COMPOUND OBJECT DETECTION USING REGION CO-OCCURRENCE STATISTICS

Selim Aksoy

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

Krzysztof Koperski, Carsten Tusk, Giovanni Marchisio

DigitalGlobe, Inc.
Research and Development
1601 Dry Creek Drive, Longmont, CO 80501, USA
{kkopersk,ctusk,gmarchis}@digitalglobe.com

ABSTRACT

Detection of heterogeneous objects that cannot be delineated as a single region through image segmentation is a difficult task in the analysis of very high spatial resolution images. We describe an approach that uses individual region occurrence and pairwise co-occurrence histograms in image windows using logistic regression classifiers that simultaneously perform feature selection and learn classification models from a small number of examples. The proposed generic method is used to learn a sparse discriminative model to localize different compound objects in large image scenes. Experiments using WorldView-2 data show that the method can successfully detect objects like school, retail, park, and residential areas using similar parameter settings.

Index Terms— Object detection, facility detection, region co-occurrence, logistic regression, regularization

1. INTRODUCTION

Increasing spatial and spectral resolution in the new generation optical sensors has enabled the acquisition of new details that were not previously visible in satellite images. However, these details have also limited the application of conventional object detection techniques that are based on traditional image segmentation and classification algorithms which expect the objects of interest to appear as homogeneous regions. Even though some objects such as buildings, roads and trees that have relatively homogeneous spectral content and consistent shape can be recognized with a certain accuracy in these images, detection of more complex objects such as schools, power plants, shopping malls, recreational grounds is still very difficult due to their heterogeneous content.

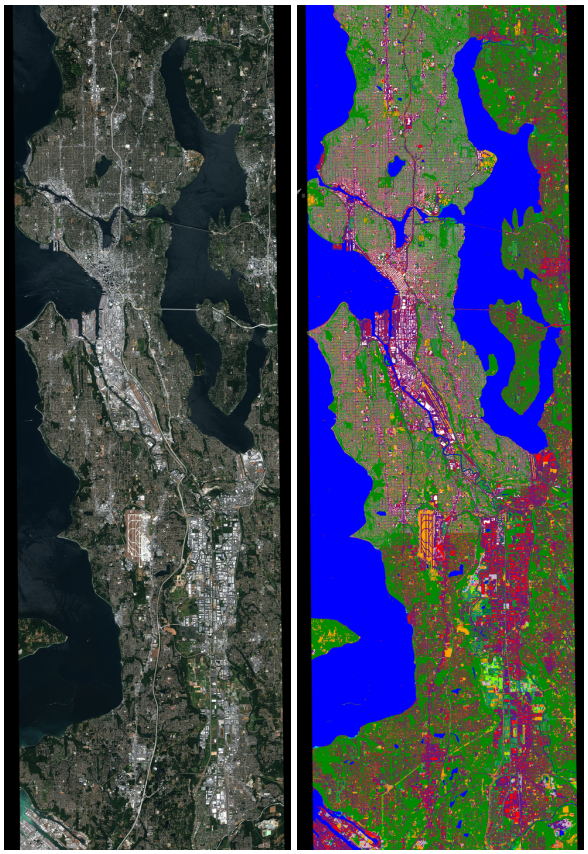
Previous work on the detection of heterogeneous structures in very high spatial resolution images involves the detection of specific structures such as airports [1] and orchards [2]. However, these methods are not generalizable because they rely on the specific characteristics of these objects. More generic methods include Gaussian mixture density estimation using histogram of Gabor texture features for the detection of golf courses and harbors [3], two-level classification where the outputs of pixel-based classifiers in the first level are given as input to the classifier in the second level for the detection of high schools [4], and clustering of histograms of visual words obtained using spectral, texture, and edge-based features for the detection of nuclear plants, coal power plants, and airports [5]. Nevertheless, these methods do not explicitly model the spatial structures that comprise the objects of interest. An alternative is described in [6] where signatures based on the spatial relationships of pairs of regions were used for latent topic discovery. Yet, such methods tend to capture very specific information, and may need very careful adjustment of the parameter settings for different applications.

This paper describes our work on the detection of compound heterogeneous objects that are modeled using combinations of simpler homogeneous regions. After obtaining the regions using image segmentation (Section 3), we represent each image window using the frequencies of the occurrence of individual region classes (Section 4) as well as their pairwise co-occurrence frequencies (Section 5). Then, we use logistic regression with a sparsity constraint to train a binary classifier that also identifies important features for each object type (Section 6). We illustrate the proposed method in the detection of school, retail, park, and residential area objects (Section 7) using a WorldView-2 image of King County in the Washington State, USA (Section 2).

2. DATA SET

The data set used in this paper consists of a multispectral WorldView-2 image with a size of $28,920 \times 9,804$ pixels and 2 m spatial resolution covering the King County, including Seattle, in the Washington State, USA. Figure 1(a) shows the true-color image. The reference data for object detection were obtained from the Open Street Map by querying the shape database using the tags "school", "retail", "residential", and "park". Table 1 shows the summary of the reference data. The Open Street Map data are known to have some inconsistencies and errors, but were used in the experiments because of the unavailability of any other GIS data.

This work was supported by DigitalGlobe, Inc. S. Aksoy was also supported in part by a Fulbright Visiting Scholar Grant.



(a) (b)

Fig. 1. Seattle data (King County) from a WorldView-2 image with a size of $28,920 \times 9,804$ pixels. (a) True-color image. (b) Land cover classification and segmentation.

3. REGION SEGMENTATION

Our image segmentation method is based on pixel classification. Feature vectors that included spectral values, ratios between spectral bands, mean values in 3×3 pixel neighborhoods, Gabor texture, and histograms of 20 spectral clusters within the 9-pixel neighborhood were used with a random forest classifier with 20 trees to classify each pixel into one of 16 classes: water, swimming pool, shrub, barren, sand, tidal flat, wetland, dry grass, pasture hay, cultivated crop, green grass, roof, road, shadow, tree, clearcut. The results of initial land cover classification were post-processed using a mixture of minimum mapping unit processing, topological rules, and majority analysis, and were converted into a region segmentation. We also added another class consisting of buildings obtained from the City of Seattle GIS database. The building shape layer was converted to a binary mask, the pixels corresponding to the mask were set to a new class called building, and the corresponding regions (connected components) were treated like the rest of the classes. The 17-class region segmentation and classification result is shown in Figure 1(b).

Table 1. Summary of the reference data. The number of positive examples (N) and the sizes (height $\times$ width) of the smallest, average, and largest reference object bounding boxes are shown.

Object	N	Smallest	Average	Largest
School	148	11×10	123×103	256×245
Retail	144	7×10	117×112	372×471
Park	491	5×1	154×171	$1,044 \times 1,346$
Residential	344	6×4	201×200	$1,653 \times 2,272$

4. REGION OCCURRENCE HISTOGRAM

After the regions are obtained, the first set of features consists of region occurrence histograms. These histograms are computed for a given window where each bin in the histogram stores the number of regions that are in that window and belong to a particular land cover class.

In order to capture the size information and differentiate between regions that belong to the same land cover class but have significantly different sizes, we further divided each land cover class into several sub-categories. First, we computed the distribution of size values using all regions in the data. This distribution was approximated in a parametric form using a Gamma density whose parameters were computed using maximum likelihood estimation. The Gamma density was selected on the basis of empirical evaluation of the size distribution that showed an exponential behavior. After the parametric density was obtained, a non-uniform quantizer was constructed using levels that were selected according to the cumulative probabilities (e.g., 0.25, 0.50, 0.75, 0.90, etc.) computed from the estimated Gamma density, and the size values were quantized using these levels. Thus, the length of the region occurrence histogram as a feature vector that incorporates both land cover and size information becomes $c \times q$ where c is the number of land cover classes and q is the number of size quantization levels.

5. REGION CO-OCCURRENCE HISTOGRAM

The region occurrence histograms described above ignore the spatial arrangements of the regions inside the window. We chose to model the spatial relationships of the regions using their co-occurrence statistics. Our choice of this second-order model was motivated by the success of the co-occurrence features in texture modeling and the combinatorial growth of the number of possible relationships for higher-order models.

Given c land cover classes and q size quantization levels, each bin $(i, j), i, j \in \{1, \ldots, c \times q\}$, in the region co-occurrence matrix stores the frequency of pairs of regions occurring in a window, one with land cover class and size level combination i and the other with land cover class and size level combination j. We assume that the regions that are close enough are related, and count only the pairs whose distances

are less than a maximum distance threshold. The distance between two regions is computed as the smallest distance between their boundary pixels. Since the co-occurrence matrices are symmetric, we only use the main diagonal and the upper triangular parts, and append them into a feature vector with length $(c \times q)(c \times q + 1)/2$. Finally, we append the region occurrence histogram and the region co-occurrence histogram together, and normalize each histogram component to zero mean and unit variance to construct the feature vector given as input to the learning and classification process.

6. LEARNING AND CLASSIFICATION

In this paper, we pose the object detection task in a binary setting where a given image window is classified as containing a target object of interest or not along with an associated probability of this decision. The binary classification is performed using logistic regression that is a popular discriminative classifier that assumes a parametric form for the posterior probability and directly estimates its parameters from the training data without any need for the assumption or the estimation of the class conditional probability distributions.

Let $\mathbf{x} \in \mathbb{R}^d$ denote the feature vector and $y \in \{-1, 1\}$ denote the corresponding binary class variable where 1 represents the target object class and -1 represents the background. The logistic model has the form

$$p(y = 1|\mathbf{x}; \mathbf{w}, w_0) = \frac{1}{1 + \exp(-\mathbf{w}^T\mathbf{x} - w_0)} \qquad (1)$$

where $\mathbf{w}$ is the weight vector and w_0 is the intercept [7]. Given labeled data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, the maximum likelihood estimates of $\mathbf{w}$ and w_0 can be found by solving

$$\min_{\mathbf{w},w_0} \sum_{i=1}^{n} \log\Big(1 + \exp\big(-y_i(\mathbf{w}^T\mathbf{x}_i + w_0)\big)\Big). \qquad (2)$$

When the number of training examples (n) is not large enough compared to the number of features (d), as in our case, the logistic regression classifier tends to suffer from the over-fitting problem in which the resulting model has many features with relatively large weights that memorize the peculiarities of the training data. A standard method to prevent over-fitting is regularization where an extra term that penalizes large weights is added to the cost function used in estimation. The l_1-regularization has shown great empirical success in the literature, particularly due to its sparsity-inducing property that leads to solutions with fewer nonzero parameter values [8]. Thus, the learning process implicitly performs feature selection while optimizing the cost function to estimate the parameters. The l_1-regularized formulation corresponds to the solution of

$$\min_{\mathbf{w},w_0} \sum_{i=1}^{n} \log\Big(1 + \exp\big(-y_i(\mathbf{w}^T\mathbf{x}_i + w_0)\big)\Big) + \lambda\|\mathbf{w}\|_1 \quad (3)$$

Table 2. Summary of the classification results. The number of size quantization levels (SQL) and the maximum distance threshold (MDT) that resulted in the best performance are shown together with the corresponding area under the ROC curve (AUC, as mean $\pm$ std).

Object	SQL	MDT	AUC
School	5	20	0.9639 ± 0.0157
Retail	5	20	0.9691 ± 0.0161
Park	5	20	0.9436 ± 0.0137
Residential	5	50	0.8975 ± 0.0186

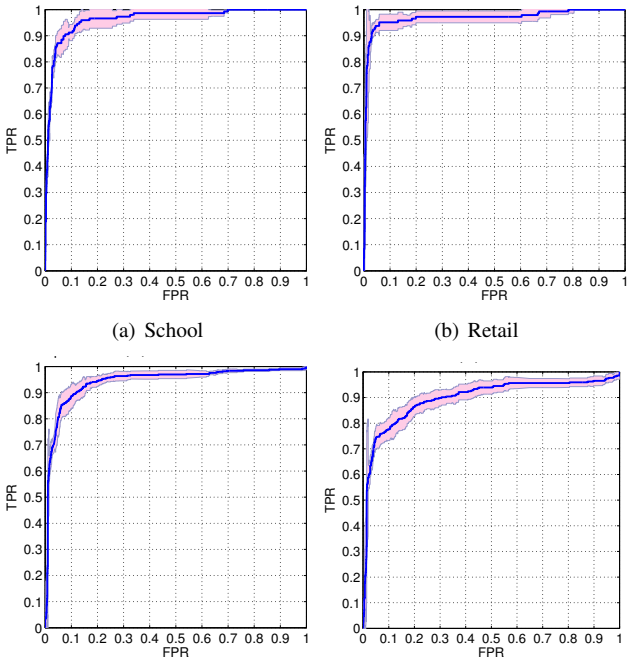
where λ is the regularization parameter. We use the l_1-ball constrained smooth convex optimization formulation [9] to obtain the weight vector $\mathbf{w}$ and the intercept w_0 as the solution of (3). Then, given a test window $\mathbf{x}$, the discriminative logistic regression classifier chooses the target class 1 if $p(y = 1|\mathbf{x}; \mathbf{w}, w_0) > 0.5$, or equivalently, $\mathbf{w}^T\mathbf{x} + w_0 > 0$.

7. EXPERIMENTS

We used 5-fold cross-validation to train the classifier and evaluate its accuracy. The training stage consisted of the estimation of the logistic regression parameters $\mathbf{w}$, w_0, and λ. We considered three different numbers of size quantization levels (1, 5, 10) and three different maximum distance thresholds (5, 20, 50). In addition to the positive examples corresponding to the reference object masks summarized in Table 1, for each object class, we randomly sampled $2,000$ windows of 100×100 pixels from the image to construct the set of negative examples.

Quantitative performance evaluation was done by using true positive and false positive rates computed from confusion matrices to construct ROC curves based on different thresholds on the posterior in (1). The area under the ROC curve was used as the criterion to compare different parameter settings. Qualitative performance evaluation was performed via visual inspection of the classification maps computed using 100×100 pixel sliding windows with 10 pixel increments.

Table 2 summarizes the classification results using 5-fold cross-validation. Figure 2 presents the resulting ROC curves. We observed that the best performing number of size quantization levels was the same (5) for all object classes. The best performing maximum distance threshold was obtained as 20 pixels for school, retail, and park, while 50 pixels gave the best results for residential due to the large neighborhoods required for the co-occurrence context in larger residential areas. The areas under the ROC curves were considerably high on the small number of examples obtained from the Open Street Map. We observed that the false positive rates at high levels of true positive rates can be higher in practice when the classifiers are applied to the whole satellite scene. Detection examples for schools are shown in Figure 3.



(a) School (b) Retail

(c) Park (d) Residential

Fig. 2. ROC curves for different object classes. y-axis is the true positive rate and x-axis is the false positive rate. The red bands around the curves show the confidence intervals computed from the cross-validation folds.

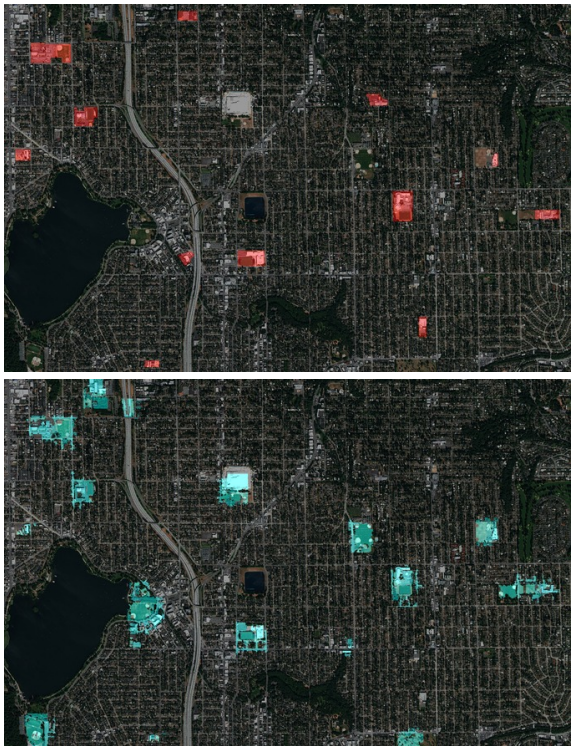


Fig. 3. School detection results in a $1,760 \times 2,717$ pixel residential neighborhood. There are 12 schools in the reference data (red overlay), and the algorithm detected 16 schools (green overlay) among which 8 are true positives.

8. CONCLUSIONS

We described an algorithm for the detection of heterogeneous objects in very high spatial resolution images using occurrence and co-occurrence histograms of regions in image windows as features for a logistic regression classifier. Experiments were performed using shape data from the Open Street Map. Future work includes more detailed evaluation of different parameters and object classes.

9. REFERENCES

[1] D. Liu, L. He, and L. Carin, "Airport detection in large aerial optical imagery," in *ICASSP*, May 17–21, 2004, vol. 5, pp. 761–764.

[2] S. Aksoy, I. Z. Yalniz, and K. Tasdemir, "Automatic detection and segmentation of orchards using very high-resolution imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 8, pp. 3117–3131, August 2012.

[3] S. Bhagavathy and B. S. Manjunath, "Modeling and detection of geospatial objects using texture motifs," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3706–3715, December 2006.

[4] N. R. Harvey et al., "Detection of facilities in satellite imagery using semi-supervised image classification and auxiliary contextual observables," in *SPIE Visual Information Processing XVIII*, Orlando, Florida, April 13, 2009, vol. 7341.

[5] R. R. Vatsavai, A. Cheriyadat, and S. Gleason, "Supervised semantic classification for nuclear proliferation monitoring," in *IEEE Applied Imagery Pattern Recognition Workshop*, Washington, DC, October 13–15, 2010.

[6] C. Vaduva, I. Gavat, and M. Datcu, "Latent Dirichlet allocation for spatial analysis of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2770–2786, May 2013.

[7] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, pp. 267–288, 1996.

[9] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, 2009, pp. 547–556.