Estimating the Spatial Extents of Geospatial Objects Using Hierarchical Models

Yi Yang and Shawn Newsam
Electrical Engineering & Computer Science
University of California at Merced
yyang6,snewsam@ucmerced.edu

Abstract

The goal of this work is to estimate the spatial extents of complex geospatial objects such as high schools and golf courses. Gazetteers are deficient in that they currently specify the spatial extents of these objects using a single latitude/longitude point. We propose a framework that uses readily available high resolution overhead imagery to estimate the boundaries of known object instances in order to update the gazetteers. Key to our approach is a hierarchical object model with three levels. The lowest level characterizes an object using local invariant features; an intermediate, latent level characterizes the land-use/land-cover (LULC) classes that constitute an object; and, the top level models an object as a distribution over these classes.

We evaluate our approach using a manually labeled ground truth dataset of four object types: high schools, golf courses, mobile home parks, and Costco shopping centers.

1. Introduction

Advances in technology continue to increase our ability to capture and store overhead imagery such as that taken from satellite or aerial platforms. While this data has great potential to benefit society, our ability to analyze this imagery has not scaled proportionally and so automated methods for extracting useful information are needed. Significant progress has been made over the last several decades in automating the analysis of overhead imagery but the bottleneck to realizing the true value of this data remains the need for manual inspection which is time intensive.

The work in this paper represents a step towards the automated analysis of high resolution overhead imagery, in particular for the detection of complex geospatial objects that are composed of multiple land-use/land-cover (LULC) classes such as high schools and golf courses. We focus here though on the *preliminary problem of estimating the spatial extents of known object instances using a novel hierarchical geospatial object model*. Computing the spatial footprints of known objects is itself an important and challenging problem. It also provides a constrained setting in which to develop and evaluate the object models before tackling the more challenging problem of detecting novel object instances.

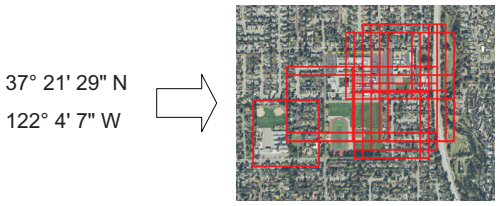


Figure 1. The goal of this work is to estimate the spatial extent of complex geospatial objects. Gazetteers are deficient in that they currently specify the spatial extents of objects *using only a single latitude/longitude point* as shown on the left for this high school. We propose a method to estimate the true spatial extent from such a point using a hierarchical object model. The results of our technique, indicated on the right by the union of the red regions, can then be used to update the gazetteer.

Our work is motivated by the fact that current gazetteers, geographic dictionaries of what-is-where on the surface of the Earth, are deficient in that *the spatial extents of the archived objects are limited to a single point, a latitude/longitude pair*. While the systems include provisions for storing at least a bounding box representation, this information has simply never been acquired or computed. As the development team of the University of California at Santa Barbara Alexandria Digital Library (ADL) gazetteer points out [13], "for a digital library application, the spatial extent of the feature, either approximately with a bounding box or more accurately with a polygonal representation, is better, but there are no large sets of gazetteer data with spatial extents." They go on to state that "establishing the standards that will enable the sharing of gazetteer data will help harvest data from many sources, but ultimately deriving spatial locations and extents from digital mapping products and other sources automatically will be needed." Our work in this paper does just as the ADL gazetteer development teams proposes: we leverage readily available high resolution overhead imagery to estimate the spatial extents of known object instances with minimal user supervision.

We propose a novel hierarchical geospatial object model

with three levels. At the lowest level, local invariant features are used to represent the pixel level information in the image. A latent intermediate level characterizes tiled image regions using a set of LULC classes. Finally, the top level represents the geospatial objects as distributions over the underlying LULC classes.

We demonstrate our approach using a manually labeled ground truth dataset containing four object types: high schools, golf courses, mobile home parks, and Costco shopping centers. This dataset is created from license-free aerial orthoimagery obtained from the United States Geological Survey (USGS) National Map and will be made publicly available to other researchers. To our knowledge, it is the first dataset of its kind that can be openly distributed.

The salient aspects of our work are as follows:

- A image based solution that addresses a significant deficiency in current gazetteers.
- A novel hierarchical model with a latent LULC level.
- A framework which leverages image and non-image data to update the gazetteers. This framework requires very few labeled training images.
- A manually labeled ground truth dataset which will be made publicly available to other researchers[1].

2. Related Work

While we are not aware of similar work on modeling widely varying geospatial objects composed of multiple LULC classes, aspects of our work are related to the following research areas in computer vision and remote sensing.

First, there has been significant effort and success in object recognition in standard (non-overhead) imagery by the computer vision community over the last decade particularly using local invariant features. An overview of this work is beyond the scope of this paper but a good survey can be found in [21].

Another related area is combining image and non-image data to improve image understanding. (In our case, the gazetteer records are the non-image data.) In particular, computer vision researchers have exploited various forms of meta-data associated with image collections to learn visual object models. Berg et al. [5] data mine a large collection of captioned images of faces from online news sources to train a recognition system for commonly occurring people. Barnard et al. [3] develop an object recognizer using 10,000 images of works of art along with associated free text which varies greatly from physical description to interpretation and mood. And, Li et al. [16] turn the search paradigm around by using search results from the Google image search engine to learn visual models for a variety of object categories.

Researchers working in the geographic information sciences have likewise proposed a number of ways to leverage non-image data sources to improve remote sensed image understanding. Using satellite or aerial imagery to maintain road networks has always held great appeal but automatically extracting roads is a challenging task. An obvious way to improve road extraction, at least for known roads, is to use existing vectorized road networks as seeds [32, 2, 11]. Researchers have also incorporated other information to improve road extraction, such as using digital surface models to account for gaps between road segments due to shadows [4]. Automated building extraction is another appealing use of remote sensed imagery. Agouris et al. [1] propose a SpatioTemporal Gazetteer that incorporates aerial imagery as well as existing vector datasets of extracted outlines and thematic datasets (building blueprints, building usage records) to automatically detect changes to the spatial footprints of buildings using template matching.

Finally, the remote sensing community has begun to realize the potential of *local invariant features* for image analysis especially in high resolution imagery. A number of methods have been developed to perform image matching for registration [15, 9, 18, 10, 29] and change detection [14, 27]. Closer to the work presented in this paper, researchers have investigated local features for detection and classification. Sirmacek and Unsalan [22, 23, 24] use local features to detect buildings and urban areas in 1m resolution IKONOS imagery. Xu et al. [30] compare quantized color and texture features with local features for classifying 0.25m resolution aerial image regions into four LULC classes. Chen et al. [8] also compare local features with standard color and texture features to classify 0.5m Digital Globe imagery into 19 LULC classes. Skurikhin [26] investigates attention based saliency detection to perform local feature based classification of 0.5m resolution Digital Globe and Google Earth imagery into anthropogenic or natural regions. Gleason et al. [12] and Vatsavai et al. [28] use quantized local features to detect complex geospatial objects such as nuclear and coal power plants in 1m resolution Digital Globe imagery. Ozdemir and Aksoy [20] investigate graph-based spatial arrangements of quantized local features to classify 1m resolution Ikonos imagery into eight LULC classes. And, Bordes and Prinet [6] investigate spatial correlograms of quantized local features to classify high resolution Digital Globe imagery into eight LULC classes.

3. Approach Overview

An overview of the proposed approach is shown in figure 2. First, gazetteers are queried for object instances, for example all high schools in a geographic region such as a city. The point locations of these objects are then used to retrieve high resolution images from online repositories. The spatial extents of the objects are manually labeled in a small subset of the images to form a training set which is used to learn the object models. The model is used to estimate the spatial extents of the objects in the target images. Fi-

[1]The dataset is available at http://vision.ucmerced.edu/datasets.

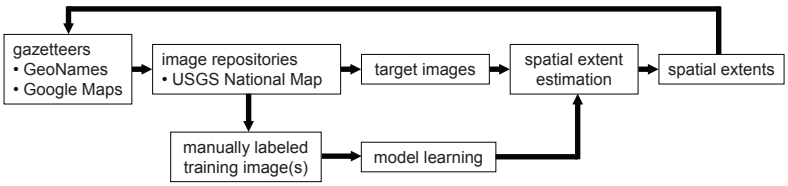


Figure 2. An overview of our proposed approach for estimating the spatial extents of geospatial objects.

nally, the gazetteers are updated with the spatial extents of the originally queried objects.

3.1. Data Sources - Gazetteers

Gazetteers are geographic dictionaries that record what-is-where on the surface of the Earth. We utilize two gazetteers in this work. First GeoNames[2], an online worldwide gazetteer compiled from several dozen sources including other gazetteers such as the USGS Geographic Names Information System. It contains over 7.5 million features (objects) categorized into nine top-level classes which are further subcategorized into 645 feature codes. All the data is accessible free of charge through a number of webservices as well as a daily database export. The GeoNames web interface allows fuzzy search using geographic names, locations, features codes, and feature classes. Queries to GeoNames return a single latitude/longitude point as the spatial extent of an object.

We also treat Google Maps as a gazetteer in that it allows us to perform location-based searches for geospatial objects such as Costco shopping centers. We further use the Google Maps Geocoding API[3] to translate the street addresses provided by Google Maps into latitude/longitude points.

3.2. Data Sources - Image Repositories

Our image-based spatial extent estimation is made possible by the availability of high resolution overhead imagery. We limit our study area to the US and use the USGS National Map Seamless Data Server[4] interface to automatically download imagery. This interface accepts spatial queries for a range of data collections including High Resolution Orthoimagery of major US urban areas at 3-inch, 6-inch, 1-foot, and 2.5-foot spatial resolutions, and the US Department of Agriculture (USDA) National Agriculture Imagery Program imagery of the conterminous United States at 1-meter or 2-meter spatial resolutions.

Images are retrieved from the National Map using a simple rectangular query region specified by its bounding latitude and longitude values. In our case, the single latitude/longitude point from the gazetteer serves as the center of a region whose size is chosen to ensure that the retrieved image contains the target object. This size is chosen empirically in the experiments below based on the observed sizes of sample objects. A single size is picked for each object type and then fixed for all the retrievals. Note that the gazetteer point does not always fall inside the object due to data collection, geo-registration, or other errors.

[2]http://www.geonames.org

[3]http://code.google.com/apis/maps/documentation/geocoding

[4]http://seamless.usgs.gov

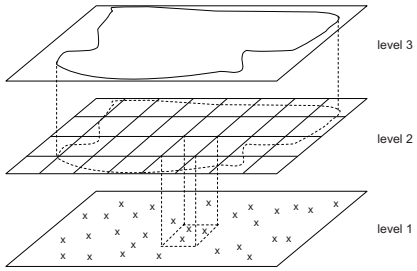


Figure 3. The three levels of our hierarchical model. Level 1 represents the object using quantized SIFT features shown here as x's. BOVW histograms are computed for image tiles and SVM classifiers are used to assign LULC labels to the tiles in level 2. The distribution of the LULC classes in level 3 constitutes the final object model.

4. Hierarchical Object Model

The three levels of the hierarchical model are shown in figure 3. We now describe each of the levels in detail.

4.1. Level 1 - Local Invariant Features

We use local invariant features to characterize the objects at the lowest level of the hierarchy. These features are designed to be robust to image variations caused by geometric image transformations such as scaling and rotation as well as to photometric distortions caused by variation in illumination, etc. They have proven to be effective for a range of computer vision applications.

Extracting local invariant features is a two-step process. First, a detection step locates salient points that are identifiable under different viewing conditions. This process ideally locates the same regions in an object or scene regardless of viewpoint or illumination. Second, these locations are described by a descriptor that is distinctive yet invariant to viewpoint and illumination.

We choose David Lowe's Scale Invariant Feature Transform (SIFT) [17] as our local invariant feature detector and descriptor. The SIFT detector, like most local feature detectors, results in a large number of feature points. This density is important for robustness but presents a representation challenge particularly since the SIFT descriptors have

128 dimensions. We adopt a standard bag-of-visual-words (BOVW) [25] approach to summarize the descriptors by quantizing and aggregating the features without regard to their location. We first construct a visual dictionary by performing k-means clustering on a large number of SIFT features (from a dataset different from that used to train the object models). This dictionary is then used to quantize the individual SIFT points into "visual words" by simply assigning the label of the closest cluster centroid. We aggregate the quantized features at the image tile level using a BOVW histogram

$$BOVW = [t_1, t_2, \ldots, t_V] \ ,$$

where t_v is the number of occurrences of visual word v in a tile and V is the dictionary size. The BOVW histogram is normalized to have unit L1 norm to account for the difference in the number of interest points between tiles.

We use 256x256 pixel tiles in all the experiments below.

4.2. Level 2 - Latent LULC Classes

An intermediate, latent level bridges the gap between the low-level local invariant features and the high-level objects. Specifically, LULC labels are assigned to image tiles using support vector machines (SVMs).

We leverage our recent work [31] on LULC classification. In that work, we used a large ground truth dataset to train SVM classifiers for a number of LULC classes. We demonstrated that the BOVW histograms outperform color histograms and texture features through extensive evaluation.

We use a one-against-all strategy to perform multi-class SVM classification. We also use the probabilistic output option of the LIBSVM package [7]. Specifically, for each tile i in an image, we compute the probability distribution over the M LULC classes as

$$P(tile_i) = [p_1, p_2, ..., p_M] \ ,$$

where p_m corresponds to the probability that tile i is assigned to the mth class by the SVM classifiers. The SVM classifiers take as input the BOVW histograms from level 1. We normalize $P(tile_i)$ so that $\sum^M p_m = 1$.

In order to reduce the effect of tile (mis)alignment, we perform the LULC labeling on tiles which overlap by 50 percent. Thus, each 128x128 pixel *block* appears in four 256x256 pixel tiles. We apply a smoothing mechanism to the LULC class distribution at the block level

$$P(block_j) = \frac{1}{4} \sum P(tile_i) \ , \qquad (1)$$

where the sum is taken over the four tiles in which block j appears.

To summarize, our final representation at level 2 in the hierarchy is a probability distribution $P(block_j)$ over M LULC classes for each 128x128 pixel block j.

4.3. Level 3 - Object Model

The top level of our representation also models the objects as probability distributions over LULC classes. The distributions corresponding to different object types can be easily learned from one or more training samples. Given N training samples encompassing a set of $\mathbb{U}$ blocks labeled at level 2, we compute

$$P(object) = \frac{1}{|\mathbb{U}|} \sum_{block_j \in \mathbb{U}} P(block_j) \ , \qquad (2)$$

where $P(block_j)$ is computed using equation 1 and $|\mathbb{U}|$ is the cardinality of $\mathbb{U}$.

5. Spatial Extent Estimation

The primary goal of this paper is to estimate the spatial extent of known object instances. Again, in the context of our problem, the gazetteer provides a single latitude/longitude point for the object. This point is used to download a target image T large enough to encompass the object. An object model $P(object)$ is then used to estimate the spatial extent of the object as follows.

First, we extract and quantize SIFT features from the target image using the same visual dictionary as in level 1 of the object model. We then compute the BOVW histograms for overlapping 256x256 pixel tiles and the multi-class SVM classifiers are used to to compute the LULC class distributions for each of the tiles. The LULC class distributions are then computed for each 128x128 pixel block using equation 1.

The problem now reduces to determining the contiguous set of blocks that are most similar to the object model. We simplify this search by 1) scoring overlapping square windows each containing a fixed number of blocks, 2) applying a threshold to the scores, and 3) computing the final spatial extent as the union of the selected windows.

Specifically, we slide a square window of size wxw blocks over the image in increments of one block. For each window location, we compute the probability distribution of the window over the LULC classes:

$$P(window) = \frac{1}{w^2} \sum_{block_j \in window} P(block_j) \ , \qquad (3)$$

where $P(block_j)$ is computed using equation 1. We then compute the similarity between the window and the object model $D(window, object)$ using the intersection measure

$$D(window, object) = \sum^M min(P(window)[m], P(object)[m])) \ , \qquad (4)$$

where $[m]$ indicates the mth component and M is the number of LULC classes. If $D(window, object)$ is above a

threshold θ, we label all the blocks in the window as belonging to the target object. (We discuss the setting of θ below.) Finally, after each window location has been visited, we compute the spatial extent of the object as the union of all the selected blocks.

6. Experimental Results

We demonstrate our approach using an evaluation dataset consisting of four object types: high schools, golf courses, mobile home parks, and Costco shopping centers.

6.1. Dataset

We use the first stage of the framework in figure 2 to identify the locations of target objects and their corresponding images. The GeoNames gazetteer is used to identify 44 high schools, 27 golf courses, and 23 mobile home parks, and Google Maps is used to identify 18 Costco shopping centers. The National Map Seamless Data Server is then used to download 1-foot resolution orthoimagery using a large query region to ensure the images contain the target objects. The images are in the RGB colorspace.

A ground truth dataset is created by manually delineating the target objects using a polygon representation. This labeling was done by undergraduates in our lab with no knowledge of the proposed approach. We also compute the rectangular, axis aligned bounding boxes of the target objects using the polygonal boundaries.

SIFT features are extracted from each of the images and quantized using a visual dictionary consisting of 100 visual words. In previous work [19], we showed that a dictionary of this size represents a good balance between efficiency and accuracy. A BOVW histogram is computed for overlapping 256x256 pixel tiles.

Tile-level LULC distributions are computed using a set of SVMs corresponding to 18 LULC classes: agricultural, airplane, baseball diamond, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, runway, sparse residential, and tennis courts. Finally, block-level LULC distributions are computed using equation 1 and object-level distributions are computed using equation 2.

6.2. Evaluation

We quantitatively evaluate the results by computing two values. First, how much of the true spatial extent is selected, and second, how much of the estimated spatial extent does not belong to the true spatial extent. We want the first to be large and the second to be small. These values are similar to true positive and false positive rates but it does not make sense to compute the standard false positive rate since it is sensitive to the size of the target images retrieved from the gazetteers.

Given a target image with ground truth spatial extent L_{true}, and estimated spatial extent L_{est} (corresponding to a specific setting of the window similarity threshold θ), we compute the true location rate (TLR) as

$$TLR = \frac{|L_{est} \bigcap L_{true}|}{|L_{true}|}\ , \tag{5}$$

and compute the false location rate (FLR) as

$$FLR = \frac{|L_{est}| - |L_{est} \bigcap L_{true}|}{|L_{true}|} \tag{6}$$

where $|\cdot|$ indicates the area of a region in pixels and $\bigcap$ indicates set intersection. TLR ranges from 0 to 100 percent while *FLR can exceed 100 percent*.

We consider two cases of our problem. First, where the ground truth spatial extent is a polygon. In this case, L_{true} is the set of 128x128 pixel blocks contained within the ground truth *polygon* (a block is considered to be inside a polygon if the majority of its area is) and L_{est} is the set of blocks computed in section 5. We also consider the case where the ground truth spatial extent is a *bounding box* (derived from a the ground truth polygon). In this case, L_{est} is the bounding box encompassing the set of blocks computed in section 5.

6.3. Experiments

We perform an extensive set of experiments where we use each ground truth item to learn an object model and then apply the model to estimate the spatial extents of objects in the remaining images. That is, if we have N ground truth instances of an object, we perform N-fold cross-validation wherein each of the N instances is used to train a model which is then applied to the remaining $N-1$ images. We perform this separately for the four object types.

We evaluate the effect of a number of design parameters including the size of the window used in the spatial extent estimation, and whether the ground truth spatial extent is a polygon or bounding box.

6.4. Results

The threshold θ that is used to determine whether a window is sufficiently similar to the object model during the estimation step is a key parameter. We therefore create the equivalent of an ROC curve showing how TLR and FLR vary as θ is decreased from 1 to 0. Based on these curves, we pick and fix a value of θ for each object type that achieves a good tradeoff between TLR and FLR. We then compute the mean and standard deviation of TLR and FLR over all trials in the N-fold cross-validation where N again is the number of ground truth instance of an object type. This is a total of $(N)(N-1)$ trials. We do this for each of the four object types and for the different design parameter settings. These results are summarized in table 1.

The columns of this table indicate the size (in pixels) of the window used in the spatial extent estimation. The top section of results correspond to the case where the ground truth spatial extent is a polygon and the estimated spatial extent is the union of all windows determined to be similar to the object model.

The bottom two sections of this table correspond to the case where the ground truth spatial extent is a bounding box. We show two subcases here. The first, termed BB all, corresponds to the case where the estimation step results in multiple disconnected regions (see figure 5(b) for example). We here compute the final spatial extent estimation as the union of the bounding boxes of the individual regions. In the second case, termed BB best, we apply a simple heuristic to choose the region that is most similar to the model using the intersection measure in equation 4.

Several trends can be observed in table 1. First, as the window size increases, both TLR and FLR increase. This makes sense as more blocks will be labeled as belonging to the object. Unfortunately, FLR increases faster than TLR so it is difficult to choose the optimal window size based purely on these results. This is a subject for future investigation. The other trend is that while both TLR and FLR are lower for the cases where the ground truth is a bounding box, FLR generally decreases faster than TLR. This shows that our approach does better at estimating a bounding box spatial extent that a more precise polygon. This makes sense but is significant because most gazetteers only include provisions for a bounding box representation (which is a fixed sized representation versus a polygon representation which has variable length).

Overall, our approach is shown to be effective especially given that our models are trained using *a single training image thus keeping user supervision very minimal*. We typically are able to estimate more than fifty percent of the true spatial extent with a false location estimation of smaller than the area of the target object.

Finally, results for three samples of each object type are shown in figures 4-7. In these results, the yellow polygons indicate the ground-truth spatial extents and the union of the red regions indicate the estimated spatial extents for the empirically chosen threshold value θ.

7. Discussion

The most salient aspect of our hierarchical model is the latent intermediate level. First, by characterizing the LULC classes that constitute an object, it allows our approach to bridge the gap between the low-level features and the high-level objects. Second, it allows us to model complex objects which are composed of multiple LULC classes. And, finally, its effectiveness is due as much to it modeling the LULC classes that do not appear in an object as those that do. In particular, the large number of LULC classes allows the windowing step to readily reject background regions that have high proportions of classes which do not appear in the object. Such discrimination would not be possible using binary single-class LULC classifiers (and then again, would only be applicable to homogeneous objects).

8. Conclusion and Future Work

We presented a framework that leverages readily available high resolution overhead imagery to estimate the spatial extents of geospatial objects using a hierarchical model. We demonstrated the approach using a challenging ground truth dataset of four object types.

Future work on this problem includes automating the selection of the threshold parameter θ possibly based on the expected sizes of the objects; exploring combining local invariant features with other low-level features such as color; extending the object model to incorporate the spatial distribution of the LULC classes; and using the model to detect novel object instances in newly acquired imagery.

9. Acknowledgements

This work was funded in part by NSF grant IIS-0917069 and a Department of Energy Early Career Scientist and Engineer/PECASE award. Any opinions, findings, and conclusions or recommendations expressed in this work are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] P. Agouris, K. Beard, G. Mountrakis, and A. Stefanidis. Capturing and modeling geographic object change: A spatiotemporal gazetteer framework. *Photogrammetric Engineering & Remote Sensing*, 66(10):1241–1250, 2000.

[2] P. Agouris, S. Gyftakis, and A. Stefanidis. Using a fuzzy supervisor for object extraction within an integrated geospatial environment. *International Archives of Photogrammetry and Remote Sensing*, 32(III/1):191–195, 1998.

[3] K. Barnard, P. Duygulu, and D. Forsyth. Clustering art. In *CVPR*, 2001.

[4] A. Baumgartner, W. Eckstein, H. Mayer, C. Heipke, and H. Ebner. Context-supported road extraction. *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, II:299–308, 1997.

[5] T. Berg, A. Berg, J. Edwards, M. Maire, R. White, Y.-W. Teh, E. Learned-Miller, and D. Forsyth. Names and faces in the news. In *CVPR*, 2004.

[6] J. Bordes and V. Prinet. Mixture distributions for weakly supervised classification in remote sensing images. In *BMVC*, 2008.

[7] C.-C. Chang and C.-J. Lin. *LIBSVM: A library for support vector machines*, 2001.

[8] L. Chen, W. Yang, K. Xu, and T. Xu. Evaluation of local features for scene classification using VHR satellite images. In *Urban Remote Sensing Joint Event*, 2011.

[9] J. Dai, W. Song, L. Pei, and J. Zhang. Remote sensing image matching via Harris detector and SIFT discriptor. In *International Congress on Image and Signal Processing*, volume 5, pages 2221–2224, 2010.

[10] L. Dorado-Munoz, M. Velez-Reyes, A. Mukherjee, and B. Roysam. A vector SIFT operator for interest point detection in hyperspectral imagery. In *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2010.

Table 1. Quantitative results. The columns indicate the window size in pixels used in the spatial extent estimation. TLR (true location rate) indicates how much of an object's true footprint is estimated. FLR (false location rate) indicates how much of the estimated region does not belong to the object's true footprint. Please see section 6.4 for additional details. All values are percentages.

			256		384		768		896		1408		2432	
			TLR	**FLR**	**TLR**	**FLR**	**TLR**	**FLR**	**TLR**	**FLR**	**TLR**	**FLR**	**TLR**	**FLR**
Polygon	**Costco**	**Mean**	31.6	113	45.0	156	72.9	309						
		Std.	6.39	37.5	9.32	42.8	13.1	31.9						
	MHP	**Mean**			41.6	168	64.4	300	68.5	352				
		Std.			7.28	79.2	9.01	95.5	7.55	81.7				
	GC	**Mean**			71.1	157			83.6	228	86.8	290		
		Std.			13.1	58.4			9.17	63.6	7.31	59.4		
	HS	**Mean**							53.1	77.8	79.6	124	91.1	331
		Std.							8.29	27.3	9.02	26.0	4.35	15.4
BB all	**Costco**	**Mean**	39.0	79.4	49.6	103	73.4	178						
		Std.	8.52	27.8	9.02	30.5	10.9	19.5						
	MHP	**Mean**			42.4	130	61.4	206	64.9	244				
		Std.			7.76	69.1	7.47	85.0	6.10	80.7				
	GC	**Mean**			81.3	137			85.9	160	87.8	186		
		Std.			7.67	59.7			4.74	52.1	4.17	50.5		
	HS	**Mean**							51.5	53.5	74.8	82.0	88.2	205
		Std.							7.88	22.5	7.53	23.5	4.14	13.2
BB best	**Costco**	**Mean**	24.4	35.0	39.1	66.3	67.0	159						
		Std.	8.06	13.3	10.1	19.3	11.8	20.7						
	MHP	**Mean**			22.5	26.8	43.5	132	52.8	181				
		Std.			5.52	10.5	9.48	60.2	8.19	71.6				
	GC	**Mean**			43.8	56.5			75.1	105	82.0	162		
		Std.			8.98	43.9			6.29	48.4	5.29	47.1		
	HS	**Mean**							44.2	30.4	69.2	71.8	88.0	205
		Std.							8.13	11.9	8.40	18.5	4.30	12.7

[11] P. Doucette, P. Agouris, M. Musavi, and A. Stefanidis. Automated extraction of linear features from aerial imagery using Kohonen learning and GIS data. In *ISD '99: Selected Papers from the International Workshop on Integrated Spatial Databases, Digital Inages and GIS*, 1999.

[12] S. Gleason, R. Ferrell, A. Cheriyadat, R. Vatsavai, and S. De. Semantic information extraction from multispectral geospatial imagery via a flexible framework. In *IEEE International Geoscience and Remote Sensing Symposium*, 2010.

[13] L. L. Hill, J. Frew, and Q. Zheng. Geographic names: The implementation of a gazetteer in a georeferenced digital library. *D-Lib*, 5(1), 1999.

[14] C. Huo, Z. Zhou, Q. Liu, J. Cheng, H. Lu, and K. Chen. Urban change detection based on local features and multiscale fusion. In *IEEE International Geoscience and Remote Sensing Symposium*, 2008.

[15] X. Jianbin, H. Wen, and W. Yirong. An efficient rotation-invariance remote image matching algorithm based on feature points matching. In *IEEE International Geoscience and Remote Sensing Symposium*, 2005.

[16] L.-J. Li, G. Wang, and L. Fei-Fei. Optimol: Automatic Online Picture collecTion via Incremental MOdel Learning. In *CVPR*, 2007.

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[18] A. Mukherjee, M. Velez-Reyes, and B. Roysam. Interest points for hyperspectral image data. *IEEE Trans. on Geoscience and Remote Sensing*, 47(3):748–760, 2009.

[19] S. Newsam and Y. Yang. Geographic image retrieval using interest point descriptors. In *Advances in Visual Computing 2007, Lecture Notes in Computer Science (LNCS)*, volume 4842, pages 275–286, 2007.

[20] B. Ozdemir and S. Aksoy. Image classification using subgraph histogram representation. In *ICPR*, 2010.

[21] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors. *Toward Category-Level Object Recognition*, volume 4170 of *LNCS*. Springer, 2006.

[22] B. Sirmacek and C. Unsalan. Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. on Geoscience and Remote Sensing*, 47(4):1156–1167, April 2009.

[23] B. Sirmacek and C. Unsalan. Urban area detection using local feature points and spatial voting. *IEEE Geoscience and Remote Sensing Letters*, 7(1):146–150, 2010.

[24] B. Sirmacek and C. Unsalan. A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Trans. on Geoscience and Remote Sensing*, 49(1):211–221, 2011.

[25] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, 2003.

[26] A. Skurikhin. Visual attention based detection of signs of anthropogenic activities in satellite imagery. In *IEEE Applied Imagery Pattern Recognition Workshop*, 2010.

[27] F. Tang and V. Prinet. Computing invariants for structural change detection in urban areas. In *Urban Remote Sensing Joint Event*, 2007.

[28] R. R. Vatsavai, A. Cheriyadat, and S. Gleason. Unsupervised semantic labeling framework for identification of complex facilities in high-resolution remote sensing images. In *Proceedings of the International Conference on Data Mining–Workshops*, 2010.

[29] Z. Xiong and Y. Zhang. A novel interest-point-matching algorithm for high-resolution satellite images. *IEEE Trans. on Geoscience and Remote Sensing*, 47(12):4189–4200, 2009.

[30] S. Xu, T. Fang, D. Li, and S. Wang. Object classification of aerial images with bag-of-visual words. *IEEE Geoscience and Remote Sensing Letters*, 7(2):366–370, April 2010.

[31] Y. Yang and S. Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *ACM SIGSPATIAL GIS*, 2010.

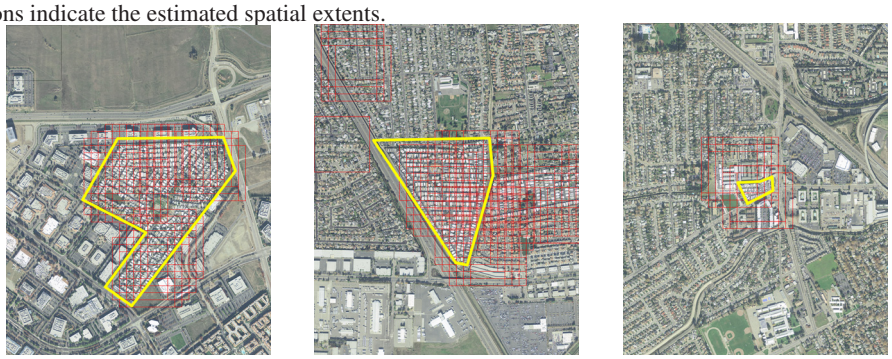
[32] C. Zhang. Towards an operational system for automated updating of road databases by integration of imagery and geodata. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4):166–186, 2004.

(a) (b) (c)

Figure 4. Three sample images of high schools. The yellow polygons indicate the manually delineated ground truth spatial extents. The unions of the red regions indicate the estimated spatial extents.

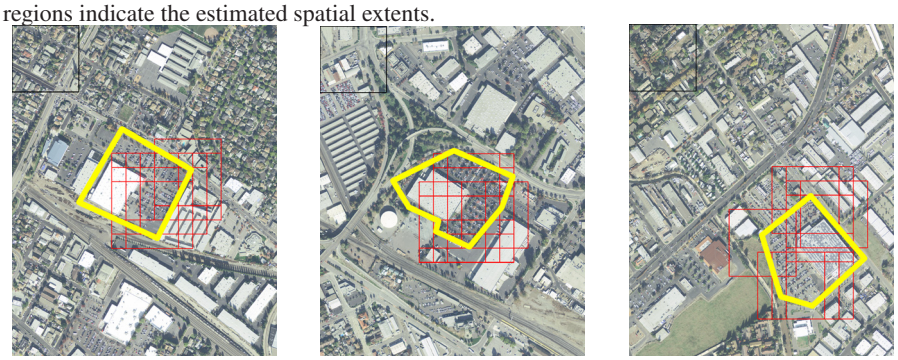
(a) (b) (c)

Figure 5. Three sample images of golf courses. The yellow polygons indicate the manually delineated ground truth spatial extents. The unions of the red regions indicate the estimated spatial extents.



(a) (b) (c)

Figure 6. Three sample images of mobile home parks. The yellow polygons indicate the manually delineated ground truth spatial extents. The unions of the red regions indicate the estimated spatial extents.



(a) (b) (c)

Figure 7. Three sample images of Costco shopping centers. The yellow polygons indicate the manually delineated ground truth spatial extents. The unions of the red regions indicate the estimated spatial extents.