel disagreement between classified images from different resolutions indicates that the resolution effects on classification results are more than they appear to be on the overall accuracy. In fact, a large portion of the confusion in the classification results at different resolutions is moderated in the overall accuracy.

4. Summary and conclusions

In this paper the resolution effects on training statistics and classification accuracy were reported for a case study of an urban environment using simulated multi-spectral image data. The average TD, and the overall and individual Kappa values obtained at different resolutions were compared both with and without texture features. The difference between the classified images at different resolutions was also discussed.

Some of the major findings from the experimental results are as follows.

1. Resolution effects on basic statistics of training data were found to be similar to those found in the literature. In particular, the spatial autocorrelation coefficient and standard deviation (or variance) of brightness values of training data decreased with increasing spatial resolution, which, in turn, was moderated by a positive pattern of dependency among neighbouring pixels.
2. No single resolution yielded the highest classification accuracy in a study area having different landscape types. When only spectral bands are used in classification, the selection of an appropriate resolution depends on the spatial structures within and between classified classes. The more heterogeneous are the land use/cover classes and the more fragmented are the

Table 4. Percentage disagreement between overall classification accuracy results at different resolutions and the difference between their overall accuracies.

	4 m	8 m	12 m	16 m	20 m	24 m
4 m	0	32.96	41.67	45.26	47.83	49.99
8 m	6.59	0	34.55	36.6	41.23	43.2
12 m	6.71	0.02	0	32.19	36.01	37.96
16 m	7.4	0.81	0.69	0	32.23	35.89
20 m	7.63	1.04	0.72	0.23	0	30.91
24 m	9.13	2.54	2.42	1.73	1.5	0

The upper-right half (above the diagonal) of the table lists the percentage of pixels classified into different classes at different spatial resolutions. The lower-left half (below the diagonal) shows the differences between overall accuracies obtained at different spatial resolutions.

landscapes, the finer the resolution required. The highest overall classification accuracy was achieved at 24 m for the Sorrento Valley and 20 m for the Del Mar study area, when no texture features were used.

3. In most cases, adding texture information as an additional input with spectral bands increased classification accuracy. However, the improvement of classification accuracy appears to be dependent on the resolution level applied. Texture was more effective in improving the classification accuracy of land use classes at finer resolution levels. When the spatial resolution exceeded a certain level, adding texture did not lead to higher classification accuracy.
4. Window size was important in texture extraction. In general, a larger window size was needed at finer resolution than at coarse resolution. For spatially heterogeneous classes, a larger window size appears to be required. On the other hand, for spectrally homogeneous classes, a small window size yielded higher accuracy. The most significant improvement was achieved at 8 m with a window size of 3×3 for Sorrento Valley and 4 m with a window size of 5×5 for Del Mar.
5. Spatial resolution had much more influence on the spatial distribution of classification errors than on the overall classification accuracy. This means that while changing the overall accuracy slightly, images of different resolution may yield dramatically different spatial distributions of classified land use/cover.

References

ANDERSON, J. R., HARDY, E., ROACH, J., and WITMER, R., 1976, A Land Use and Land Cover Classification System for Use with Remote Sensor Data. USGS (United States Geological Survey) Professional Paper #964 (ISSN No. 1044-9612), USGS, Reston, Virginia.

ARBIA, G., BENEDETTI, R., and ESPA, G., 1996, Effects of the MAUP on image classification. *Geographical Systems*, **3**, 123–141.

BIAN, L., and BUTLER, R., 1999, Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric Engineering and Remote Sensing*, **65**(1), 73–84.

CAMPBELL, J. B., 1981, Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing*, **47**(3), 355–363.

CHEN, D., and STOW, D. A., 2002, The effect of training strategies on supervised classification at different spatial resolutions. *Photogrammetric Engineering and Remote Sensing*, **68**(11), 1155–1161.

CHEN, J. M., 1999, Spatial scaling of a remotely sensed surface parameter by contexture. *Remote Sensing of Environment*, **69**, 30–42.

CIHLAR, J., 2000, Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, **21**, 1097–1114.

COHEN, J., 1965, A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1), 37–46.

COULTER, L., STOW, D., KIRACOFE, B., LANGEVIN, C., CHEN, D., DAESCHNER, S., SERVICE, D., and KAISER, J., 1999, Deriving current land use information for metropolitan transportation planning through integration of remotely sensed data and GIS. *Photogrammetric Engineering & Remote Sensing*, **65**(11), 1293–1300.

CUSHNIE, J. L., 1987, The interactive effect of spatial resolution and degree of interval variability within land-cover types on classification accuracies. *International Journal of Remote Sensing*, **8**, 15–29.

GONG, P., and HOWARTH, P. J., 1990a, An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric Engineering and Remote Sensing*, **56**(5), 597–603.

GONG, P., and HOWARTH, P. J., 1990b, The use of structural information for improving land-cover classification accuracies at the rural–urban fringe. *Photogrammetric Engineering and Remote Sensing*, **56**(1), 67–73.

Gong, P., and Howarth, P. J., 1992, Frequency-based contextual classification and gray-level vector reduction for land-use identification. *Photogrammetric Engineering and Remote Sensing*, **58**(4), 423–437.

Hlavka, C. A., and Livingston, G. P., 1997, Statistical models of fragmented land cover and the effects of coarse spatial resolution on the estimation of area with satellite sensor imagery. *International Journal of Remote Sensing*, **18**, 2253–2259.

Hodgson, M. E., 1998, What size window for image classification? A cognitive perspective. *Photogrammetric Engineering and Remote Sensing*, **64**(8), 797–806.

Johnson, D. D., and Howarth, P. J., 1987, The effects of spatial resolution on land cover/use theme extraction from airborne digital data. *Canadian Journal of Remote Sensing* **13**(2), 68–74.

Kontoes, C. C., Raptis, V., Lautner, M., and Oberstadler, R., 2000, The potential of kernel classification techniques for land use mapping in urban areas using 5m-spatial resolution IRS-1C imagery. *International Journal of Remote Sensing*, **21**, 3145–3151.

Lakshmana Rao, K. M., 1996, Regional network planning and the development of rural transportation using remote sensing techniques. *International Journal of Remote Sensing*, **17**, 3453–3466.

Lam, N. S-N., and Quattrochi, D.A., 1992, On the issues of scale, resolution, and fractal analysis in the mapping sciences. *Professional Geographer*, **44**(1), 88–98.

Latty, R. S., Nelson, R., Markham, B., Williams, D., Toll, D., and Irons, J., 1985, Performance comparison between information extraction techniques using variable spatial resolution data. *Photogrammetric Engineering and Remote Sensing*, **51**(9), 1159–1170.

Marceau, D. J., Howarth, P. J., and Gratton, D. J., 1994a, Remote Sensing and the measurement of Geographical Entities in a forest environment 1: The scale and spatial aggregation problem. *Remote Sensing of Environment*, **49**, 93–104.

Marceau, D. J., Howarth, P. J., and Gratton, D. J., 1994b, Remote Sensing and the measurement of Geographical Entities in a forest environment 2, The optimal spatial resolution. *Remote Sensing of Environment*, **49**, 105–117.

Markham, B. L., and Townshend, J. R. G., 1981, Land cover classification accuracy as a function of sensor spatial resolution. *Proceedings of the 15th International Symposium on Remote Sensing of Environment*, pp. 1075–1090. 3

Martin, L. R. G., Howarth, P. J., and Holder, G., 1988, Multispectral classification of land use at the rural–urban fringe using SPOT data. *Canadian Journal of Remote Sensing*, **14**(2), 72–79.

Quattrochi, D. A., and Goodchild, M. F. (eds.), 1997, *Scale in Remote Sensing and GIS* (Boca Raton, FL: CRC/Lewis Publishers).

Stehman, S., 1992, Comparison of Systematic and Random Sampling for Estimating the Accuracy of Maps Generated from Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing*, **58**(9), 1343–1350.

Swain, P. H., and Davis, S. M. (eds.), 1978, *Remote Sensing: The Quantitative Approach* (New York: McGraw-Hill).

Toll, D. L., 1984, An evaluation of simulated TM data and Landsat MSS data for determining suburban and regional land use and land cover. *Photogrammetric Engineering and Remote Sensing*, **50**(12), 1713–1724.

Toll, D. L., 1985, Effect of Landsat TM sensor parameters on land cover classification. *Remote Sensing of Environment*, **17**(2), 129–140.

Townshend, J., and Justice, J., 1981, Information extraction from remotely sensed data, a user view. *International Journal of Remote Sensing*, **2**, 313–329.

Treitz, P., and Howarth, P., 2000, High spatial resolution remote sensing data for forest ecosystem classification: An examination of spatial scale. *Remote Sensing of Environment*, **72**, 268–289.

Wang, J., Treitz, P. M., and Howarth, P. J., 1992, Road network detection from SPOT imagery for updating geographical information systems in the rural–urban fringe. *International Journal of Remote Sensing*, **6**, 141–157.

Woodcock, C. E., and Strahler, A. H., 1987, The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**, 311–332.

Authors Queries

Journal: **International Journal of Remote Sensing**
Title: **Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case**
Authors: **D. Chen *et al.***

Dear Author
During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof. Many thanks for your assistance

Query Reference	Query	Remarks
1	Treiz and Howarth 2000 – please check spelling of Treiz. Is Treitz?	
2	Please supply IFOV in full use.	
3	Markham and Townshend, 1981 – please supply details of publisher and place of publication.	
4	In Figure 3 are all resolutions in metres?	
5	In Figure 4 are both resolutions in metres?	
6	In Figure 7 caption, can put 'resolutions' in place of a second 'results'?	