The Use of Census Data in Urban Image Classification

Victor Mesev

Abstract

A supervised classification strategy containing a suite of techniques that allow the linking of urban land cover from remotely sensed data with urban functional characteristics from population census data is outlined and demonstrated. For a stronger link, census tract data are also interpolated into more disaggregated and more precise raster-based surfaces using GIS. Census data in tabular and surface format are then used to modify maximum-likelihood classifications through stratified class a priori *probabilities, and in terms of assisting the selection of training samples and contextual post-classification sorting. The strategy is applied to the classification of housing density of four settlements in the United Kingdom. The results show high site-specific accuracy, and improvements in class area estimates.*

Introduction

Urban areas are undoubtedly one of the most challenging surfaces for image classification. Yet the dynamic nature of settlements means that the classification of urban areas has perhaps the greatest potential among the widest audience. For example, practitioners of urban monitoring, management, planning, and land-use zoning activities all need detailed information on the morphology, and especially functional use, of urban land at frequent time intervals. These needs have in part been met by image classifications which have allowed the consistent interpretation of the physical structure of urban land cover, albeit at somewhat small-scale distinctions of urban/non-urban, built/non-built, and, at best, categorical building sizes and shapes (Forster, 1993). Limitations to more detailed urban properties are of course the result of the inherent spatial variabilities and composition heterogeneity of urban surfaces, both of which lead to pixels with multiple class membership (Forster, 1985; Haack *et al.*, 1987). As a consequence, per-pixel image classifications have become less favored than approaches that seek to examine the textural, contextual, and spatial (Barnsley and Barr, 1996) properties and patterns of neighboring pixels, as well as sub-pixel class member compositions using, for example, fuzzy sets or assuming linear relationships between pixel values and land-cover proportions.

However, most of these extensions to per-pixel methods have had somewhat variable degrees of success in producing classifications that are more accurate than simple per-pixel categorization. Indeed, given the additional calculations needed to perform neighborhood or sub-pixel classifications, similar accuracy results have been routinely produced from standard per-pixel methods guided by ancillary information (Harris and Ventura, 1995), including a modification of the per-pixel algorithm itself to take into account non-spectral information on the structure and characteristics of urban areas (Mesev *et al.*, 1998). Ancillary data could also possibly be used to represent the functional characteristics of urban areas which could then be linked with urban structural properties derived from remote sensing. If urban monitoring and planning is desired at the city and regional levels, what is needed is a strategy which could introduce functional-related data, for instance, housing type and population, into standard image processing to produce land-cover classifications that are more consistent with both the physical layout of urban areas as well as their socio-economic distributions.

This paper will outline a strategy within which such additional urban-related data from the United Kingdom Census of Population (OPCS, 1991) can be used to construct a unique urban classification strategy (Figure 1) as well as contribute to research on remote sensing/GIS integration. Census data and interpolated surfaces of census data are incorporated into standard supervised image processing at three stages: before, during, and after classification (Hutchinson, 1982): i.e.,

- Before, or pre-classification, census surfaces are used to assist the selection of class training samples;
- After classification, census surfaces are used to assist in post-classification sorting; and
- During classification, census data in tabular form are normalized and used as class *a priori* probabilities in a Bayesian-modified maximum-likelihood estimator.

Census Data and Census Data Surfaces

Before this three-forked strategy is examined, it's important to first define and conceptually justify the use of tabular and, more importantly, surface-based census data. Socio-economic and housing data from national population censuses are represented as aggregated spatial units, known as census tracts (in the UK these are enumeration districts (EDs)). If research is at the city level, a collection of interrelated census tracts can be accurately and reliably used to calculate a number of functional attributes (for instance, population size, social composition, housing type, etc.). However, as census attributes are assumed to be uniform within a census tract, there is no relation between a settlement's physical structure and its functional characteristics. In other words, within a census tract land cover and land use are unrelated and indistinguishable. What is needed is a means of disaggregating the census tract and filtering out areas that are non-built and non-residential. Remotely sensed data are a convenient source for separating built form from open spaces and vegetation, and, to some degree, separating buildings that may or may not be used for residence. Where image data fall short, support can be given from census data which are spatially manipulated to determine more accurately the location of

School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, United Kingdom (t.v.mesev@bris.ac.uk).

Photogrammetric Engineering & Remote Sensing, Vol. 64, No. 5, May 1998, pp. 431–438.

0099-1112/98/6405–431$3.00/0

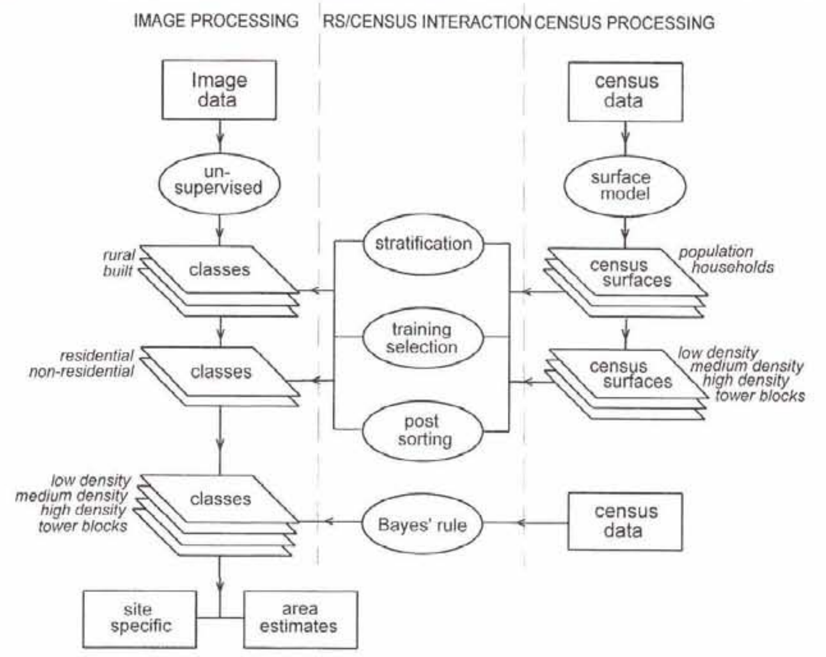

Figure 1. Methodology: Interaction between census data and image classification.

residential land use. The basis for this manipulation is the use of population-weighted centroids which act as pointers to where within the tract the greatest concentration of residential land use is located (Figure 2). By using distances between centroids, a surface can be interpolated which takes into account residential density and, therefore, an approximation of the residential, not urban, geography of the settlement.

Surface models have long been used to represent point data as more continuous distributions which are essentially scale-free from rigid zonal collection units, and are used to reveal more apparent trends and patterns. The arrival of GIS has facilitated the easier manipulation and more efficient storage of surface models, thereby realizing immense potential for analyzing large amounts of population-related datasets at consistent scales and frequent temporal intervals. The surface model used in this paper is derived from work by Martin and Bracken (1991), and is essentially a smoothing, point-based areal interpolation algorithm with a weighting factor based on intercentroid distances. It is able to transform spatial data from irregular zonal units (census tracts) (Figure 2) into regular units (surface cells) at finer levels of spatial disaggregation (Figure 3). Basically, the census value associated with each ED centroid is distributed spatially according to a simple model of distance-decay, implemented by centering a moving window (kernel) over the cells containing each centroid. For any cell i, the variable Y_i is allocated as

$$Y_i = \sum_{j=1}^{c} A_j Z_{ij} \quad (1)$$

where Y_i is the estimated value in the ith cell of the output surface, A_j is the value of the variable assigned to the jth centroid (where c is the total number of centroids in the model area), and Z_{ij} is a weighting of cell i relative to centroid j (based on the distance-decay assumptions). The surface produced has been documented to closely approximate (over 90 percent) the spatial dimensions of urban areas represented from remotely sensed imagery (Bracken and Martin, 1991). It was also selected as the most appropriate means for preparing and representing census data to be used in urban image classifications for three reasons: it provides the basis for population-weighted centroids, it takes into account localized intercentroidal distance decay functions, and it preserves the total volume under the surface. By calculating each census tract centroid as the population *center-of-mass*, the model ensures that most surface cells fall within the residential morphology of an urban area, thereby excluding non-built areas and non-residential urban land use. In calculating intercentroidal distances (with a decay function), and by ensuring that the total population is preserved, each cell is allocated a representative density value which is an estimated proportion of the whole area.

With the assistance of standard GIS data query operations, the raster-based surface model is able to reconstruct a much larger-scale geography of residential patterns, with each cell able to represent the density values of population and housing attributes. For each cell, census variables are represented as probability values at two scales, local and global. Local probabilities define the relative proportions of each variable a, composed of J states (for discrete) or J values (for continuous) of each cell Y_i of the surface model, and assuming $0 \leq \Pr(Y_i) \leq 1$, where $i = 1, 2, \ldots, I$; and $\Sigma_i \Pr(Y_i) = 1$. These local probabilities associated with discrete surface cells are used at both the training stage, as well as for post-classification sorting. Global probabilities are independent of the surface model and determine the relative proportions for each variable within a defined area of the image, or stratum, and again assume $0 \leq \Pr(a_j) \leq 1$, where $j = 1, 2, \ldots, J$; and $\Sigma_j \Pr(a_j) = 1$. These global probabilities associated with variables a_j are used directly in the Bayes' modified rule as weights, or *a priori* probabilities, for each spectral class.

Classification Strategy

Training Area Selection

Census surfaces using local probabilities bring into urban image classification for the first time a reliable and consistent means with which to select training samples based on census

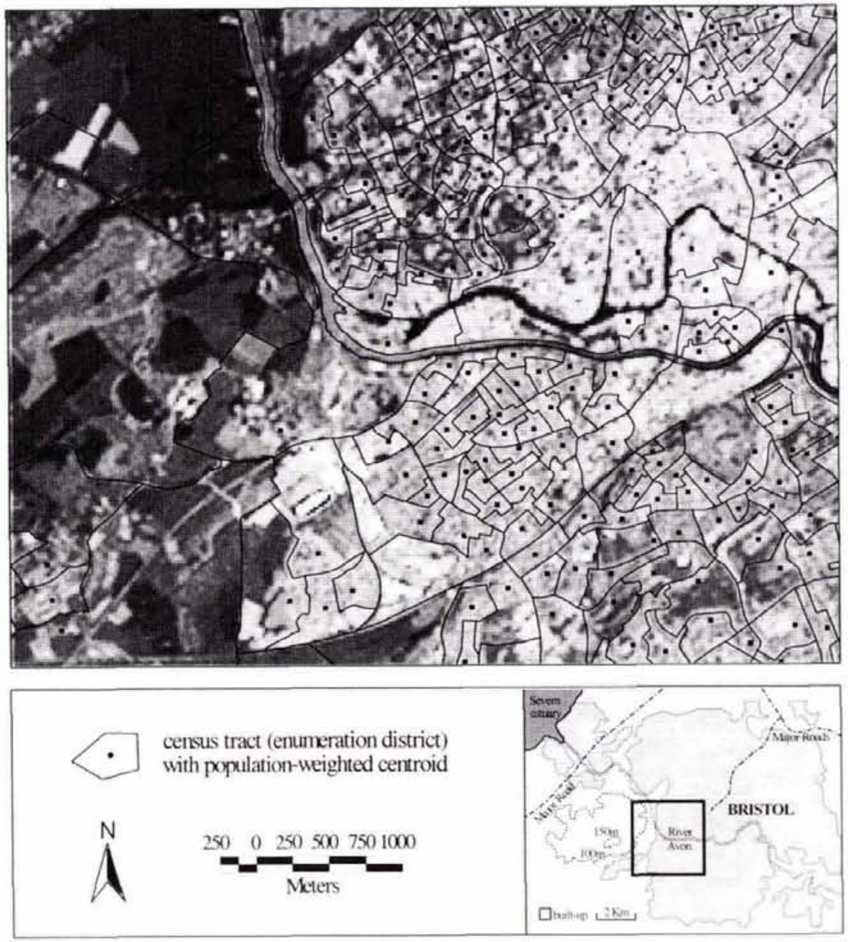

Figure 2. Census tracts (enumeration districts in the UK) with population-weighted centroids of Bristol, England. Note that the sizes of the tracts represent residential density, and these closely coincide with different areas of the SPOT HRV-XS background image. Larger in the downtown area and rural periphery, but much smaller in the suburbs. Also, centroids are population weighted and must fall within a built area.

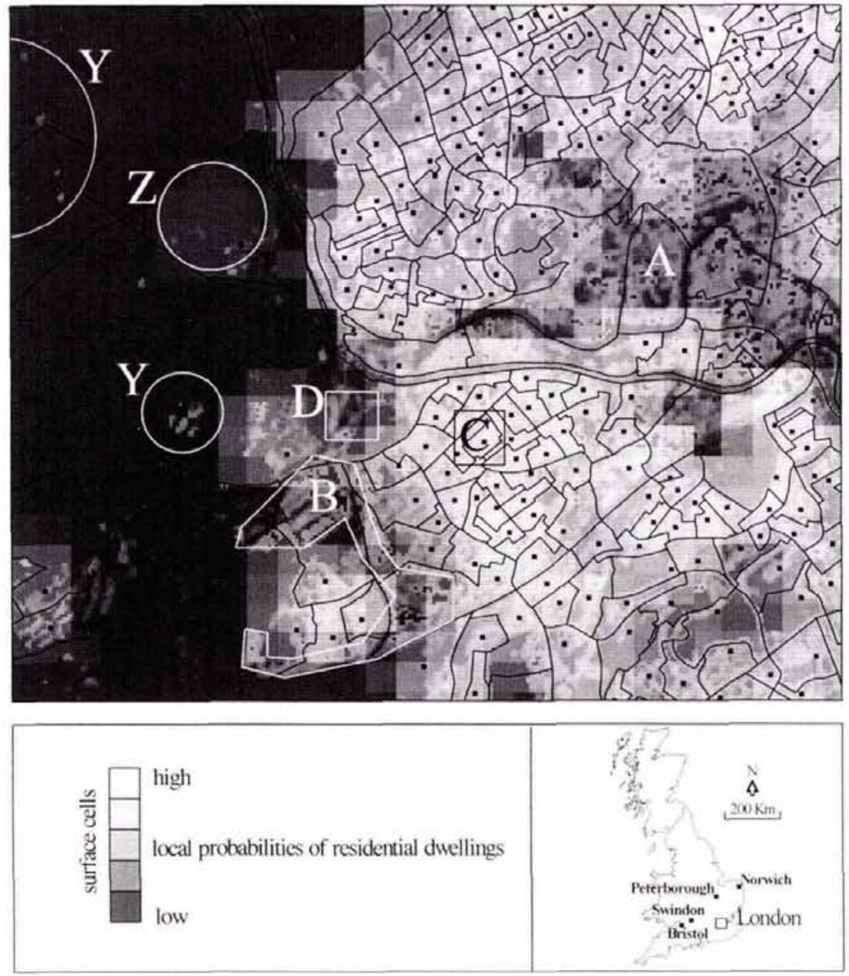

Figure 3. Generalized training samples and post-classification sorting using the residential census surface and SPOT image. Non-residential areas, represented by low residential probabilities, are readily apparent at A (downtown) and B (industrial estate); whilst high probabilities are evident in suburbs C and D. Y indicates areas classified as residential but not coincident with the surface (bare soil), and Z is represented by the surface but not coincident with the residential class (heavily forested community).

distributions. The process of signature extraction, however, still relies on interpretation skills, but these surfaces are a way of increasing the amount of information on urban functional characteristics available to the analyst. This additional information can then be used to differentiate between similar urban classes as well as to improve the quality of the class label. When selecting class signatures, there is a trade-off between having a sufficiently large sample size to ensure accurate statistical parameters used by classifiers, and restrictive enough to ensure class separability. It is essential not to exclude any important pixels that would contribute to the representation of the class, but it is equally essential from a computational standpoint not to include redundant pixels in the classification. Therefore, there needs to be a balance between sample size and sample error (Curran and Williamson, 1986), because the generation of representative training statistics is sometimes more important for obtaining accurate classifications than is the selection of classifier algorithm itself. Generating training statistics for spectral classes that are spectrally distinct and mutually exclusive do not pose problems. Difficulties arise with images of urban areas where heterogeneous land cover leads to severe spectral variability and where uni-modal class statistics are not easily selected.

It is argued here that, although the spatial resolutions between image pixels and surface cells are different, their mismatch still enables the generation of sensible training statistics. The following line of argument may clarify this. SPOT-XS images are commonly resampled and represented at a 20-m spatial resolution, while the surface model produces cells with an optimal 200-m representation (for standard British urban areas) (Martin and Bracken, 1991). This means that, when spatially integrated, a single surface cell will cover 100 image pixels. As training samples are generated from contiguous groups of pixels representing areas, not points, this resolution imbalance is more of a benefit than a hindrance. Training samples of land-cover characteristics can be collected within the spatial limits of surface cells representing associated census functional local probabilities (Quarmby *et al.*, 1988). However, it must be stressed that the surface model is only an approximation to the residential geography of an urban area. As such, the analyst should only use the model as a flexible framework within which to base sampling decisions.

Post-Classification Sorting

The other use of census surfaces is in post-classification sorting. This is simply the resolution of potentially misclassified pixels, and the role of the surfaces is to verify the classification of the residential land-use category from other built land covers. In this capacity, the surface model closely follows theoretical assumptions in the use of contextual model-based approaches to the segmentation of satellite imagery. The essential reasoning of such approaches is primarily based on the ability to incorporate context more formally into the interpretation of a scene. The method proposed here involves the use of census surfaces as simple, spatially approximate contextual devices for "weeding out" potentially misclassified residential pixels. The implementation relies on the spatial registration of the residential classified stratum with the residential surface to common geometrical co-ordinates. Acting very much as a template, surface cells are then used to determine whether pixels classified as residential fit within the approximate spatial pattern of the local probabilities of the residential surface. If pixels clearly do not conform, they are then labeled as potentially misclassified, and will either await further verification (e.g., ground survey, topographic maps, or aerial photographs) or will be eliminated. Most of these eliminated pixels are expected to be outside the main urban areas, representing quarries, exposed soil, and other land-cover types with reflectance characteristics similar to built structures.

Modified Maximum-Likelihood Classifier

The third, and final component of the classification strategy is the Bayesian modification to the conventional maximum-likelihood (ML) classifier. The Bayes' classifier has already been successfully applied to the physical landscape (Strahler, 1980; Maselli, *et al.*, 1992; Foody *et al.*, 1992), but not until most recently has it become operational in an urban context (Sadler and Barnsley, 1990; Mesev *et al.*, 1998). This is an interesting anomaly because it is exactly when classes are closely related, as in urban environments, that modified ML *a priori* probabilities have been documented to have the most effect (Haralick and Fu, 1983; Tom and Miller, 1984; Mather, 1985). As long as *a priori* probabilities are reliably calculated using ancillary information and applied to feature space that contains mutually exclusive classes, classifications should be more accurate. To preserve these assumptions, *a priori* probabilities will be calculated with reference to census data and applied to stratified feature space that contains all mutually exclusive classes. A brief synopsis of the modification to the ML algorithm should be helpful.

As a parametric classifier, the ML algorithm relies on each training sample being represented by a Gaussian probability density function, completely described by the mean vector and variance-covariance matrix using all available spectral bands. Given these parameters, it is possible to compute the statistical probability of a pixel vector being a member of each spectral class (Thomas *et al.*, 1987). The goal is to assign the most likely class w_j, from a set of N classes, $w_i, \ldots, w_N$, to any feature vector $\mathbf{x}$ in the image. A feature vector $\mathbf{x}$ is the vector $(x_1, x_2, \ldots, x_M)$, composed of pixel values in M features (usually, spectral bands). The most likely class w_j for a given feature vector $\mathbf{x}$ is the one with the highest *a posteriori* probability $\Pr(w_j|\mathbf{x})$. Therefore, all $\Pr(w_j|\mathbf{x})$, $j \in [1 \ldots N]$ are calculated, and w_j with the highest value is selected (Fukunaga and Hummels, 1987). The calculation of $\Pr(w_j|\mathbf{x})$ is based on Bayes' theorem: i.e.,

$$\Pr(w_j|\mathbf{x}) = \frac{\Pr(\mathbf{x}|w_j) \times \Pr(w_j)}{\Pr(\mathbf{x})}. \tag{2}$$

On the left-hand side is the *a posteriori* probability that a pixel with feature vector $\mathbf{x}$ should be classified as belonging to class w_j. The right-hand side is based on Bayes' theorem, where $\Pr(\mathbf{x}|w_j)$ is the conditional probability that some feature vector $\mathbf{x}$ occurs in a given class, in other words, the probability density of w_j as a function of $\mathbf{x}$. Supervised classifications, such as the ML, derive this information from training samples by parametrically assuming normal class probability densities and estimating the mean vector and covariance matrix. The modification to the ML is $\Pr(w_j)$ which is the *a priori* probability of the occurrence of w_j irrespective of its feature vector, and as such is open to estimation by *a priori* knowledge external to the remotely sensed image. External *a priori* knowledge will typically include information on the distribution and relative areas covered by each class in the study scene, which is most readily handled by a GIS. It follows that the accuracy of class priors is at best equal to the quality of ancillary *a priori* knowledge. To further clarify in image classification terms, *a priori* probabilities can be visualized as a means of shifting decision boundaries to produce larger volumes in M-dimensional feature space for classes that are expected to be large and smaller volumes for classes that are expected to be small. The denominator in Equation 2, $\Pr(\mathbf{x})$, is the unconditional probability density which is used to normalize the numerator such that

$$\Pr(\mathbf{x}) = \sum_{j=1}^{N} \Pr(\mathbf{x}|w_j) \times \Pr(w_j). \tag{3}$$

Typically, ML classifiers assume *a priori* probabilities to

Table 1. Categorical Residential Land-Use Densities

	Residential Density Categories (dwellings/ha.)			
Settlement	low density	medium density	high density	tower blocks
Bristol	< 18	18 - 25	26 - 69	70 >
Norwich	< 22	22 - 33	34 - 75	76 >
Swindon	< 20	20 - 29	30 - 129	130 >
Peterborough	< 21	21 - 34	35 - 147	147 >
Average	< 20	21 - 30	31 - 105	105 >

be equal and assign each $\Pr(w_j)$ a value of 1.0. However, variations in *a priori* probabilities can be an important remedy for the problem of spectrally overlapping urban classes. If a feature vector **x** has probability density values that are significantly different from zero for several classes, it is not inconceivable for that pixel to belong to any of these classes. When selecting a class solely on the basis of its spectral characteristics, a large probability of error inevitably results. The use of appropriate *a priori* probabilities, based on reliable supplementary information, is one way to reduce this error in class assignments. Moreover, it would seem intuitively more sensible to suggest that some classes are more likely to occur than others. In this paper, these *a priori* probabilities will be generated from census global probabilities, and applied to four urban images which have been stratified using training samples and post-classification sorting based on census surfaces.

Four Settlements in the United Kingdom

Study Areas and Materials

The classification strategy outlined above will now be tested on four medium-sized settlements in the United kingdom using a SPOT HRV(XS) image for the city of Bristol (320,000 population; 145,000 households), taken on 17 May 1988; and two Landsat 5 (TM) scenes of Swindon (147,000 and 65,000), taken on 30 October 1988, and of Norwich (202,000 and 90,000) and Peterborough (126,000 and 53,000), taken on 15 July 1989. Census data are taken, and surfaces generated, from the UK 1991 Census of Population (OPCS, 1991). All processing is carried out using UNIX-based ERDAS (IMAGINE, 8.2) proprietary software (ERDAS, 1996), along with a suite of purpose-written C and FORTRAN programs.

Methodology

The methodology is based on a series of hierarchical stratifications whereby each scene is systematically classified into a succession of urban land covers/land uses. The first segmentation is generated by a standard ERDAS ISODATA unsupervised classification, and produces a general binary distinction between built and non-built land covers. The built stratum is then classified into residential and non-residential land use using the modified maximum-likelihood estimator and supervised by training samples based on the residential census surface. Figure 3 illustrates how superimposing the built image stratum (shown darker) and the residential surface (shown brighter and coarser) can reveal clear distinctions between residential and non-residential land use in Bristol. Indeed, the surface is such a reasonable approximation with the image that reliable training areas can be easily inferred, for example, non-residential areas corresponding to the central business district (downtown) (labeled A), and an outlying industrial estate (B). Residential density categories are also trained using census surfaces in the same manner. This time, the residential image stratum is stacked with census surfaces representing four levels of residential density. These are low, medium, high, and tower blocks, and are calculated by the average number of dwellings per hectare, which in turn are derived from the average ratios between census counts and census tract areas (Table 1). In Figure 3, training samples for these density categories are selected by surfaces cells of high local probability of occurrence. For instance, the area marked (C) may be used for high density, and (D) for low density.

The same residential surface is also used to assist post-classification sorting of the residential stratum by removing misclassified pixels. The circles, labeled (Y) in Figure 3, represent features that are not covered by the surface, and, upon further investigation using ground and collateral information, these were deemed not as non-residential but as exposed soil and agricultural fields. The area labeled (Z) illustrates a third scenario and corresponds to an area of surface representation but scarce image representation. An investigation into this case later found that localized dispersed housing was partially hidden by a heavily forested habitat and hence misclassified from the SPOT image.

Once sufficient class signatures are collected for residential, non-residential, and each of the four residential density categories, the built and residential strata of the SPOT image are classified using these samples, and global census probabilities in the form *a priori* probabilities within the maximum-likelihood discriminant function. In the UK 1991 Population of Census (OPCS, 1991), residential density is measured by reference to the type of building as well as the number of dwellings per census tract. The majority of building types include detached housing, which corresponds to low density, semi-detached to medium, terraced to high, and apartments to tower blocks. As a consequence, a scaling ratio is needed to convert these census categories into ones of density categories and, at the same time, preserve relative areal proportions. Using stereoscopic aerial photographs, 20 samples of dwelling type sizes were generated, and average relative size ratios between dwelling types were constructed. The ratios were 1 detached dwelling to 1.5 semi-detached, 1 detached to 2.25 terraced, 1 semi-detached to 1.5 terraced, and 1 detached to 10 apartments (Table 2). Although these are approximations, they are still more realistic than assuming absolute linear building type relationships. The *a priori* probabilities are then calculated and inserted as follows.

Consider z_k as the census variable "residential building type" (where k: 1 = detached, 2 = semi-detached, 3 = terrace, and 4 = apartment blocks). When stratified into exclusively residential feature space, the four classes will have A pixels with feature values x_i, where $x_1, \ldots, x_A$ are not necessarily mutually exclusive. The objective is to find the probability that a random pixel (within the residential stratum of the image) will be a member of a spectral class w_j (where j: 1 = low density, 2 = medium density, 3 = high density, 4 = tower blocks), given its density vector of observed measurements **x**, in m-dimensional feature space *and* that it belongs to ancillary class z_k, described as

$$\Pr(w_j \mid \mathbf{x}, z_k). \tag{4}$$

Table 2. Census *a priori* Probabilities and Scaling Factors

	Census *a priori* probabilities			
Settlement	low density	medium density	high density	tower blocks
Bristol	0.1145	0.3779	0.4686	0.0390
Norwich	0.4240	0.3027	0.2393	0.0339
Swindon	0.3014	0.3642	0.3138	0.0210
Peterborough	0.3625	0.3500	0.2648	0.0227
Scaling	1.00	1.50	2.25	10.00

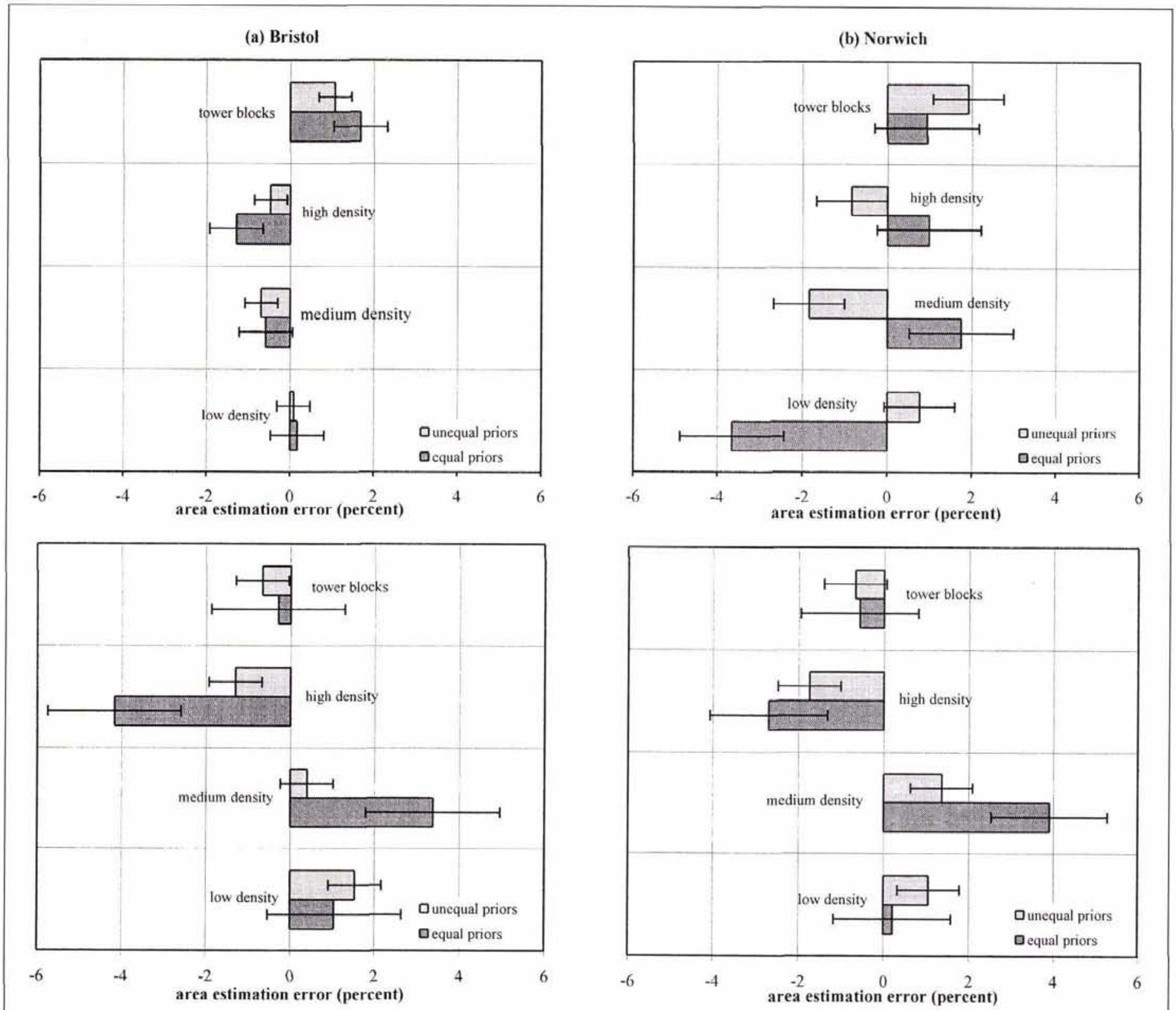

Figure 4. Area estimation errors under conditions of equal and unequal *a priori* probability classifications.

It is also assumed that the effects of z_k are external to the original generation of the mean vector and covariance matrix of w_j. As a result, the likelihood function $\Pr(w_j \mid \mathbf{x})$ is unaltered by the introduction of z_k, but is simply modified by the conditional probability

$$\Pr(w_j \mid z_k). \tag{5}$$

This is a process of identifying the association between spectral class w_j with census variable z_k. For example, the spectral class labelled as low density residential would be directly associated with a conditional probability of the census variable "detached dwellings." In effect, w_1 is weighted by the probability of z_1, producing the *a priori* probability of $\Pr(w_1)$. In our example we assume that the *a priori* probabilities of each of the four dwelling types exist in inclusive *M*-dimensional feature space, so that $\Pr(w_1) + \Pr(w_2) + \Pr(w_3) + \Pr(w_4) = 1.0$. The probability densities $d_{i1} = \Pr(\mathbf{x}_i \mid w_1)$, $d_{i2} = \Pr(\mathbf{x}_i \mid w_2)$, $d_{i3} = \Pr(\mathbf{x}_i \mid w_3)$, $d_{i4} = \Pr(\mathbf{x}_i \mid w_4)$, are known for each pixel. Let l_{i1} be the shorthand for the *a posteriori* probability $\Pr(w_1 \mid x_i, z_1)$ that pixel *i* belongs to class w_1, and let p_j be the shorthand for the *a priori* probabilities. The Bayesian modified ML is now represented as

$$l_{i1} = \frac{d_{i1} p_1}{d_{i1} p_1 + d_{i2} p_2 + d_{i3} p_3 + d_{i4} p_4}. \tag{6}$$

Likewise, $l_{i2} = \Pr(w_2 \mid x_i, z_2)$, $l_{i3} = \Pr(w_3 \mid x_i, z_3)$, and $l_{i4} = \Pr(w_4 \mid x_i, z_4)$ may also be calculated, and, of course, the sum of the four *a posteriori* probabilities equals 1.0,

$$l_{ij} = \frac{d_{ij} p_j}{\sum_{j=1}^{4} d_{ij} p_j}. \tag{7}$$

Results

The complete methodology is repeated for all four settlements. From Figure 4, it is readily apparent that classifications generated using census-assisted training area selection, post-classification sorting, and modified *a priori* probabilities,

Table 3. Site-Specific Accuracy Assessment of Bristol, Classified Under (a) Equal *a priori* Probabilities and Unsupervised Methods (b) Census-Assisted Unequal *a priori* Probabilities, and Training Sample Selection and Post-Classification Sorting. Note: Number of Correct Sample Points in Bold and Percentage Correct in Parentheses; with Classification Strata Shown Between Bold Lines, and 250 Sample Points in Each Stratum

(a)

CLASSIFIED	REFERENCE DATA							
Kappa coefficient	*K = 0.939; n = 250*		K = 0.713; n = 250		*K = 0.607; n = 250*			
categories	*non-built*	*built*	non-resid.	residential	*low*	*medium*	*high*	*blocks*
non-built	86 (96)	*0*						
built	*7*	*157 (100)*	0	0				
non-residential		1	**39** (74)	3	*0*	*0*	*0*	*0*
residential		6	14	**187** (98)				
low density			*0*		*22 (76)*	*3*	*2*	*0*
medium density			*2*		*2*	*72 (82)*	*13*	*3*
high density			*14*		*3*	*12*	***81** (82)*	*5*
tower blocks			*1*		*2*	*1*	*3*	*9 (53)*

(b)

CLASSIFIED	REFERENCE DATA							
Kappa coefficient	*K = 0.939; n = 250*		K = 0.786; n = 250		*K = 0.737; n = 250*			
categories	*non-built*	*built*	non-resid.	residential	*low*	*medium*	*high*	*blocks*
non-built	86 (96)	*0*						
built	*7*	*157 (100)*	0	0				
non-residential		1	**47** (89)	2	*0*	*0*	*0*	*0*
residential		6	6	**188** (99)				
low density			*0*		*25 (86)*	*2*	*1*	*0*
medium density			*2*		*1*	***81** (92)*	*7*	*2*
high density			*14*		*2*	*5*	***90** (91)*	*5*
tower blocks			*1*		*1*	*0*	*1*	*10 (59)*

in most cases, resulted in more accurate areal estimates with lower percentage errors and smaller standard errors. Reference data on the areal coverages of residential density categories were calculated directly from the population census. However, there do not seem to be any clear distinctions between density categories, with no systematic pattern to the errors. Perhaps multitemporal images and further case studies would provide some indication of whether some land uses have inherent propensity to over or under predict class areas. Only then can adjustments be made both to the Bayes' modifier, class definitions, and scaling factors.

A detailed site-specific accuracy assessment was conducted to examine more closely the effects of the overall methodology. Table 3 summarizes the results from the Bristol stratified classifications under (a) standard supervised methods, and (b) *a priori* probabilities modified using census data, and training sample selection and post-classification sorting using the residential and residential density category census surfaces. The strata were each verified by 250 random samples of ground truth points, collected at distinguishable locations in the image (usually at main road junctions), by manual observation. The built stratum was generated by an unsupervised classification, and so the results are the same in both (a) and (b). However, kappa coefficients for classifications using the census-assisted methods improved markedly for both the residential/non-residential, and residential density strata. In the former, the greatest improvement was in a number of samples correctly classified as non-residential (mostly on the urban periphery). A larger increase in the kappa, from 0.607 to 0.737, was calculated for the residential density stratum, which can be attributed to a combination of more informed training areas and responsive class *a priori* probabilities. These results are certainly comparable with other urban land-use classification studies, and seem to suggest that further research on the use of additional information on urban functional characteristics (such as census data) should be seriously investigated, or at least welcomed, into urban image classifications.

Conclusions

This paper has called for greater integration between functional data and remote sensing in the much neglected field of urban image classification. It has argued how such attribute data (handled by GIS) can make substantial in-roads into the quality of information that can be used to supervise training sample selection, post-classification sorting, as well as the application of stratified census probabilities in a modified maximum- likelihood classifier. On the whole, tabular census and census surfaces have brought a number of benefits to standard image classification by

- Improving the *a priori* probabilities of spectrally overlapping urban categories.
- Providing a raster-based data structure for easier integration with image data.

- By allowing surfaces to be unconstrained by zonal boundaries and transforming point-based centroids into areal surfaces, they thereby satisfy conditions for training sample selection.
- Representing urban functional attributes within the physical layout of the urban built form. This ultimately will provide scope for representing other human-related data, including postal information and address-matching.
- Creating a fully automated, nationally available vehicle for consistent local- and regional-scale information. These extensive small-scale classification improvements can assist small-scale spatial analyses of the form and structure of entire urban systems, both spatially and temporally.

The repercussions and spin-offs from this work are particularly aimed at urban monitoring and analysis at the city-wide and regional level. Surfaces of human-related activities will allow various aspects of demographic, socio-economic, and housing characteristics to be modeled within the structural limits of urban morphologies outlined by image data.

Acknowledgments
This work was partly funded by a Research Fellowship (number H53627501295) awarded by the UK Economic and Science Research Council (ESRC).

References

Barnsley, M.J., and S.L. Barr, 1996. Inferring Urban Land Use from Satellite Sensor Images Using Kernel-Based Spatial Reclassification, *Photogrammetric Engineering &nd Remote Sensing*, 62(8): 949–958.

Curran, P.J., and H.D. Williamson, 1986. Sample Size for Ground and Remotely-Sensed Data, *Remote Sensing of Environment*, 20: 31–41.

ERDAS, 1996. *ERDAS Imagine (8.2): User's Manual*, Earth Resources Data Analysis Systems Inc., Atlanta, Georgia.

Foody, G.M., N.A Campbell, N.M. Trodd, and T.F. Wood, 1992. Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-Likelihood Classification, *Photogrammetric Engineering & Remote Sensing*, 58(9):1335–1341.

Forster, B.C., 1985. An Examination of Some Problems and Solutions in Monitoring Urban Areas from Satellite Platforms, *International Journal of Remote Sensing*, 6:139–151.

———, 1993. Coefficient of Variation as a Measure of Urban Spatial Attributes, Using SPOT HRV and Landsat TM Data, *International Journal of Remote Sensing*, 14:2403–2409.

Fukunaga, K., and D.M. Hummels, 1987. Bayes Error Estimation Using Parzen and K-NN Procedures, *IEEE Transactions PAMI*, 9: 634–643.

Haack, B., N. Bryant, and S. Adams, 1987. An Assessment of Landsat MSS and TM Data for Urban and Near-Urban Land-Cover Digital Classification, *Remote Sensing of Environment*, 21:201–213.

Haralick, R.M., and K. Fu, 1983. Pattern Recognition and Classification, *Manual of Remote Sensing* (R. Colwell, editor). American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia.

Harris, P.M., and S.J. Ventura, 1995. The Integration of Geographic Data with Remotely Sensed Imagery to Improve Classification in an Urban Area, *Photogrammetric Engineering & Remote Sensing*, 61(8):993–998.

Hutchinson, C.F., 1982. Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement, *Photogrammetric Engineering & Remote Sensing*, 48(2):123–130.

Langford, M., and D.J. Unwin, 1994. Generating and Mapping Population Density Surfaces Within a Geographical Information System, *The Cartographic Journal*, 31:21–26.

Martin, D.J., and I. Bracken, 1991. Techniques for Modelling Population-Related Raster Databases, *Environment and Planning A*, 23: 1069–1075.

Maselli, F., C. Conese, L. Petkov, and R. Resti, 1992. Inclusion of Prior Probabilities Derived from a Nonparametric Process into the Maximum-Likelihood Classifier, *Photogrammetric Engineering & Remote Sensing*, 58(2):201–207.

Mather, P.M., 1985. A Computationally-Efficient Maximum Likelihood Classifier Employing Prior Probabilities for Remotely-Sensed Data, *International Journal of Remote Sensing*, 6(2):369–376.

Mesev, T.V., B. Gorte, and P.A. Longley, 1998. Modified Maximum Likelihood Classifications and their Application to Urban Remote Sensing, *Remote Sensing and Urban Analysis* (J.P. Donnay and M.J. Barnsley, editors), Taylor & Francis, London (in press).

Michalak, W.Z., 1993. GIS in Land Use Change Analysis: Integration of Remotely-Sensed Data into GIS, *Applied Geography*, 13:28–44.

OPCS (Office of Population Censuses and Surveys), 1991. *1991 Census User's Guides, 1-45*, General Register Office for Scotland, UK.

Quarmby, N.A., J.L. Cushnie, and R.J. Saull, 1988. Remote Sensing for Real World Applications, *Proceedings of Image Processing '88*, Blenheim Online, London, pp. 151–162.

Sadler, G.J., and M.J. Barnsley, 1990. *Use of Population Density Data to Improve Classification Accuracies in Remotely-Sensed Images of Urban Areas*, South East Regional Research Laboratory Working Report Number 22, Birkbeck University, London.

Strahler, A.H., 1980. The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely-Sensed Data, *Remote Sensing of Environment*, 10:135–163.

Thomas, I.L., V.M. Benning, and N.P. Ching, 1987. *Classification of Remotely-Sensed Images*, IOP, Bristol.

Tom, C.H., and L.D. Miller, 1984. An Automated Land-Use Mapping Comparison of the Bayesian Maximum Likelihood and Linear Discriminant Analysis Algorithms, *Photogrammetric Engineering & Remote Sensing*, 50(2):193–207.

(Received 15 February 1997; accepted 8 September 1997; revised 5 November 1997)