Spatial Techniques for Image Classification*

Selim Aksoy
Bilkent University
Department of Computer Engineering
Bilkent, 06800, Ankara, Turkey
saksoy@cs.bilkent.edu.tr

Abstract

The constant increase in the amount and resolution of remotely sensed imagery necessitates development of intelligent systems for automatic processing and classification. We describe a Bayesian framework that uses spatial information for classification of high-resolution images. First, spectral and textural features are extracted for each pixel. Then, these features are quantized and are used to train Bayesian classifiers with discrete non-parametric density models. Next, an iterative split-and-merge algorithm is used to convert the pixel level classification maps into contiguous regions. Then, the resulting regions are modeled using the statistical summaries of their spectral, textural and shape properties, and are used with Bayesian classifiers to compute the final classification maps. Experiments with three ground truth data sets show the effectiveness of the proposed approach over traditional techniques that do not make strong use of region-based spatial information.

1 Introduction

The amount of image data that is received from satellites is constantly increasing. For example, nearly 3 terabytes of data are being sent to Earth by NASA's satellites every day [1]. Advances in satellite technology and computing power have enabled the study of multi-modal, multi-spectral, multi-resolution and multi-temporal data sets for applications such as urban land use monitoring and management, GIS and mapping, environmental change, site suitability, agricultural and ecological studies. Automatic content extraction, classification and content-based retrieval have become highly desired goals for developing intelligent systems for effective and efficient processing of remotely sensed data sets.

There is an extensive literature on classification of remotely sensed imagery using parametric or non-parametric statistical or structural techniques with many different features [2]. Most of the previous approaches try to solve the content extraction problem by building pixel-based classification and retrieval models using spectral and textural features. However, a recent study [3] that investigated classification accuracies reported in the last 15 years showed that there has not been any significant improvement in the performance of classification methodologies over this period. The reason behind this problem is the large semantic gap between the low-level features used for

*This work was supported by the TUBITAK CAREER Grant 104E074 and European Commission Sixth Framework Programme Marie Curie International Reintegration Grant MIRG-CT-2005-017504.

classification and the high-level expectations and scenarios required by the users. This semantic gap makes a human expert's involvement and interpretation in the final analysis inevitable, and this makes processing of data in large remote sensing archives practically impossible. Therefore, practical accessibility of large remotely sensed data archives is currently limited to queries on geographical coordinates, time of acquisition, sensor type and acquisition mode [4].

The commonly used statistical classifiers model image content using distributions of pixels in spectral or other feature domains by assuming that similar land cover/use structures will cluster together and behave similarly in these feature spaces. However, the assumptions for distribution models often do not hold for different kinds of data. Even when nonlinear tools such as neural networks or multi-classifier systems are used, the use of only pixel-based data often fails the expectations.

An important element of image understanding is the spatial information because complex land structures usually contain many pixels that have different feature characteristics. Remote sensing experts also use spatial information to interpret the land cover because pixels alone do not give much information about image content. Image segmentation techniques [5] automatically group neighboring pixels into contiguous regions based on similarity criteria on pixels' properties. Even though image segmentation has been heavily studied in image processing and computer vision fields, and despite the early efforts [6] that use spatial information for classification of remotely sensed imagery, segmentation algorithms have only recently started receiving emphasis in remote sensing image analysis. Examples of image segmentation in the remote sensing literature include region growing [7] and Markov random field models [8] for segmentation of natural scenes, hierarchical segmentation for image mining [9], region growing for object level change detection [10] and fuzzy rule-based classification [11], and boundary delineation of agricultural fields [12].

We model spatial information by segmenting images into spatially contiguous regions and classifying these regions according to the statistics of their spectral and textural properties and shape features. To develop segmentation algorithms that group pixels into regions, first, we use non-parametric Bayesian classifiers that create probabilistic links between low-level image features and high-level user-defined semantic land cover/use labels. Pixel level characterization provides classification details for each pixel with automatic fusion of its spectral, textural and other ancillary attributes [13]. Then, each resulting pixel level classification map is converted into a set of contiguous regions using an iterative split-and-merge algorithm [13, 14] and mathematical morphology. Following this segmentation process, resulting regions are modeled using the statistical summaries of their spectral and textural properties along with shape features that are computed from region polygon boundaries [15, 14]. Finally, non-parametric Bayesian classifiers are used with these region level features that describe properties shared by groups of pixels to classify these groups into land cover/use categories defined by the user.

The rest of the chapter is organized as follows. An overview of feature data used for modeling pixels is given in Section 2. Bayesian classifiers used for classifying these pixels are described in Section 3. Algorithms for segmentation of regions are presented in Section 4. Feature data used for modeling resulting regions are described in Section 5. Application of the Bayesian classifiers to region level classification is described in Section 6. Experiments are presented in Section 7 and conclusions are given in Section 8.

2 Pixel Feature Extraction

The algorithms presented in this chapter will be illustrated using three different data sets:

1. *DC Mall*: HYDICE (Hyperspectral Digital Image Collection Experiment) image with $1,280 \times 307$ pixels and 191 spectral bands corresponding to an airborne data flightline over the Washington DC Mall area.

   The *DC Mall* data set includes 7 land cover/use classes: roof, street, path, grass, trees, water, and shadow. A thematic map with ground truth labels for 8,079 pixels was supplied with the original data [2]. We used this ground truth for testing and separately labeled 35,289 pixels for training. Details are given in Figure 1.

2. *Centre*: DAIS (Digital Airborne Imaging Spectrometer) and ROSIS (Reflective Optics System Imaging Spectrometer) data with $1,096 \times 715$ pixels and 102 spectral bands corresponding to the city center in Pavia, Italy.

   The *Centre* data set includes 9 land cover/use classes: water, trees, meadows, self-blocking bricks, bare soil, asphalt, bitumen, tiles, and shadow. The thematic maps for ground truth contain 7,456 pixels for training and 148,152 pixels for testing. Details are given in Figure 2.

3. *University*: DAIS and ROSIS data with 610×340 pixels and 103 spectral bands corresponding to a scene over the University of Pavia, Italy.

   The *University* data set also includes 9 land cover/use classes: asphalt, meadows, gravel, trees, (painted) metal sheets, bare soil, bitumen, self-blocking bricks, and shadow. The thematic maps for ground truth contain 3,921 pixels for training and 42,776 pixels for testing. Details are given in Figure 3.

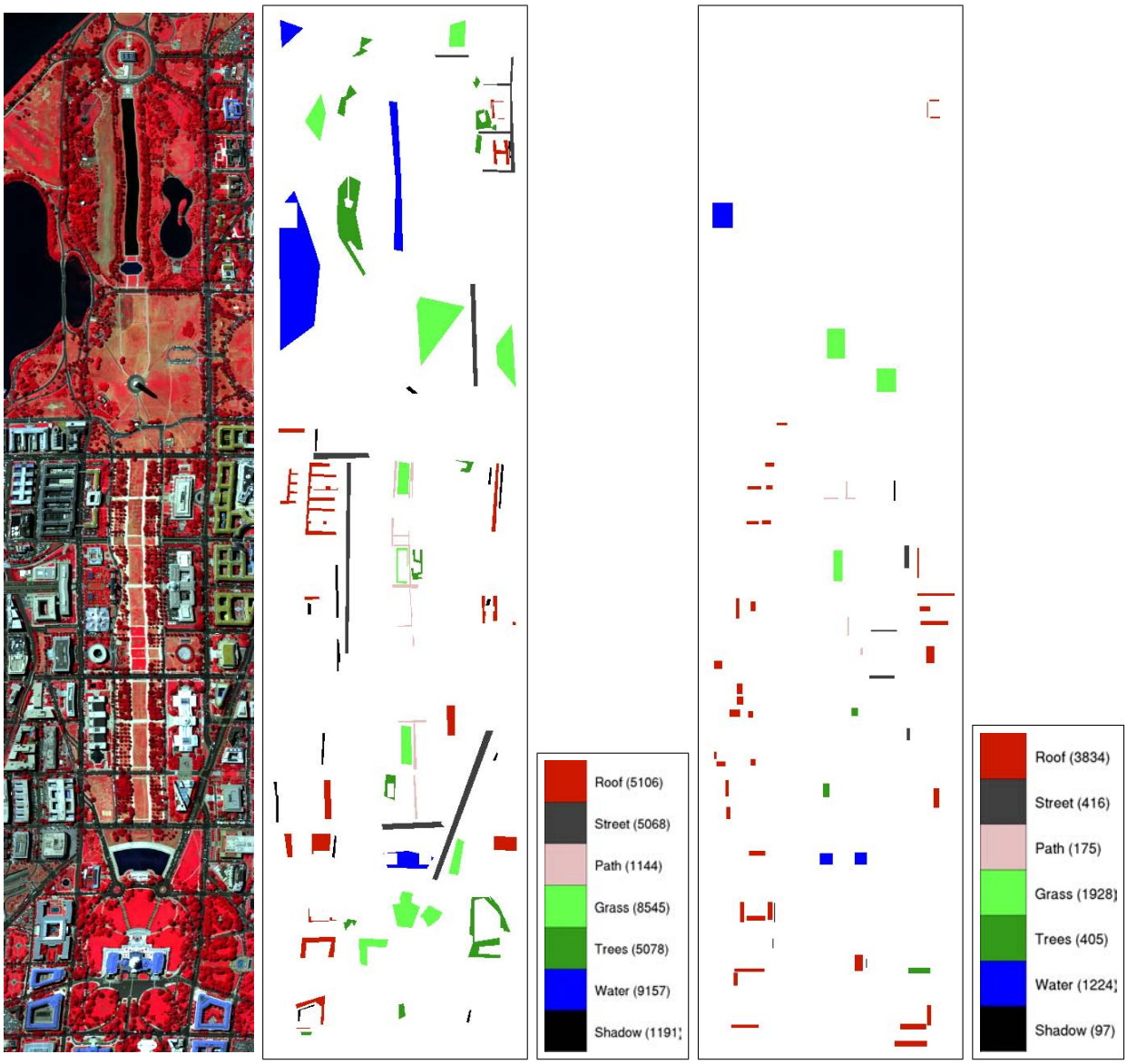
The Bayesian classification framework that will be described in the rest of the chapter supports fusion of multiple feature representations such as spectral values, textural features, and ancillary data such as elevation from DEM. In the rest of the chapter, pixel level characterization consists of spectral and textural properties of pixels that are extracted as described below.

To simplify computations and to avoid the curse of dimensionality during the analysis of hyper-spectral data, we apply Fisher's linear discriminant analysis (LDA) [16] that finds a projection to a new set of bases that best separate the data in a least-squares sense. The resulting number of bands for each data set is one less than the number of classes in the ground truth.

We also apply principal components analysis (PCA) [16] that finds a projection to a new set of bases that best represent the data in a least-squares sense. Then, we keep the top 10 principal components instead of the large number of hyper-spectral bands. In addition, we extract Gabor texture features [17] by filtering the first principal component image with Gabor kernels at different scales and orientations shown in Figure 4. We use kernels rotated by $n\pi/4, n = 0, \ldots, 3$, at 4 scales resulting in feature vectors of length 16. In previous work [13], we observed that, in general, micro-texture analysis algorithms like Gabor features smooth noisy areas and become useful for modeling neighborhoods of pixels by distinguishing areas that may have similar spectral responses but have different spatial structures.

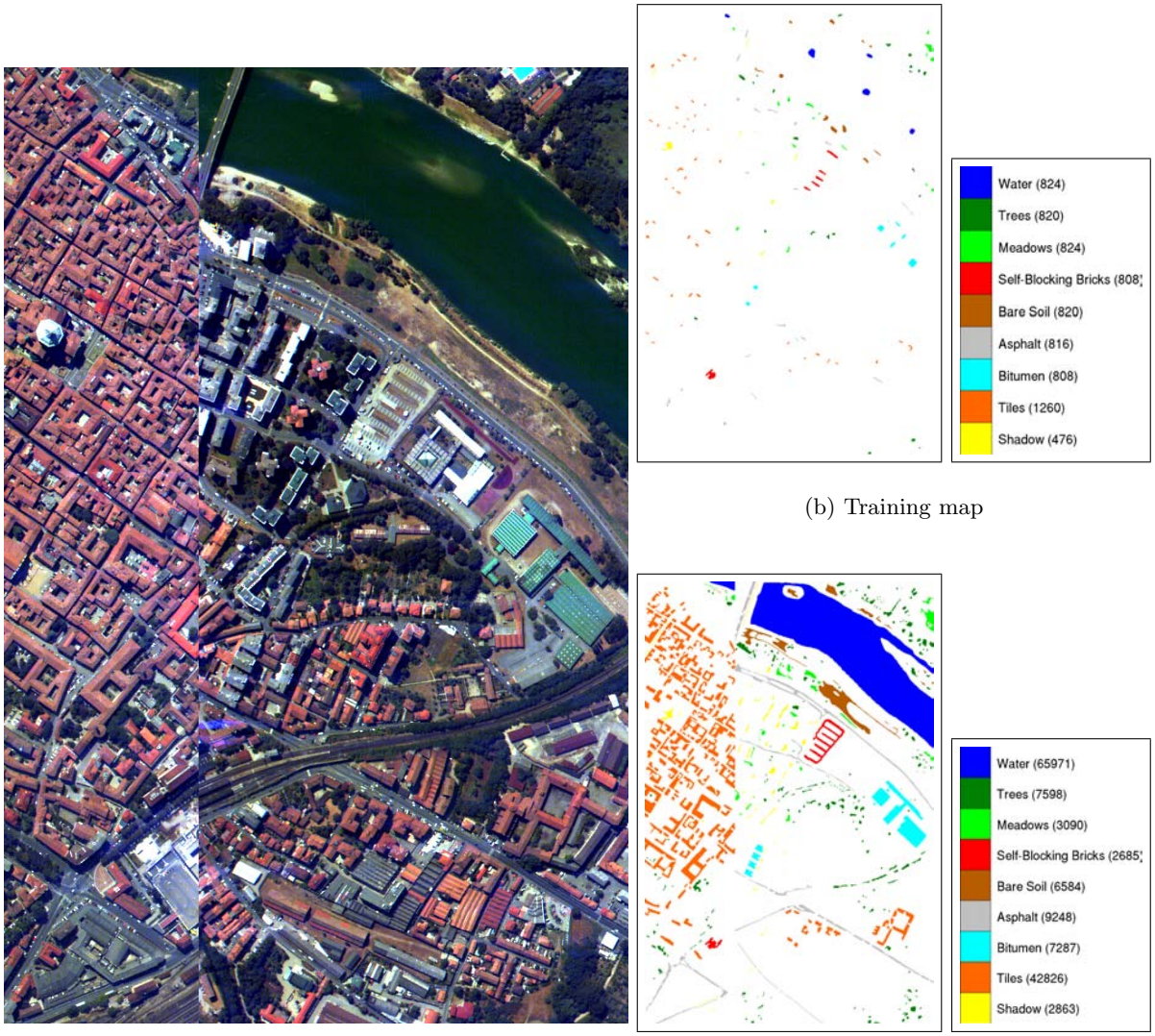
Finally, each feature component is normalized by linear scaling to unit variance [18] as

$$\tilde{x} = \frac{x - \mu}{\sigma} \tag{1}$$



(a) *DC Mall* data (b) Training map (c) Test map

Figure 1: False color image of the *DC Mall* data set (generated using the bands 63, 52 and 36) and the corresponding ground truth maps for training and testing. The number of pixels for each class are shown in parenthesis in the legend.

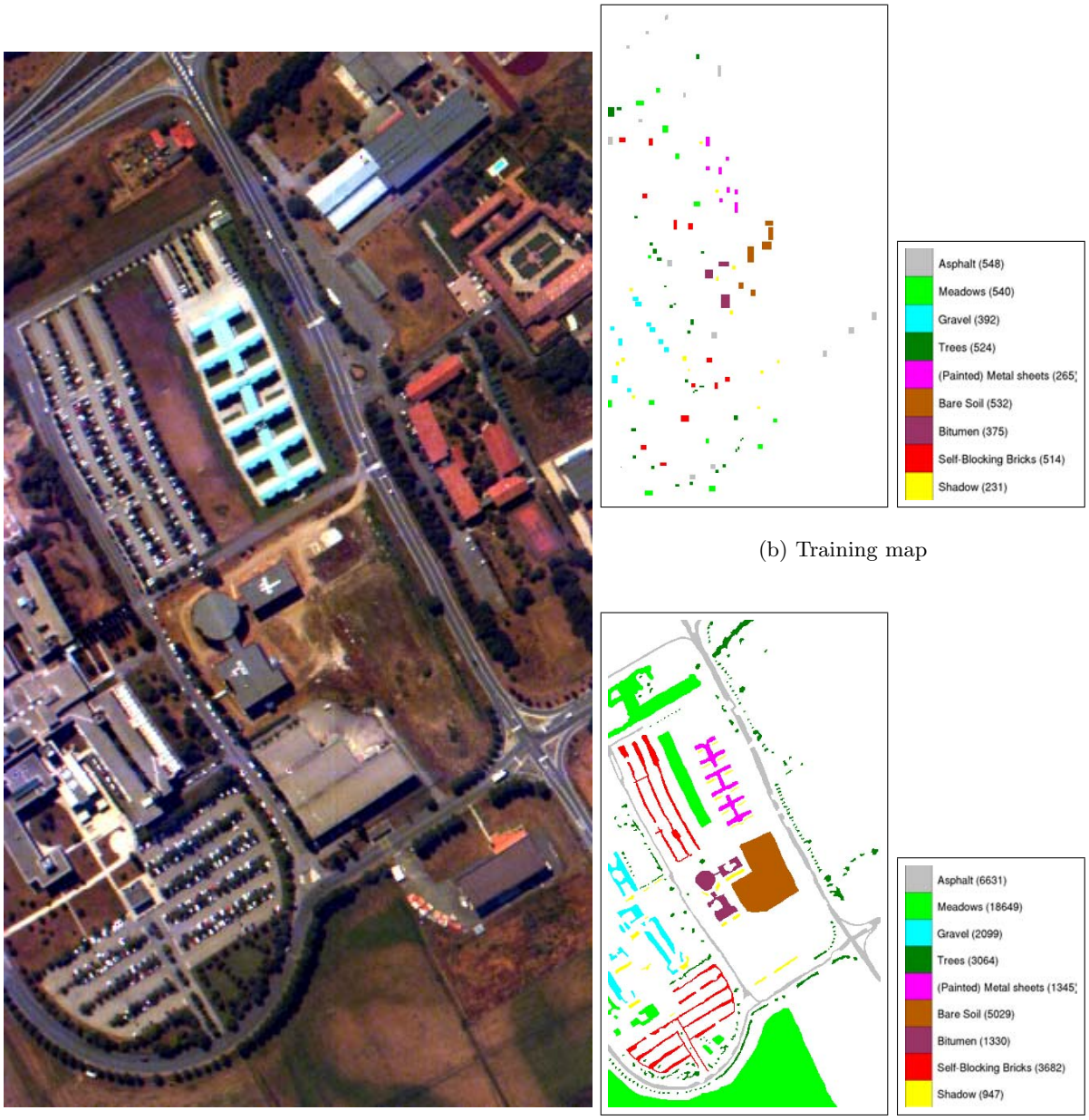


(a) *Centre* data

(b) Training map

(c) Test map

Figure 2: False color image of the *Centre* data set (generated using the bands 68, 30 and 2) and the corresponding ground truth maps for training and testing. The number of pixels for each class are shown in parenthesis in the legend. (A missing vertical section in the middle was removed.)



(a) *University* data

(b) Training map

(c) Test map

Figure 3: False color image of the *University* data set (generated using the bands 68, 30 and 2) and the corresponding ground truth maps for training and testing. The number of pixels for each class are shown in parenthesis in the legend.

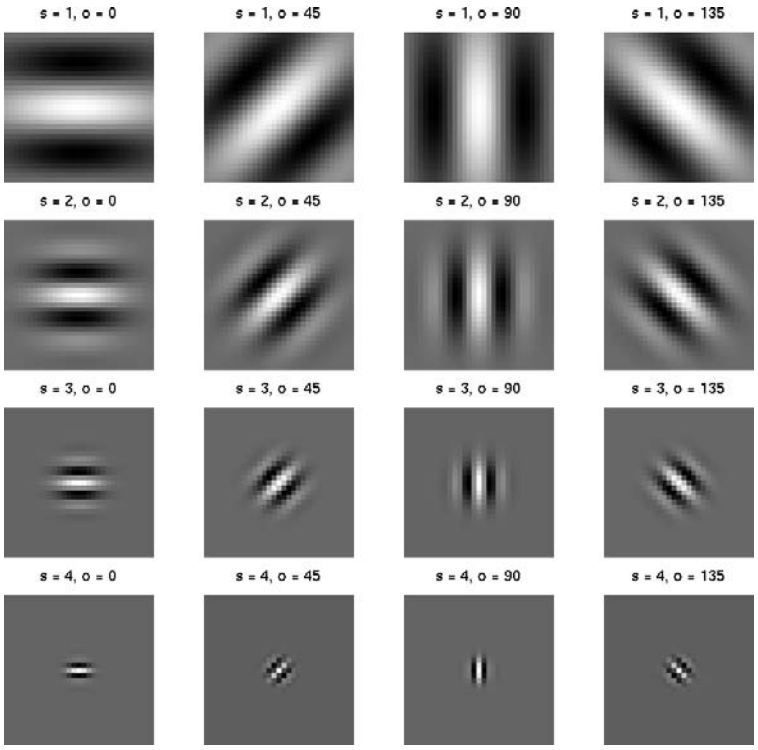


Figure 4: Gabor texture filters at different scales ($s = 1, \ldots, 4$) and orientations ($o \in \{0°, 45°, 90°, 135°\}$). Each filter is approximated using 31×31 pixels.

where x is the original feature value, $\tilde{x}$ is the normalized value, μ is the sample mean, and σ is the sample standard deviation of that feature, so that the features with larger ranges do not bias the results. Examples for pixel level features are shown in Figures 5-7.

3 Pixel Classification

We use Bayesian classifiers to create subjective class definitions that are described in terms of easily computable objective attributes such as spectral values, texture, and ancillary data [13]. The Bayesian framework is a probabilistic tool to combine information from multiple sources in terms of conditional and prior probabilities. Assume there are k class labels, $w_1, \ldots, w_k$, defined by the user. Let $x_1, \ldots, x_m$ be the attributes computed for a pixel. The goal is to find the most probable label for that pixel given a particular set of values of these attributes. The degree of association between the pixel and class w_j can be computed using the posterior probability

$$\begin{aligned} & p(w_j|x_1, \ldots, x_m) \\ & \quad = \frac{p(x_1, \ldots, x_m|w_j)p(w_j)}{p(x_1, \ldots, x_m)} \\ & \quad = \frac{p(x_1, \ldots, x_m|w_j)p(w_j)}{p(x_1, \ldots, x_m|w_j)p(w_j) + p(x_1, \ldots, x_m|\neg w_j)p(\neg w_j)} \\ & \quad = \frac{p(w_j)\prod_{i=1}^{m} p(x_i|w_j)}{p(w_j)\prod_{i=1}^{m} p(x_i|w_j) + p(\neg w_j)\prod_{i=1}^{m} p(x_i|\neg w_j)} \end{aligned} \tag{2}$$

under the conditional independence assumption. The conditional independence assumption simplifies learning because the parameters for each attribute model $p(x_i|w_j)$ can be estimated separately. Therefore, user interaction is only required for the labeling of pixels as positive (w_j) or negative ($\neg w_j$) examples for a particular class under training. Models for different classes are learned separately from the corresponding positive and negative examples. Then, the predicted class becomes the one with the largest posterior probability and the pixel is assigned the class label

$$w_j^* = \arg \max_{j=1,\ldots,k} p(w_j|x_1, \ldots, x_m). \tag{3}$$

We use discrete variables and a non-parametric model in the Bayesian framework where continuous features are converted to discrete attribute values using the unsupervised k-means clustering algorithm for vector quantization. The number of clusters (quantization levels) is empirically chosen for each feature. (An alternative is to use a parametric distribution assumption, e.g., Gaussian, for each individual continuous feature but these parametric assumptions do not always hold.) Schröder *et al.* [19] used similar classifiers to retrieve images from remote sensing archives by approximating the probabilities of images belonging to different classes using pixel level probabilities. In the following, we describe learning of the models for $p(x_i|w_j)$ using the positive training examples for the j'th class label. Learning of $p(x_i|\neg w_j)$ is done the same way using the negative examples.

For a particular class, let each discrete variable x_i have r_i possible values (states) with probabilities

$$p(x_i = z|\boldsymbol{\theta_i}) = \boldsymbol{\theta_{iz}} > 0 \tag{4}$$

where $z \in \{1, \ldots, r_i\}$ and $\boldsymbol{\theta_i} = \{\boldsymbol{\theta_{iz}}\}_{z=1}^{r_i}$ is the set of parameters for the i'th attribute model. This corresponds to a multinomial distribution. Since maximum likelihood estimates can give unreliable

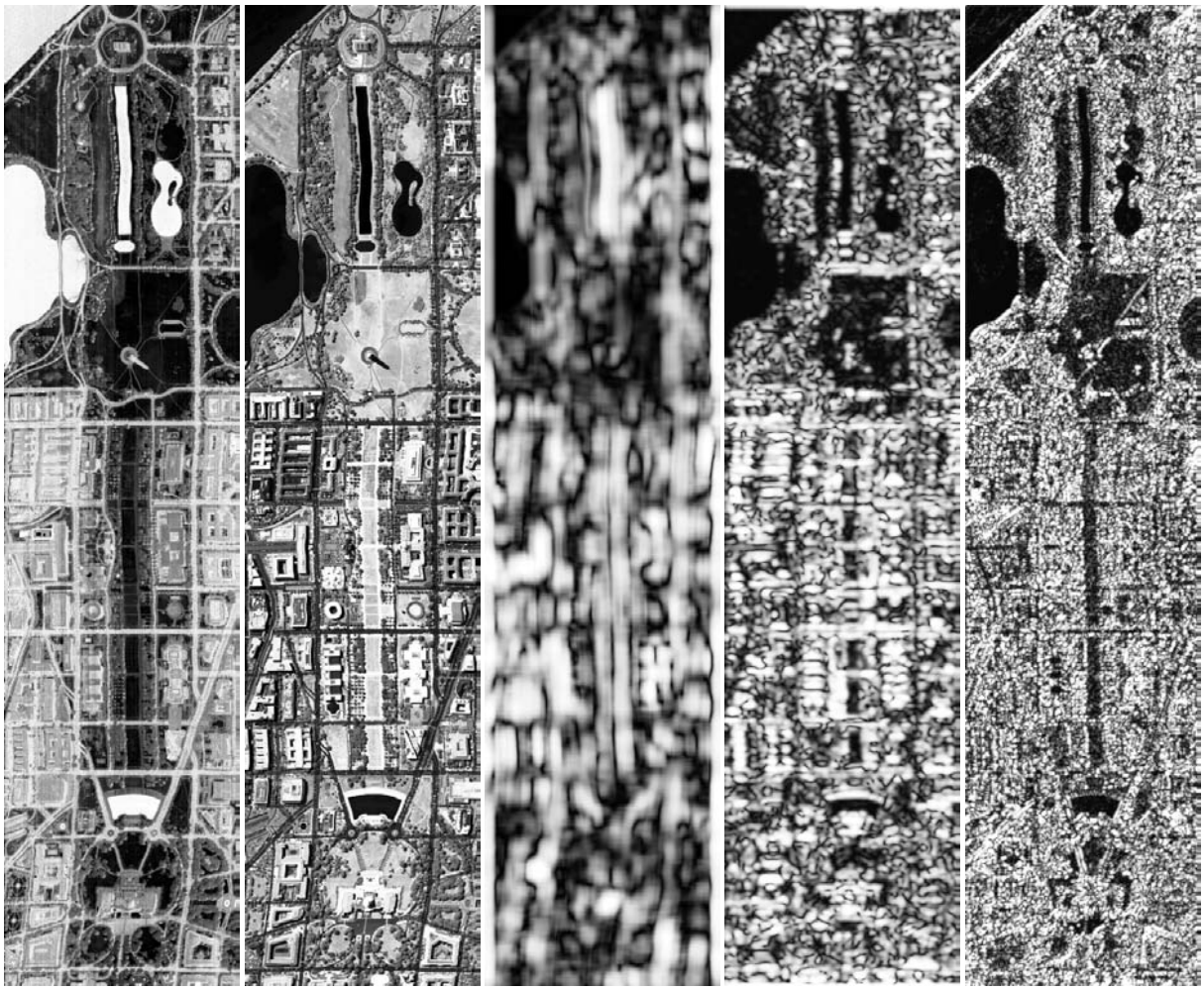


Figure 5: Pixel feature examples for the *DC Mall* data set. From left to right: the first LDA band, the first PCA band, Gabor features for 90 degree orientation at the first scale, Gabor features for 0 degree orientation at the third scale, and Gabor features for 45 degree orientation at the fourth scale. Histogram equalization was applied to all images for better visualization.

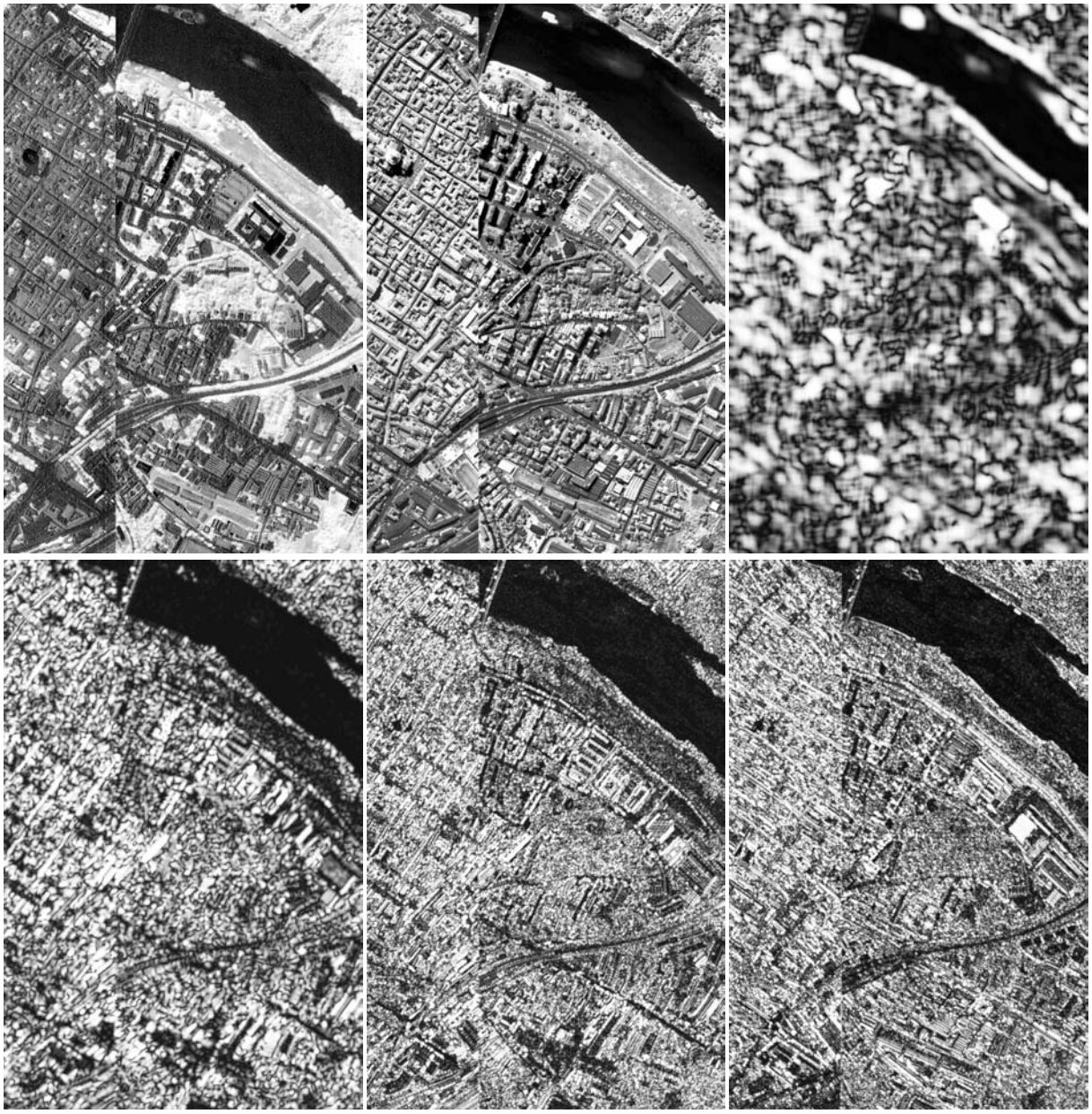


Figure 6: Pixel feature examples for the *Centre* data set. From left to right, first row: the first LDA band, the first PCA band, Gabor features for 135 degree orientation at the first scale; second row: Gabor features for 45 degree orientation at the third scale, Gabor features for 45 degree orientation at the fourth scale, and Gabor features for 135 degree orientation at the fourth scale. Histogram equalization was applied to all images for better visualization.

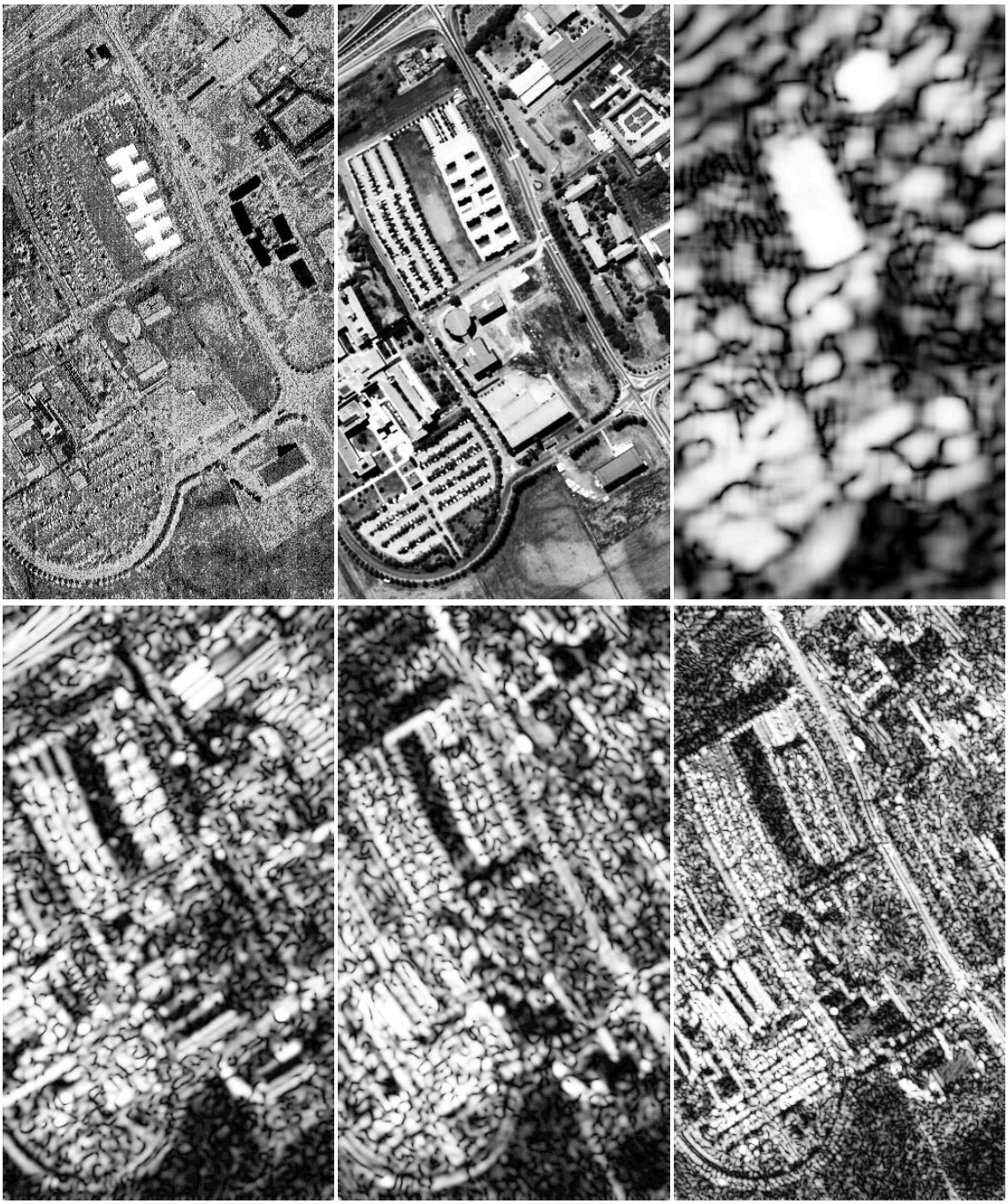


Figure 7: Pixel feature examples for the *University* data set. From left to right, first row: the first LDA band, the first PCA band, Gabor features for 45 degree orientation at the first scale; second row: Gabor features for 45 degree orientation at the third scale, Gabor features for 135 degree orientation at the third scale, and Gabor features for 135 degree orientation at the fourth scale. Histogram equalization was applied to all images for better visualization.

results when the sample is small and the number of parameters is large, we use the Bayes estimate of $\boldsymbol{\theta_{iz}}$ that can be computed as the expected value of the posterior distribution.

We can choose any prior for $\boldsymbol{\theta_i}$ in the computation of the posterior distribution but there is a big advantage to use conjugate priors. A conjugate prior is one which, when multiplied with the direct probability, gives a posterior probability having the same functional form as the prior, thus allowing the posterior to be used as a prior in further computations [20]. The conjugate prior for the multinomial distribution is the Dirichlet distribution [21]. Geiger and Heckerman [22] showed that if all allowed states of the variables are possible (i.e., $\boldsymbol{\theta_{iz}} > 0$) and if certain parameter independence assumptions hold, then a Dirichlet distribution is indeed the only possible choice for the prior.

Given the Dirichlet prior $p(\boldsymbol{\theta_i}) = \mathrm{Dir}(\boldsymbol{\theta_i}|\alpha_{i1}, \ldots, \alpha_{ir_i})$ where α_{iz} are positive constants, the posterior distribution of $\boldsymbol{\theta_i}$ can be computed using the Bayes rule as

$$\begin{aligned} p(\boldsymbol{\theta_i}|\mathcal{D}) &= \frac{p(\mathcal{D}|\boldsymbol{\theta_i})p(\boldsymbol{\theta_i})}{p(\mathcal{D})} \\ &= \mathrm{Dir}(\boldsymbol{\theta_i}|\alpha_{i1} + N_{i1}, \ldots, \alpha_{ir_i} + N_{ir_i}) \end{aligned} \tag{5}$$

where $\mathcal{D}$ is the training sample and N_{iz} is the number of cases in $\mathcal{D}$ in which $x_i = z$. Then, the Bayes estimate for $\boldsymbol{\theta_{iz}}$ can be found by taking the conditional expected value

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{iz}} = E_{p(\boldsymbol{\theta_i}|\mathcal{D})}[\boldsymbol{\theta_{iz}}] = \frac{\alpha_{iz} + N_{iz}}{\alpha_i + N_i} \tag{6}$$

where $\alpha_i = \sum_{z=1}^{r_i} \alpha_{iz}$ and $N_i = \sum_{z=1}^{r_i} N_{iz}$.

An intuitive choice for the hyper-parameters $\alpha_{i1}, \ldots, \alpha_{ir_i}$ of the Dirichlet distribution is the Laplace's uniform prior [23] that assumes all r_i states to be equally probable ($\alpha_{iz} = 1, \forall z \in \{1, \ldots, r_i\}$) which results in the Bayes estimate

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{iz}} = \frac{1 + N_{iz}}{r_i + N_i}. \tag{7}$$

Laplace's prior is regarded to be a safe choice when the distribution of the source is unknown and the number of possible states r_i is fixed and known [24].

Given the current state of the classifier that was trained using the prior information and the sample $\mathcal{D}$, we can easily update the parameters when new data $\mathcal{D}'$ is available. The new posterior distribution for $\boldsymbol{\theta_i}$ becomes

$$p(\boldsymbol{\theta_i}|\mathcal{D}, \mathcal{D}') = \frac{p(\mathcal{D}'|\boldsymbol{\theta_i})p(\boldsymbol{\theta_i}|\mathcal{D})}{p(\mathcal{D}'|\mathcal{D})}. \tag{8}$$

With the Dirichlet priors and the posterior distribution for $p(\boldsymbol{\theta_i}|\mathcal{D})$ given in (5), the updated posterior distribution becomes

$$p(\boldsymbol{\theta_i}|\mathcal{D}, \mathcal{D}') = \mathrm{Dir}(\boldsymbol{\theta_i}|\alpha_{i1} + N_{i1} + N'_{i1}, \ldots, \alpha_{ir_i} + N_{ir_i} + N'_{ir_i}) \tag{9}$$

where N'_{iz} is the number of cases in $\mathcal{D}'$ in which $x_i = z$. Hence, updating the classifier parameters involves only updating the counts in the estimates for $\hat{\boldsymbol{\theta}}_{\boldsymbol{iz}}$.

The Bayesian classifiers that are learned from examples as described above are used to compute probability maps for all land cover/use classes and assign each pixel to one of these classes using the maximum a posteriori probability (MAP) rule given in (3). Example probability maps are shown in Figures 8-10.

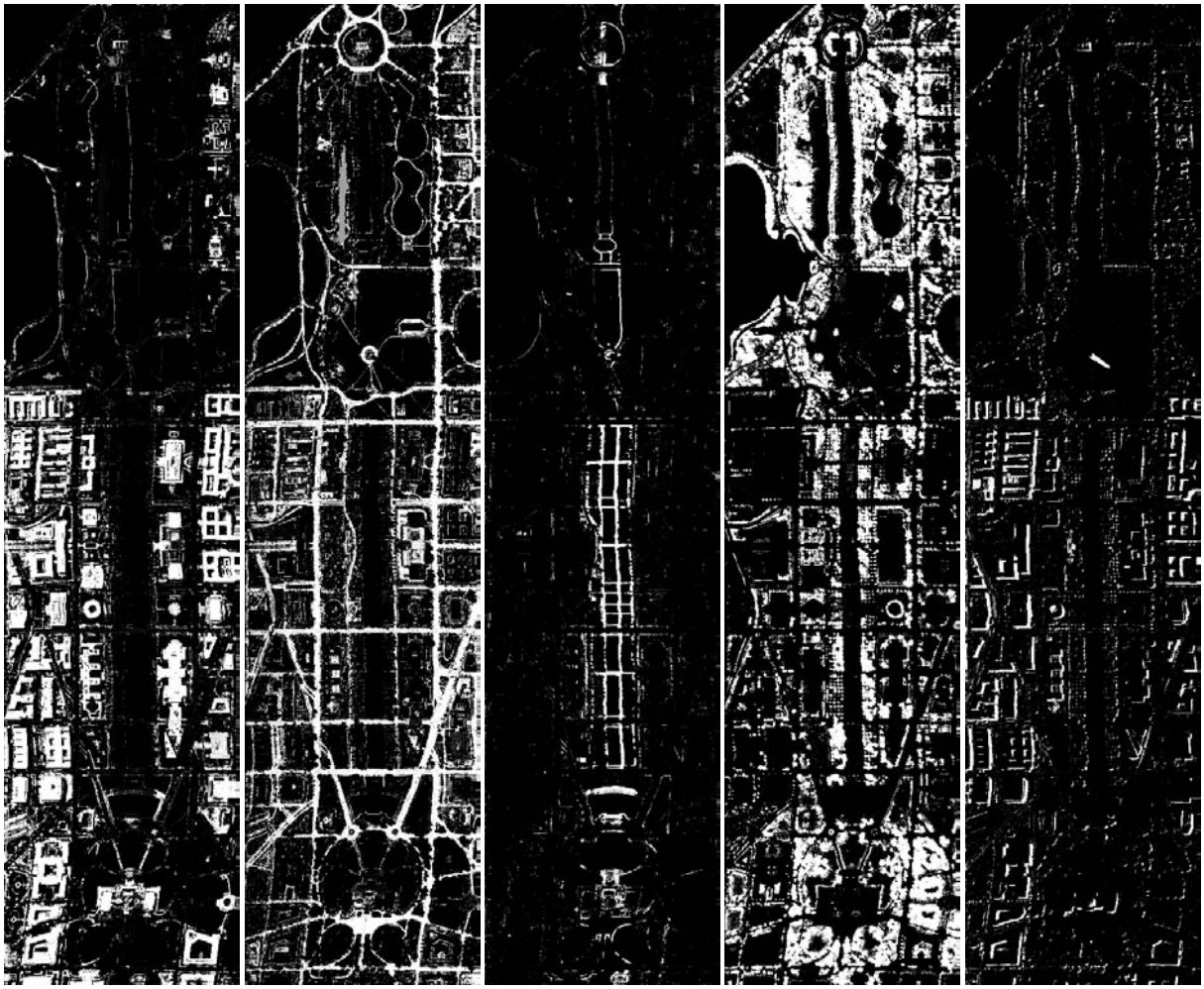


Figure 8: Pixel level probability maps for different classes of the *DC Mall* data set. From left to right: roof, street, path, trees, shadow. Brighter values in the map show pixels with high probability of belonging to that class.

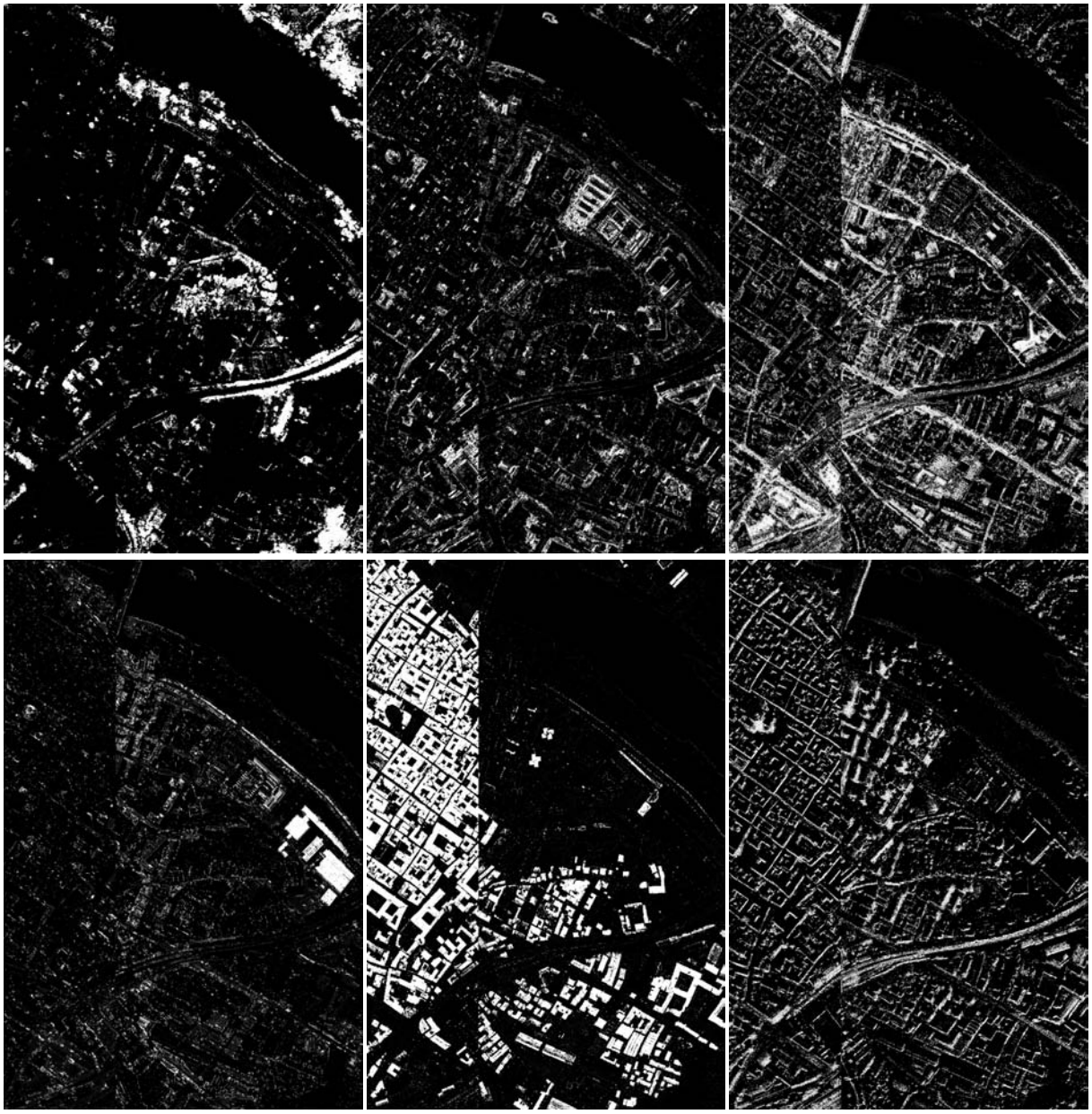


Figure 9: Pixel level probability maps for different classes of the *Centre* data set. From left to right, first row: trees, self-blocking bricks, asphalt; second row: bitumen, tiles, shadow. Brighter values in the map show pixels with high probability of belonging to that class.

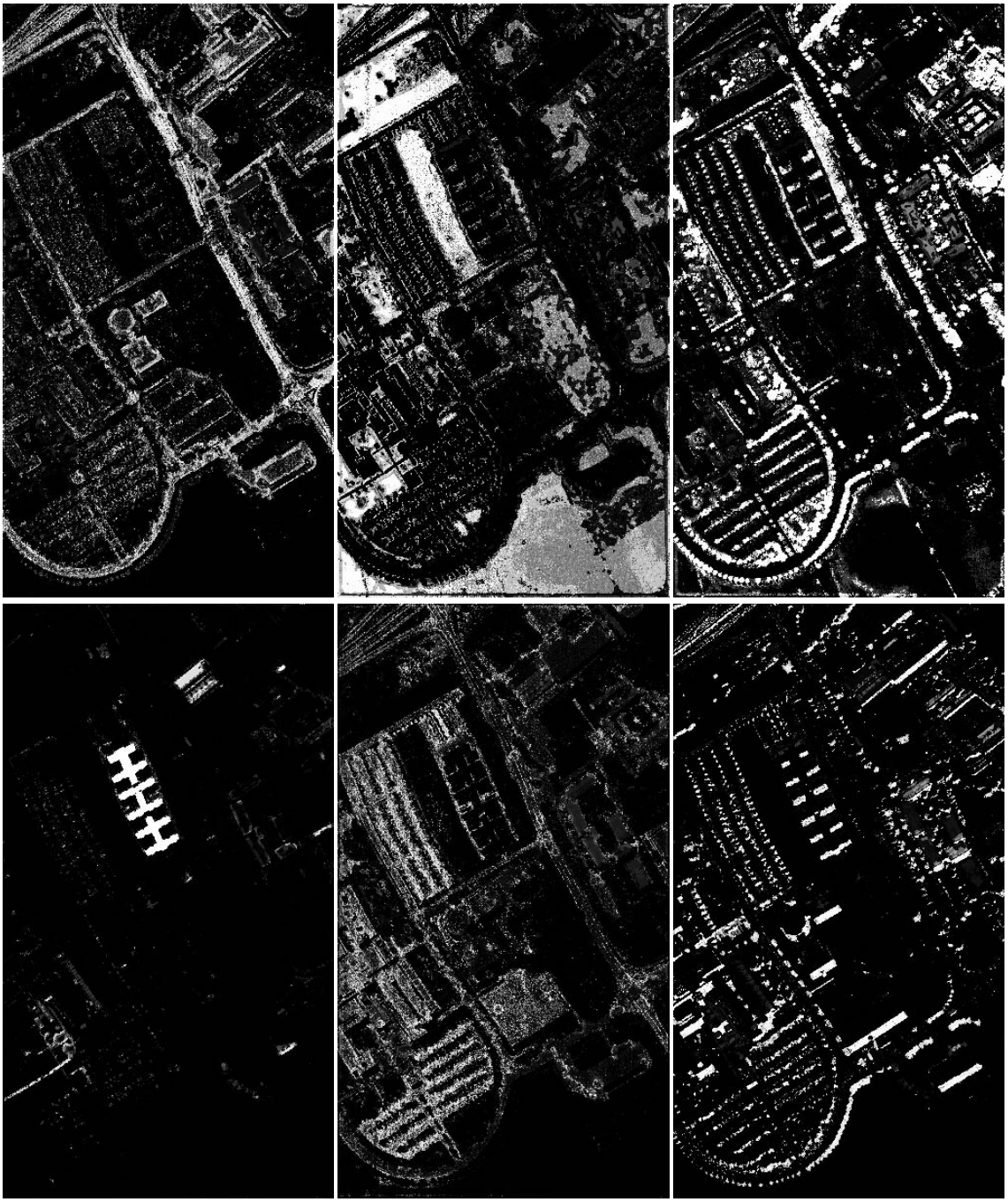


Figure 10: Pixel level probability maps for different classes of the *University* data set. From left to right, first row: asphalt, meadows, trees; second row: metal sheets, self-blocking bricks, shadow. Brighter values in the map show pixels with high probability of belonging to that class.

4 Region Segmentation

Image segmentation is used to group pixels that belong to the same structure with the goal of delineating each individual structure as an individual region. In previous work [25], we used an automatic segmentation algorithm that breaks an image into many small regions and merges them by minimizing an energy functional that trades off the similarity of regions against the length of their shared boundaries. We have also recently experimented with several segmentation algorithms from the computer vision literature. Algorithms that are based on graph clustering [26], mode seeking [27] and classification [28] have been reported to be successful in moderately sized color images with relatively homogeneous structures. However, we could not apply these techniques successfully to our data sets because the huge amount of data in hyper-spectral images made processing infeasible due to both memory and computational requirements, and the detailed structure in high-resolution remotely sensed imagery prevented the use of sampling that has been often used to reduce the computational requirements of these techniques.

The segmentation approach we have used in this work consists of smoothing filters and mathematical morphology. The input to the algorithm includes the probability maps for all classes where each pixel is assigned either to one of these classes or to the reject class for probabilities smaller than a threshold (latter type of pixels are initially marked as background). Since pixel-based classification ignores spatial correlations, the initial segmentation may contain isolated pixels with labels different from those of their neighbors. We use an iterative split-and-merge algorithm [13] to convert this intermediate step into contiguous regions as follows:

1. Merge pixels with identical class labels to find the initial set of regions and mark these regions as foreground,
2. Mark regions with areas smaller than a threshold as background using connected components analysis [5],
3. Use region growing to iteratively assign background pixels to the foreground regions by placing a window at each background pixel and assigning it to the class that occurs the most in its neighborhood.

This procedure corresponds to a spatial smoothing of the clustering results. We further process the resulting regions using mathematical morphology operators [5] to automatically divide large regions into more compact sub-regions as follows [13]:

1. Find individual regions using connected components analysis for each class,
2. For all regions, compute the erosion transform [5] and repeat:
   (a) Threshold erosion transform at steps of 3 pixels in every iteration,
   (b) Find connected components of the thresholded image,
   (c) Select sub-regions that have an area smaller than a threshold,
   (d) Dilate these sub-regions to restore the effects of erosion,
   (e) Mark these sub-regions in the output image by masking the dilation using the original image,

   until no more sub-regions are found,

3. Merge the residues of previous iterations to their smallest neighbors.

The merging and splitting process is illustrated in Figure 11. The probability of each region belonging to a land cover/use class can be estimated by propagating class labels from pixels to regions. Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be the set of pixels that are merged to form a region. Let w_j and $p(w_j|x_i)$ be the class label and its posterior probability, respectively, assigned to pixel x_i by the classifier. The probability $p(w_j|x \in \mathcal{X})$ that a pixel in the merged region belongs to the class w_j can be computed as

$$\begin{aligned} & p(w_j|x \in \mathcal{X}) \\ & \quad = \frac{p(w_j, x \in \mathcal{X})}{p(x \in \mathcal{X})} = \frac{p(w_j, x \in \mathcal{X})}{\sum_{t=1}^{k} p(w_t, x \in \mathcal{X})} \\ & \quad = \frac{\sum_{x \in \mathcal{X}} p(w_j, x)}{\sum_{t=1}^{k} \sum_{x \in \mathcal{X}} p(w_t, x)} = \frac{\sum_{x \in \mathcal{X}} p(w_j|x)p(x)}{\sum_{t=1}^{k} \sum_{x \in \mathcal{X}} p(w_t|x)p(x)} \\ & \quad = \frac{E_x\{\mathbb{I}_{x \in \mathcal{X}}(x)p(w_j|x)\}}{\sum_{t=1}^{k} E_x\{\mathbb{I}_{x \in \mathcal{X}}(x)p(w_t|x)\}} = \frac{1}{n}\sum_{i=1}^{n} p(w_j|x_i) \end{aligned} \tag{10}$$

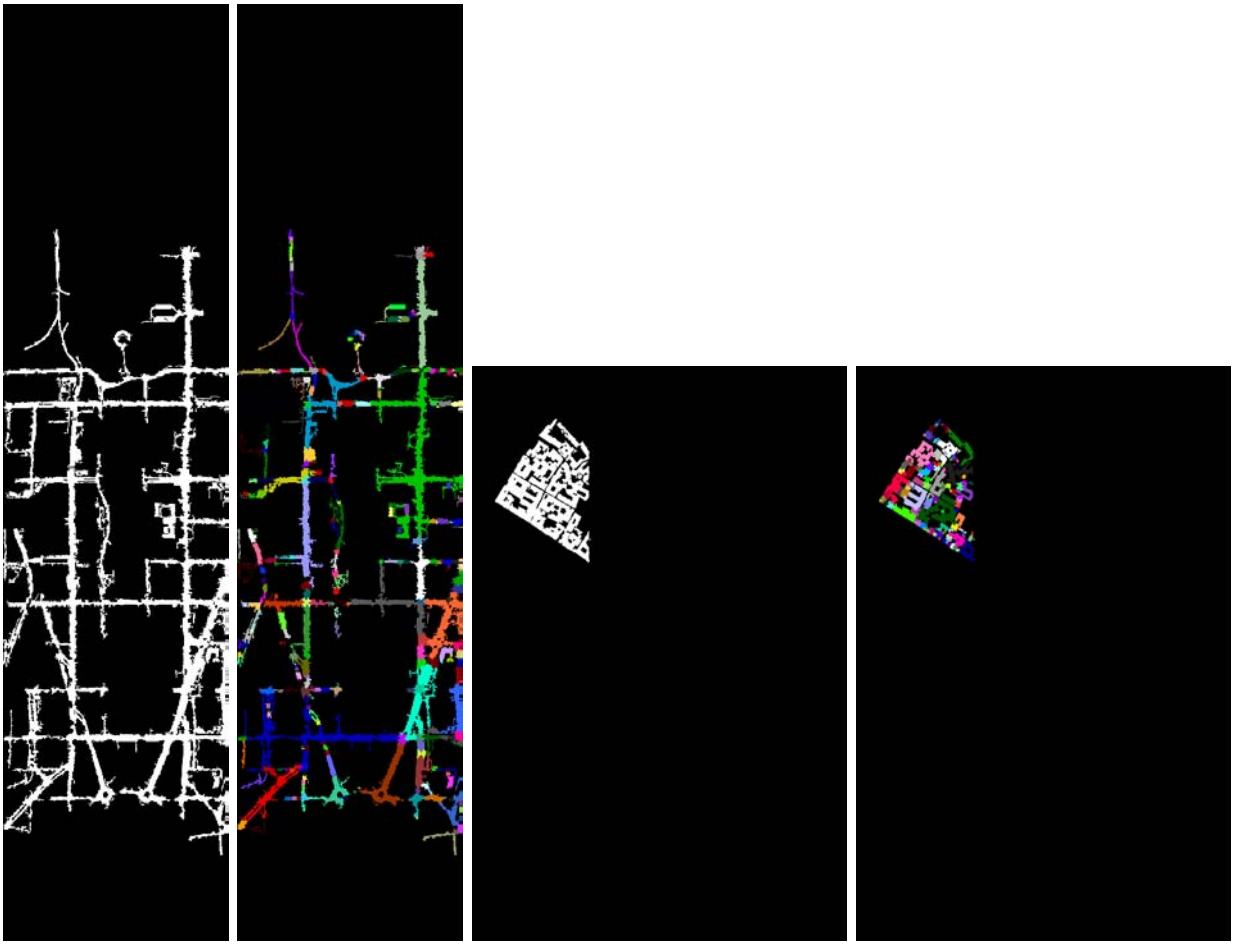
where $\mathbb{I}_A(\cdot)$ is the indicator function associated with the set A. Each region in the final segmentation are assigned labels with probabilities using (10).

5 Region Feature Extraction

Region level representations include properties shared by groups of pixels obtained through region segmentation. The regions are modeled using the statistical summaries of their spectral and textural properties along with shape features that are computed from region polygon boundaries. The statistical summary for a region is computed as the means and standard deviations of features of the pixels in that region. Multi-dimensional histograms also provide pixel feature distributions within individual regions. The shape properties [5] of a region correspond to its

- area,
- orientation of the region's major axis with respect to the x axis,
- eccentricity (ratio of the distance between the foci to the length of the major axis; e.g., a circle is an ellipse with zero eccentricity),
- Euler number (1 minus the number of holes in the region),
- solidity (ratio of the area to the convex area),
- extent (ratio of the area to the area of the bounding box),
- spatial variances along the x and y axes, and
- spatial variances along the region's principal (major and minor) axes,

resulting in a feature vector of length 10.



(a) A large connected region formed by merging pixels labeled as street in *DC Mall* data

(b) More compact sub-regions after splitting the region in (a)

(c) A large connected region formed by merging pixels labeled as tiles in *Centre* data

(d) More compact sub-regions after splitting the region in (c)

Figure 11: Examples for the region segmentation process. The iterative algorithm that uses mathematical morphology operators is used to split a large connected region into more compact sub-regions.

6 Region Classification

In the remote sensing literature, image classification is usually done by using pixel features as input to classifiers such as minimum distance, maximum likelihood, neural networks or decision trees. However, large within-class variations and small between-class variations of these features at the pixel level and the lack of spatial information limit the accuracy of these classifiers.

In this work, we perform final classification using region level information. To be able to use the Bayesian classifiers that were described in Section 3, different region-based features such as statistics and shape features are independently converted to discrete random variables using the k-means algorithm for vector quantization. In particular, for each region, we obtain 4 values from

- clustering of the statistics of the LDA bands (6 bands for *DC Mall* data, 8 bands for *Centre* and *University* data),
- clustering of the statistics of the 10 PCA bands,
- clustering of the statistics of the 16 Gabor bands,
- clustering of the 10 shape features.

In the next section, we evaluate the performance of these new features for classifying regions (and the corresponding pixels) into land cover/use categories defined by the user.

7 Experiments

Performances of the features and the algorithms described in the previous sections were evaluated both quantitatively and qualitatively. First, pixel level features (LDA, PCA and Gabor) were extracted and normalized for all three data sets as described in Section 2. The ground truth maps shown in Figures 1-3 were used to divide the data into independent training and test sets. Then, the k-means algorithm was used to cluster (quantize) the continuous features and convert them to discrete attribute values, and Bayesian classifiers with discrete non-parametric models were trained using these attributes and the training examples as described in Section 3. The value of k was set to 25 empirically for all data sets. Example probability maps for some of the classes were given in Figures 8-10. Confusion matrices, shown in Tables 1-3, were computed using the test ground truth for all data sets.

Next, the iterative split-and-merge algorithm described in Section 4 was used to convert the pixel level classification results into contiguous regions. The neighborhood size for region growing was set to 3×3. The minimum area threshold in the segmentation process was set to 5 pixels. After the region level features (LDA, PCA and Gabor statistics, and shape features) were computed and normalized for all resulting regions as described in Section 5, they were also clustered (quantized) and converted to discrete values. The value of k was set to 25 again for all data sets. Then, Bayesian classifiers were trained using the training ground truth as described in Section 6, and were applied to the test data to produce the confusion matrices shown in Tables 4-6.

Finally, comparative experiments were done by training and evaluating traditional maximum likelihood classifiers with the multivariate Gaussian with full covariance matrix assumption for each class (quadratic Gaussian classifier) using the same training and test ground truth data. The classification performances of all three classifiers (pixel level Bayesian, region level Bayesian,

Table 1: Confusion matrix for pixel level classification of the *DC Mall* data set (testing subset) using LDA, PCA and Gabor features.

		Assigned							Total	% Agree
		roof	street	path	grass	trees	water	shadow		
True	roof	3771	49	12	0	1	0	1	3834	98.3568
	street	0	412	0	0	0	0	4	416	99.0385
	path	0	0	175	0	0	0	0	175	100.0000
	grass	0	0	0	1926	2	0	0	1928	99.8963
	trees	0	0	0	0	405	0	0	405	100.0000
	water	0	0	0	0	0	1223	1	1224	99.9183
	shadow	0	4	0	0	0	0	93	97	95.8763
Total		3771	465	187	1926	408	1223	99	8079	99.0840

Table 2: Confusion matrix for pixel level classification of the *Centre* data set (testing subset) using LDA, PCA and Gabor features.

		Assigned									Total	% Agree
		water	trees	meadows	bricks	bare soil	asphalt	bitumen	tiles	shadow		
True	water	65877	0	1	0	1	7	0	0	85	65971	99.8575
	trees	1	6420	1094	5	0	45	4	0	29	7598	84.4959
	meadows	0	349	2718	0	22	1	0	0	0	3090	87.9612
	bricks	0	0	0	2238	221	139	87	0	0	2685	83.3520
	bare soil	0	9	110	1026	5186	191	59	3	0	6584	78.7667
	asphalt	4	0	0	317	30	7897	239	5	756	9248	85.3914
	bitumen	4	0	1	253	22	884	6061	9	53	7287	83.1755
	tiles	0	1	0	150	85	437	116	41826	211	42826	97.6650
	shadow	12	0	0	3	0	477	0	0	2371	2863	82.8152
Total		65898	6779	3924	3992	5567	10078	6566	41843	3505	148152	94.8985

Table 3: Confusion matrix for pixel level classification of the *University* data set (testing subset) using LDA, PCA and Gabor features.

		Assigned									Total	% Agree
		asphalt	meadows	gravel	trees	metal sheets	bare soil	bitumen	bricks	shadow		
True	asphalt	4045	38	391	39	1	105	1050	875	87	6631	61.0014
	meadows	21	14708	14	691	0	3132	11	71	1	18649	78.8675
	gravel	91	14	1466	0	0	3	19	506	0	2099	69.8428
	trees	5	76	1	2927	0	40	1	2	12	3064	95.5287
	metal sheets	0	2	0	1	1341	0	0	1	0	1345	99.7026
	bare soil	34	1032	7	38	20	3745	32	119	2	5029	74.4681
	bitumen	424	1	7	1	0	1	829	67	0	1330	62.3308
	bricks	382	45	959	2	1	87	141	2064	1	3682	56.0565
	shadow	22	0	0	0	0	0	0	2	923	947	97.4657
Total		5024	15916	2845	3699	1363	7113	2083	3707	1026	42776	74.9205

Table 4: Confusion matrix for region level classification of the *DC Mall* data set (testing subset) using LDA, PCA and Gabor statistics, and shape features.

		Assigned							Total	% Agree
		roof	street	path	grass	trees	water	shadow		
True	roof	3814	11	5	0	0	1	3	3834	99.4784
	street	0	414	0	0	0	0	2	416	99.5192
	path	0	0	175	0	0	0	0	175	100.0000
	grass	0	0	0	1928	0	0	0	1928	100.0000
	trees	0	0	0	0	405	0	0	405	100.0000
	water	0	1	0	0	0	1223	0	1224	99.9183
	shadow	1	2	0	0	0	0	94	97	96.9072
Total		3815	428	180	1928	405	1224	99	8079	99.6782

Table 5: Confusion matrix for region level classification of the *Centre* data set (testing subset) using LDA, PCA and Gabor statistics, and shape features.

		Assigned									Total	% Agree
		water	trees	meadows	bricks	bare soil	asphalt	bitumen	tiles	shadow		
True	water	65803	0	0	0	0	0	0	0	168	65971	99.7453
	trees	0	6209	1282	28	22	11	5	0	41	7598	81.7189
	meadows	0	138	2942	0	10	0	0	0	0	3090	95.2104
	bricks	0	0	1	2247	173	31	233	0	0	2685	83.6872
	bare soil	1	4	59	257	6139	11	102	0	11	6584	93.2412
	asphalt	0	1	2	37	4	8669	163	0	372	9248	93.7392
	bitumen	0	0	0	24	3	726	6506	0	28	7287	89.2823
	tiles	0	0	0	39	13	220	2	42380	172	42826	98.9586
	shadow	38	0	2	2	0	341	12	0	2468	2863	86.2033
Total		65842	6352	4288	2634	6364	10009	7023	42380	3260	148152	96.7675

Table 6: Confusion matrix for region level classification of the *University* data set (testing subset) using LDA, PCA and Gabor statistics, and shape features.

		Assigned									Total	% Agree
		asphalt	meadows	gravel	trees	metal sheets	bare soil	bitumen	bricks	shadow		
True	asphalt	4620	7	281	4	0	52	344	1171	152	6631	69.6727
	meadows	8	17246	0	1242	0	19	6	7	121	18649	92.4768
	gravel	9	5	1360	2	0	0	0	723	0	2099	64.7928
	trees	39	37	0	2941	0	4	13	14	16	3064	95.9856
	metal sheets	0	0	0	0	1344	0	0	1	0	1345	99.9257
	bare soil	0	991	0	5	0	4014	0	19	0	5029	79.8171
	bitumen	162	0	0	0	0	0	1033	135	0	1330	77.6692
	bricks	248	13	596	33	5	21	125	2635	6	3682	71.5644
	shadow	16	0	0	0	1	0	0	1	929	947	98.0993
Total		5102	18299	2237	4227	1350	4110	1521	4706	1224	42776	84.4445

Table 7: Summary of classification accuracies using the pixel level and region level Bayesian classifiers and the quadratic Gaussian classifier.

	DC Mall	*Centre*	*University*
Pixel level Bayesian	99.0840	94.8985	74.9205
Region level Bayesian	99.6782	96.7675	84.4445
Quadratic Gaussian	99.3811	93.9677	81.2792

quadratic Gaussian) are summarized in Table 7. For qualitative comparison, the classification maps for all classifiers for all data sets were computed as shown in Figures 12-14.

The results show that the proposed region level features and Bayesian classifiers performed better than the traditional maximum likelihood classifier with the Gaussian density assumption for all data sets with respect to the ground truth maps available. Using texture features, that model spatial neighborhoods of pixels, in addition to the spectral-based ones improved the performances of all classifiers. Using the Gabor filters at the third and fourth scales (corresponding to 8 features) improved the results the most. (The confusion matrices presented show the performances of using these features instead of the original 16.) The reason for this is the high spatial image resolution where filters with a larger coverage include mixed effects from multiple structures within a pixel's neighborhood.

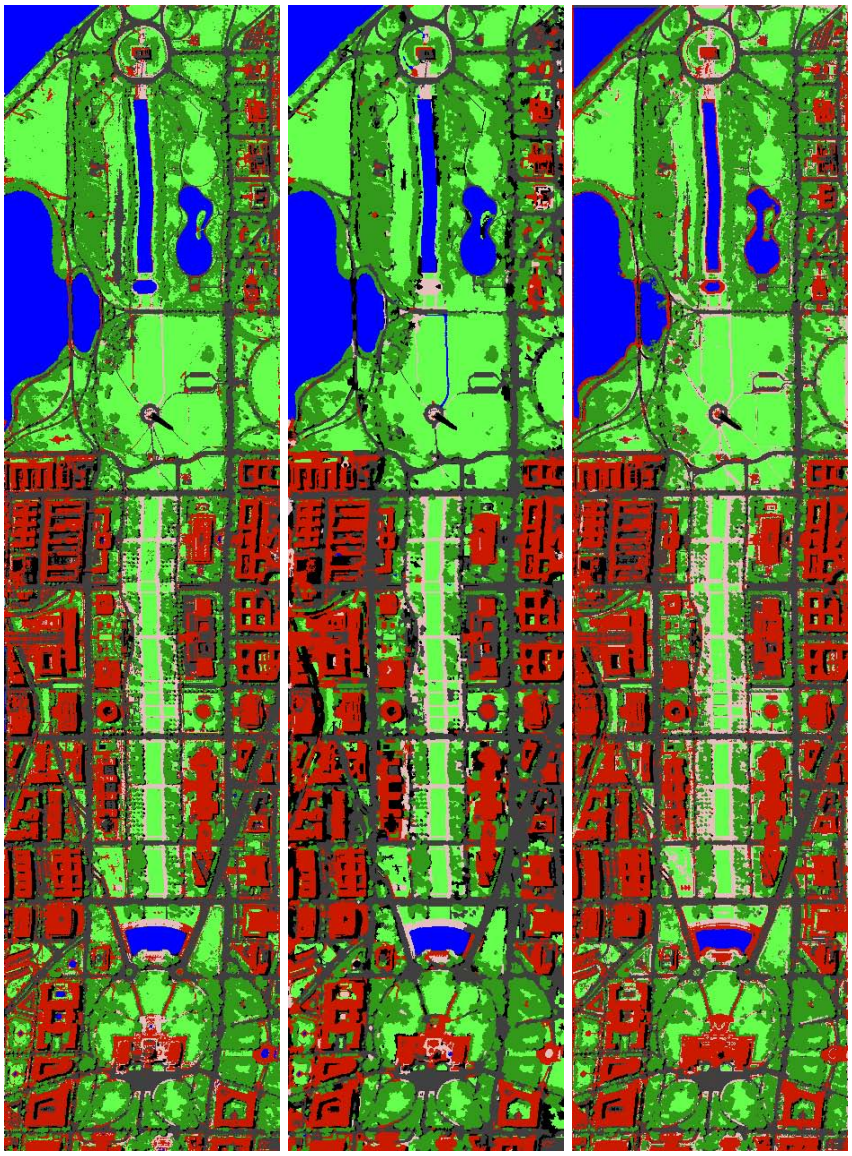
Using region level information gave the most significant improvement for the *University* data set. The performances of pixel level classifiers for *DC Mall* and *Centre* data sets using LDA- and PCA-based spectral and Gabor-based textural features were already quite high. In all cases, region level classification performed better than pixel level classifiers.

One important observation to note is that even though the accuracies of all classifiers look quite high, some misclassified areas can still be found in the classification maps for all images. This is especially apparent in the results of pixel level classifiers where many isolated pixels that are not covered by test ground truth maps (e.g., the upper part of the *DC Mall* data, tiles on the left of the *Centre* data, many areas in the *University* data) were assigned wrong class labels because of the lack of spatial information and, hence, the context. The same phenomenon can be observed in many other results published in the literature. A more detailed ground truth is necessary for a more reliable evaluation of classifiers for high-resolution imagery. We believe that there is still a large margin for improvement in the performance of classification techniques for data received from state-of-the-art satellites.

8 Conclusions

We have presented an approach for classification of remotely sensed imagery using spatial techniques. First, pixel level spectral and textural features were extracted and used for classification with non-parametric Bayesian classifiers. Next, an iterative split-and-merge algorithm was used to convert the pixel level classification maps into contiguous regions. Then, spectral and textural statistics and shape features extracted from these regions were used with similar Bayesian classifiers to compute the final classification maps.

Comparative quantitative and qualitative evaluation using traditional maximum likelihood Gaussian classifiers in experiments with three different data sets with ground truth showed that the proposed region level features and Bayesian classifiers performed better than the traditional pixel



(a) Pixel level Bayesian

(b) Region level Bayesian

(c) Quadratic Gaussian

Figure 12: Final classification maps with the Bayesian pixel and region level classifiers and the quadratic Gaussian classifier for the *DC Mall* data set. Class color codes were listed in Figure 1.

(a) Pixel level Bayesian (b) Region level Bayesian (c) Quadratic Gaussian

Figure 13: Final classification maps with the Bayesian pixel and region level classifiers and the quadratic Gaussian classifier for the *Centre* data set. Class color codes were listed in Figure 2.

(a) Pixel level Bayesian (b) Region level Bayesian (c) Quadratic Gaussian

Figure 14: Final classification maps with the Bayesian pixel and region level classifiers and the quadratic Gaussian classifier for the *University* data set. Class color codes were listed in Figure 3.

level classification techniques. Even though the numerical results already look quite impressive, we believe that selection of the most discriminative subset of features and better segmentation of regions will bring further improvements in classification accuracy. We are also in the process of gathering ground truth data with a larger coverage for better evaluation of classification techniques for images from high-resolution satellites.

Acknowledgment

The author would like to thank Dr. David A. Landgrebe and Mr. Larry L. Biehl from Purdue University, Indiana, U.S.A., for the *DC Mall* data set, and Dr. Paolo Gamba from the University of Pavia, Italy, for the *Centre* and *University* data sets.

References

[1] S. S. Durbha and R. L. King. Knowledge mining in earth observation data archives: a domain ontology perspective. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 1, September 2004.

[2] D. A. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons, Inc., 2003.

[3] G. G. Wilkinson. Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):433–440, March 2005.

[4] M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P. G. Marchetti, and S. D'Elia. Information mining in remote sensing image archives: system concepts. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2923–2936, December 2003.

[5] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.

[6] R. L. Kettig and D. A. Landgrebe. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, GE-14(1):19–26, January 1976.

[7] C. Evans, R. Jones, I. Svalbe, and M. Berman. Segmenting multispectral Landsat TM images into field units. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1054–1064, May 2002.

[8] A. Sarkar, M. K. Biswas, B. Kartikeyan, V. Kumar, K. L. Majumder, and D. K. Pal. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1102–1113, May 2002.

[9] J. C. Tilton, G. Marchisio, K. Koperski, and M. Datcu. Image information mining utilizing hierarchical segmentation. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1029–1031, Toronto, Canada, June 2002.

[10] G. G. Hazel. Object-level change detection in spectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):553–561, March 2001.

[11] T. Blaschke. Object-based contextual image classification built on image segmentation. In *Proceedings of IEEE GRSS Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pages 113–119, Washington, DC, October 2003.

[12] A. Rydberg and G. Borgefors. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11):2514–2520, November 2001.

[13] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton. Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):581–589, March 2005.

[14] S. Aksoy and H. G. Akcay. Multi-resolution segmentation and shape analysis for remote sensing image classification. In *Proceedings of 2nd International Conference on Recent Advances in Space Technologies*, Istanbul, Turkey, June 9-11 2005.

[15] S. Aksoy, K. Koperski, C. Tusk, and G. Marchisio. Interactive training of advanced classifiers for mining remote sensing image archives. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 773–782, Seattle, WA, August 22–25 2004.

[16] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2000.

[17] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[18] S. Aksoy and R. M. Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5):563–582, May 2001.

[19] M. Schroder, H. Rehrauer, K. Siedel, and M. Datcu. Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2288–2298, September 2000.

[20] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[21] M. H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, 1970.

[22] D. Geiger and D. Heckerman. A characterization of the Dirichlet distribution through global and local parameter independence. *The Annals of Statistics*, 25(3):1344–1369, 1997. MSR-TR-94-16.

[23] T. M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[24] R. F. Krichevskiy. Laplace's law of succession and universal encoding. *IEEE Transactions on Information Theory*, 44(1):296–303, January 1998.

[25] S. Aksoy, C. Tusk, K. Koperski, and G. Marchisio. Scene modeling and image mining with a visual grammar. In C. H. Chen, editor, *Frontiers of Remote Sensing Information Processing*, pages 35–62. World Scientific, 2003.

[26] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000.

[27] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.

[28] P. Paclik, R. P. W. Duin, G. M. P. van Kempen, and R. Kohlus. Segmentation of multi-spectral images using the combined classifier approach. *Image and Vision Computing*, 21(6):473–482, June 2003.