Integrating Gazetteers and Remote Sensed Imagery

Shawn Newsam
Electrical Engineering and Computer Science
University of California
Merced, CA 95343
snewsam@ucmerced.edu

Yi Yang
Electrical Engineering and Computer Science
University of California
Merced, CA 95343
yyang6@ucmerced.edu

ABSTRACT

This work explores the potential for increased synergy between gazetteers and high-resolution remote sensed imagery. These two data sources are complementary. Gazetteers provide high-level semantic information about what is where but they must be manually compiled and maintained. On the other hand, imagery can be automatically acquired but only provides low-level radiometric information. We explore ways in which these two data sources can be integrated to more fully automate geographic data management. In particular, we show how gazetteers represent a rich source of semi-supervised training data for geospatial object modelling. We also describe an example of information flow in the other direction, namely, how high-resolution imagery can be used to refine the spatial extents of geospatial objects in gazetteers.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*spatial databases and GIS*; J.2 [**Physical Science and Engineering**]: Earth and Atmospheric Sciences; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*object recognition*

General Terms

Information Integration, Appearance Modeling

Keywords

Remote sensed imagery, gazetteers, appearance models, geospatial objects, interest points, Markov random fields

1. INTRODUCTION

Thanks to advances in technology, our ability to capture and store remote sensed imagery, such as that taken from satellite or aerial platforms, continues to improve. Unfortunately, our ability to analyze this imagery has not scaled proportionally and so automated methods are needed to realize the true value of this data. One approach that is proving particularly promising for making progress on automated image understanding is to leverage (non-image) data associated with the images. The nature of this meta-data varies not only in the richness of its description but also in how directly it is related to the image content. Extremes range from a well-annotated set of images in which keywords are associated with specific image regions, to the co-occurrence of large sections of text and multiple images in the same document without explicit correspondences, such as on a Web page.

Remote sensed imagery is a particulary interesting category of images to be considered in this context since geographic location enables a great variety of meta-data to be associated with a georeferenced image. In this paper, we explore one such association, that of integrating gazetteers and high-resolution satellite imagery. We describe novel methods for how these complementary data sources can be combined to more fully automate geographic data management. We first describe how gazetteers can serve as a source of semi-supervised training data for geospatial object modelling. We then describe how high-resolution imagery can be used to refine the spatial extents of items in the gazetteer.

2. RELATED WORK

Fully automated image understanding based only on the image content–the pixel values–is likely to remain an unsolved problem. Even the most effective system we know of, the human visual system, relies heavily on experience, context, and other information external to an image. It makes sense therefore that machine vision systems should leverage non-image data sources wherever possible. Our increased ability to acquire, store, and aggregate data, along with the tremendous growth of the Internet, has made a great variety of such data sources available. Over the past decade, computer vision researchers have developed novel ways in which to fuse image and non-image data to advance automated image understanding.

Content based image retrieval (CBIR) is limited by the semantic gap between the low-level image features, such as color and texture, and the kind of higher-level annotation needed to support meaningful search. Researchers have explored various ways to overcome this by incorporating other sources of information. CBIR systems for the World Wide Web (WWW) often query associated text, such as an image's ALT tag or text near the image, in addition to the low-level image features. The CORTINA system of Quack

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ACM GIS'08 , November 5-7, 2008. Irvine, CA, USA
Copyright 2008 ACM ISBN 978-1-60558-323-5/08/11 ...$5.00.

et al. [25] combines text and image context to search over 3 million images on the WWW. Other systems exploit the structure of the WWW, such as Newsam et al.'s category based image retrieval system [22] which makes use of the hierarchical structure of the WWW pages containing the images in addition to associated text. Other approaches enable keyword search by tackling the more difficult problem of assigning keywords to images. Barnard and Forsyth in [9] learn the statistical associations between image regions and words in a large corpus of annotated images. The learned associations can then be used to annotate new images. (Interestingly, this association also enables the complementary task of auto-illustration.)

Computer vision researchers have exploited various forms of meta-data associated with image collections to learn visual object models. Berg et al. in [11] 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. in [8] 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. in [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.

Closer to the work presented in this paper is the idea of using spatial and temporal meta-data to organize personal photo collections. Researchers have explored ways in which access to large collections of snapshots can be improved by classifying and grouping photographs using not only the time stamps of individual photographs but also the temporal spacing between photographs [12]. The advent of hand held global positioning systems (GPS) and even on-camera GPS now allows photographs to be organized geospatially. While research continues in this area, such as Yahoo Research's ZoneTag project [20, 3], many of the online photo sharing portals, such as Flickr [2], already offer this capability.

Researchers working in the geographic information sciences have 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 [28, 5, 13]. 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 [10]. Automated building extraction is another appealing use of remote sensed imagery. Agouris et al. [4] 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.

There has been research effort on using non-image GIS data to recognize a broader variety of object classes but largely without implementation or experimental results. Bailloeul et al. in [6] describe the theoretical aspects of a system which uses a priori knowledge in form of outdated urban maps to control contour- and region-based segmentation of new imagery. And, Baltsavias in [7] discusses a wide range of ways in which "knowledge" can improve image analysis but, as he indicates, his definition of non-image information is extremely broad and includes concepts such as rules, models, and context, in additional to specific GIS data. (He notes that very few of the works he surveyed use priori knowledge in the form of maps, GIS or other geodatabases.) Walter and Fritsch in [26] do provide results from using GIS data to automatically derive appearance models which are then applied to imagery to verify the GIS data but their application is at the level of land use classification.

The work presented in this paper is different from previous approaches in that it leverages non-image GIS data to model a much broader range of object classes. We achieve this by integrating an extensive gazetteer, which catalogs over 200 object classes, with state-of-the-art approaches to object recognition.

Note that the discussion above, and the context of this paper, concerns techniques for integrating multiple data sources, one of which is unprocessed image data. This is different from approaches which assume the image analysis has already been carried out such as the interesting recent work by Michalowski et al. [18] on combining street vector data, phone-book records and building outlines to map postal addresses.

3. DATA SOURCES

This section describes the two data sources being considered for integration, gazetteers and high-resolution remote sensed imagery.

3.1 Gazetteers

A gazetteer is a geographic directory. It contains records indicating what-is-where on the surface of the earth. The what varies greatly both in terms of the class of object as well as its physical characteristics, such as its spatial extent. While the minimal set of fields is a name and a point location, gazetteers typically include both proper names (e.g., San Francisco Internal Airport) and object classification (e.g., Airport), more complex spatial extents, such as a bounding box or higher-order polygon, and relations to other records (e.g., part of San Francisco County).

Gazetteers are not a new concept but advances in information technology have enabled them to become more extensive, thanks to automated aggregation, and more accessible, thanks to the Internet, than ever before. In this work, we focus on the Alexandria Digital Library (ADL) gazetteer [1] which is part of the ADL project at the University of California at Santa Barbara. The ADL gazetteer is an exemplar of modern gazetteers. It was created by aggregating the United States Geological Survey's *Geographic Names Information System* (GNIS) and the National Imagery and Mapping Agency's *Geographic Names Processing System* (GNPS) [15]. It has an online browser-based map interface for interactive querying. More important for our work is the ADL Gazetteer Service Protocol which supports remote query and response functions using standard HTTP XML-formatted requests.

The ADL gazetteer contains almost six million records and thus represents one of the most extensive and diverse collections of its kind. It covers the entire world although due to the nature of its original sources, its coverage is more complete for the US. It catalogs over 200 different classes of geographic features. Further, a feature class thesaurus maps almost 1,000 non-preferred terms to these primary

classes. The feature classes are organized hierarchically under the following six root classes: administrative areas, hydrographic features, land parcels, manmade features, physiographic features, and regions. Table 5 contains a (slightly out-of-date) list of the primary feature classes and their total counts.

While the ADL gazetteer is an impressive collection of geographic data, it has several shortcomings. Some of these could potentially be compensated for by advances in automated image analysis such as the techniques presented in this paper. At the moment, the spatial extent of the records is limited to a single point, a longitude/latitude pair. While the system includes provisions for at least a bounding box representation, this information was not present in the original sources and post-ingest manual specification is prohibitively expensive. As the ADL development team points out [15], "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." Section 5 of this paper presents results from our work on this important problem.

3.2 Remote Sensed Imagery

We believe that remote sensed image analysis is poised to undergo a paradigm shift thanks to the growing availability of wide area coverage submeter pixel resolution data. The domain is faced with the prospects and challenges of object level analysis on a scale not possible before. The higher resolution imagery greatly increases the variety of objects that are now observable, at least according to theoretical bounds such as Shannon's sampling theorem. This aligns the problem more closely with the well researched computer vision challenge of generic object recognition and allows it to leverage the many advances made over the past several decades in that area. In particular, it can leverage recent successes such as a new class of image descriptors, termed interest point descriptors, which have proven effective at modelling the appearances of diverse categories of objects viewed under widely varying conditions. In recent work [23, 24, 27], we investigated the application of this new class of descriptors to content based image retrieval and land cover classification in high-resolution remote sensed imagery with encouraging results. In section 4 below, we extend this investigation to geospatial object modelling.

Meter (e.g. IKONOS) and submeter (e.g. Quickbird) pixel resolution satellite imagery has been available commercially now for nearly a decade. GeoEye-1, which is scheduled to launch in August of 2008, will provide 0.41 meter/pixel imagery. This is in addition to the growing collection of aerial imagery which boasts even higher resolution. Automated image analysis is needed to realize the full potential of these information rich data sources. We believe the integration of gazetteers and remote sensed imagery represents a promising step in this direction.

4. GEOSPATIAL OBJECT MODELLING

This section describes a framework for using prior knowledge in the form of gazetteer records to learn appearance models for a variety of geospatial objects in an unsupervised fashion. It is one of two applications presented in this paper which leverage the integration of gazetteers and remote sensed imagery. The other application, on using appearance models to improve the spatial extent associated with geospatial objects in a gazetteer, is described in section 5.

This investigation proceeds as follows. First, we use the ADL gazetteer to locate and extract image regions corresponding to 13 different classes of geospatial objects in a large collection of IKONOS satellite imagery. We then extract high-dimensional image features from these regions and show they are clustered in the feature space, a potentially sufficient condition for learning object models. We then investigate classification strategies to further demonstrate that we are able to learn appearance models of geospatial objects in an unsupervised fashion through integration of a gazetteer and remote sensed imagery.

4.1 Data

Thanks to a generous grant from Lockheed Martin Corporation, we have access to a set of 16 georeferenced IKONOS satellite images which cover over 3,000 square kilometers of the continental US. While this imagery consists of one meter panchromatic band and four meter multispectral bands, we only utilize the panchromatic imagery. These IKONOS images are mostly of metropolitan areas and include coverage of Phoenix, Los Angeles, New York City, San Diego, San Jose, Washington DC, and El Paso.

We selected a subset of object classes cataloged by the ADL gazetteer (see table 5 for the full list) using the following criteria. First, the spatial extent of the objects must be limited enough for them to be contained within a single IKONOS image. This rules out larger objects such as counties. Second, the objects must be visible in the imagery. This obvious constraint rules out subterranean objects such as caves. Third, the visual characteristics of the object taken as a whole must be distinctive. This allows different object types to share some visual features but requires that there should be some way of discriminating between them visually. Finally, since most of our imagery is of urban/suburban regions, we focused on manmade objects. This selection process resulted in the 13 geospatial object classes listed in table 1.

We then used the ADL Gazetteer Service Protocol to identify instances of these 13 geospatial object classes in our IKONOS imagery. This was achieved by issuing XML formatted queries requesting all instances of a particular object class contained within the spatial footprint of each of the images separately. Table 1 lists the total number of instances identified in all 16 images. Table 1 also indicates the class ID assignments that will be used in subsequent tables.

As mentioned earlier, a shortcoming of the ADL gazetteer (and most gazetteers) is that it specifies the spatial extents of the objects as only a single point location. Therefore, even though we can use the gazetteer to identify object instances in the IKONOS imagery, we know very little about which region in the image actually corresponds to the objects. Without additional information, the best we can do is extract a square subimage centered at the gazetteer specified location and assume the object is contained somewhere in this subimage. At the moment, the size of these subimages is set manually and varies from 300 by 300 pixels for heliports

Table 1: Appearance models were learned for the following geospatial object classes. The column labelled count refers to the total number of examples in the IKONOS imagery as found by querying the ADL gazetteer.

ID	Object Class	Count
1	Airport Features	24
2	Cemeteries	40
3	Country Clubs	12
4	Educational Facilities	694
5	Golf Courses	23
6	Harbors	6
7	Heliports	50
8	Medical Facilities	70
9	Mobile Home Parks	88
10	Parks	274
11	Railroad Features	9
12	Religious Facilities	126
13	Shopping Centers	147

to 1000 by 1000 pixels for airports with most being 600 by 600 pixels. A better estimate of the size of the subimages for different object classes could be obtained using a dataset with spatial extents. This would reduce the amount of background (non-object) regions in the subimages and therefore improve the appearance modelling.

Figure 1 contains two sample subimages from four object classes: educational facilities, golf courses, mobile home parks, and shopping centers. Again, these subimages were extracted in a completely automated fashion by cropping a fixed sized region from an IKONOS image centered at the point location specified by the ADL gazetteer. These are some of the better samples in that 1) the point locations are actually within the objects; 2) the ratio of the object to background is quite high; and 3) the objects are good visual examples of their classes. Overall, the subimages form a fairly noisy dataset. Frequently, the ADL specified point location is not within the object due to inaccuracies of the original gazetteer data sources and/or the georeferencing process. We can compensate for this by extracting larger regions but this results in a lower object to background ratio. There can also be significant variance visually between the samples for a particular class. The less-than-perfect nature of the subimage dataset make the object modelling task an interesting challenge.

4.2 Image Features

Our visual appearance models are based on a new class of image features termed interest point descriptors. These descriptors seek to capture the visually salient regions of an image under the assumption that it is these regions that will be the most discriminative for an image analysis task. Extracting interest point descriptors involves first using a saliency detector to identify the salient locations in an image and then describing the image patches at these locations using local descriptors. We use David Lowe's Scale Invariant Feature Transform (SIFT) descriptors [17] as they have been shown to be robust to image rotation and scale, and to be capable of matching images with geometric distortion and varied illumination. They have shown to outperform other local descriptors in an image matching task [19], and, in previous work [23, 24, 27], we demonstrated their effectiveness for content based image retrieval and land cover classification in high resolution remote sensed imagery.

We use the standard SIFT interest point detector and the standard histogram-of-gradients descriptor. These 128 dimension descriptors can be thought of roughly as summarizing the edge information in an image patch centered at an interest point. SIFT features can be quite dense spatially especially in complex imagery. Each of our subimages ends up being described by 6,313.8 descriptors on average.

We use a bag-of-words approach from document analysis to compactly summarize the SIFT features of a subimage. First, we construct a feature codebook by applying k-means clustering to a large collection of SIFT features selected at random from the IKONOS imagery. We then quantize the SIFT features in our subimages by assigning each one the label of its closest cluster center. Finally, we compute a single frequency feature vector for each subimage as

$$SIFT_{freq} = [t_0, t_1, \ldots, t_{k-1}] \ , \tag{1}$$

where t_i is number of occurrences of the quantized SIFT features with label i and k is the size of the quantization codebook. $SIFT_{freq}$ is similar to a term vector in document retrieval. The frequency features are scaled so their components sum to one to account for the varying number of SIFT descriptors per subimage. We use a codebook with 50 entries (i.e., $k = 50$ in the k-means clustering) as our earlier work [24] showed this number was optimal for an image retrieval task.

The cosine distance measure has shown to be effective for comparing documents represented by term vectors [14] so we use it here to compute the similarity between images. The similarity between a query image Q with counts $[q_0, \ldots, q_{k-1}]$ and a target image T with counts $[t_0, \ldots, t_{k-1}]$ is computed as

$$D_{freq}(Q,T) = \frac{\sum_{i=0}^{k-1} q_i t_i}{\sqrt{\sum_{i=0}^{k-1} q_i^2} \sqrt{\sum_{j=0}^{k-1} t_j^2}} \ . \tag{2}$$

The cosine distance measure ranges from zero (not similar) to one (very similar or perfect match).

We first investigate whether the frequency features belonging to the different object classes form clusters in the space induced by the cosine distance. We compare the average within cluster feature distance to the average between cluster feature distance. The results of this analysis are shown in table 2. We observe that the features for most classes are more similar to themselves than to the features for other classes.

4.3 Classification

We used a classification framework to further show that we are learning discriminative appearance models for the 13 object classes. We computed average classification rates using a leave-one-out approach in which we train classifiers using all the features for an object type except one selected at random which is later used as the test feature. We explored two types of classifiers, a simple nearest centroid classifier and a maximum likelihood classifier.

The nearest centroid classifier assigns a test feature to the object class with the closest centroid as computed using the

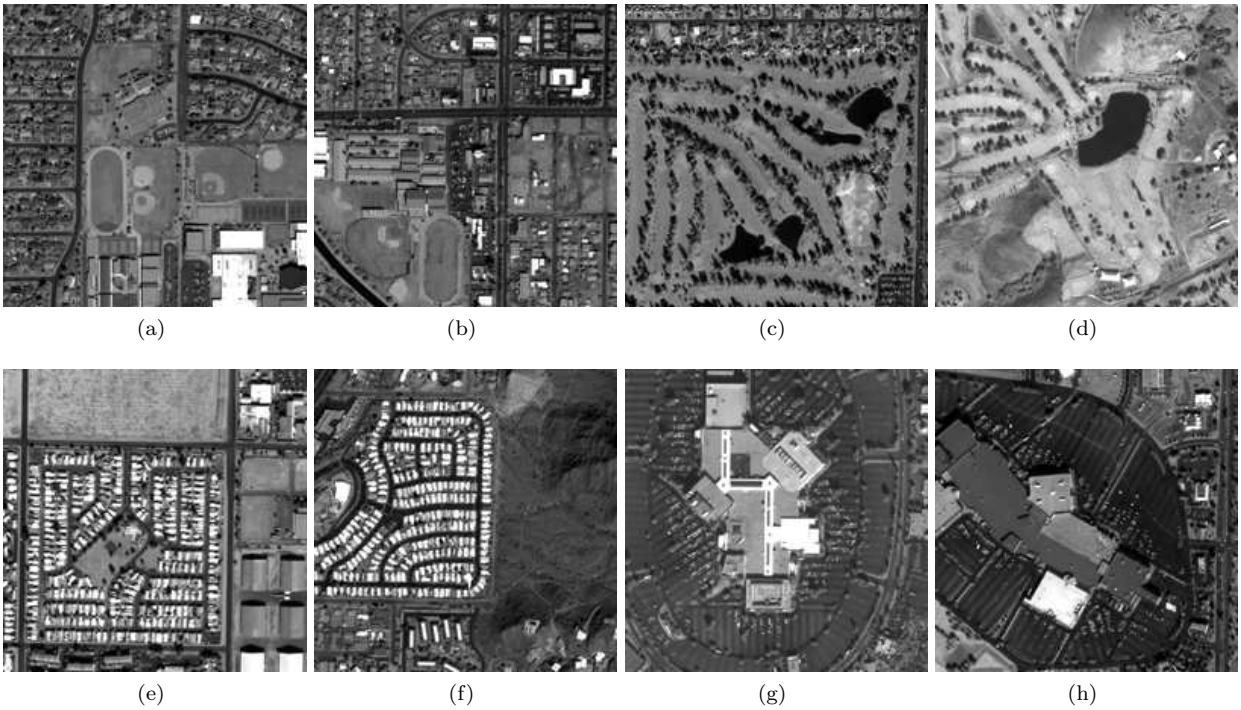


Figure 1: Sample subimages of different object classes extracted in an unsupervised fashion from IKONOS imagery using the ADL gazetteer. Two samples of each of the following classes are shown in scan order: educational facilities, golf courses, mobile home parks, and shopping centers.

cosine distance. This centroid is simply the average of the training features for an object type. A separate classifier is trained for each class. Table 3 shows the confusion matrix for 1,500-fold cross validation. The classification rates for different object classes vary significantly from a low of 0.02 for parks to 0.69 for mobile home parks, with several above 0.50. Overall, the fact that we are doing much better than chance (0.08) confirms we are learning discriminative appearance models, at least for most of the classes.

We also performed classification using a maximum likelihood classifier to better incorporate the spread of the features for an object class. We modelled the distribution of the features for each class using a von Mises-Fisher (vMF) probability density function. A vMF density can be thought of as a normal density for points on a unit hyper-sphere which is appropriate for our features since they are directional (thus compared using the cosine distance). The probability density function of a vMF distribution has the following form

$$p(x|\mu,\kappa) = c_d(\kappa)e^{\kappa\mu^T x}\ , \tag{3}$$

where d is the dimension, and μ, the distribution "mean," and κ, the distribution spread, are the parameters. We again use a leave-one-out approach in which a separate classifier is learned for each class by estimating μ and κ from the set of features with one left out. The left out samples are then classified as the object class which maximizes their likelihood:

$$c^* = \arg \max_{c \in allclasses} p(x_{sample}|\mu_c,\kappa_c)\ . \tag{4}$$

The results of performing a 1,500-fold cross validation are shown in table 4. Again, though the classification rates vary significantly between classes, several of the rates are above 0.5 confirming we are learning discriminative appearance models.

4.4 Discussion

This section described a fully automated technique for learning visual appearance models for geospatial object classes by integrating the ADL gazetteer and high-resolution IKONOS imagery. Preliminary exploration of the class feature distributions and on training both a nearest centroid and a maximum likelihood classifier indicates we are learning discriminative models even though the training images are extremely noisy.

We find the two classifiers perform quite differently. In particular, the maximum likelihood classifier, which accounts for the variation in feature spread between classes, completely miss-classifies educational facilities, harbors, railroad features, religious facilities, and shopping centers. These objects instead are predominantly classified as medical facilities. Further investigation here is needed although we observed that the estimated von Mises-Fisher distributions for these classes are more peaked (the value of κ is larger) and thus classes with wider distributions, such as medical facilities, can "sweep-up" noisy instances.

We are continuing work on this problem. We plan to use feature selection algorithms to determine which of the quantized SIFT features are the most discriminative for different classes. We also plan on extending the appearance models to capture the spatial arrangements of the features as we expect this to be a discriminative characteristic of geospatial objects.

Table 2: The features belonging to different object classes form clusters in the feature space. The value at row i, column j is the average distance of the features for object class (ID) i to the features for object class j. The higher the value, the more similar the features. Note that the features of most object classes are more similar to themselves than to the features of other object classes. The entries on the diagonal are emphasized for readability.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	**0.854**	0.837	0.813	0.847	0.814	0.859	0.835	0.851	0.848	0.839	0.878	0.839	0.870
2	0.837	**0.926**	0.902	0.922	0.896	0.896	0.853	0.916	0.896	0.913	0.920	0.919	0.907
3	0.813	0.902	**0.891**	0.896	0.884	0.873	0.836	0.892	0.873	0.889	0.894	0.895	0.885
4	0.847	0.922	0.896	**0.938**	0.878	0.912	0.865	0.932	0.917	0.924	0.933	0.938	0.929
5	0.814	0.896	0.884	0.878	**0.891**	0.863	0.830	0.876	0.854	0.875	0.885	0.875	0.869
6	0.859	0.896	0.873	0.912	0.863	**0.938**	0.869	0.915	0.882	0.903	0.925	0.924	0.921
7	0.835	0.853	0.836	0.865	0.830	0.869	**0.841**	0.869	0.861	0.857	0.882	0.864	0.880
8	0.851	0.916	0.892	0.932	0.876	0.915	0.869	**0.928**	0.912	0.919	0.931	0.934	0.928
9	0.848	0.896	0.873	0.917	0.854	0.882	0.861	0.912	**0.928**	0.902	0.916	0.911	0.920
10	0.839	0.913	0.889	0.924	0.875	0.903	0.857	0.919	0.902	**0.912**	0.922	0.925	0.915
11	0.878	0.920	0.894	0.933	0.885	0.925	0.882	0.931	0.916	0.922	**0.949**	0.931	0.937
12	0.839	0.919	0.895	0.938	0.875	0.924	0.864	0.934	0.911	0.925	0.931	**0.948**	0.930
13	0.870	0.907	0.885	0.929	0.869	0.921	0.880	0.928	0.920	0.915	0.937	0.930	**0.938**

Table 3: The confusion matrix of classification using a closest centroid classifier. 1,500-fold cross validation is performed using a leave-one-out approach. The value at row i, column j indicates the average classification rate that a sample of object class i is classified as class j. The entries on the diagonal are emphasized for readability. These results confirm that we are learning discriminative appearance models for most object classes.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	**0.58**	0.04	0.00	0.09	0.04	0.03	0.00	0.00	0.05	0.04	0.08	0.00	0.04
2	0.00	**0.34**	0.02	0.03	0.19	0.03	0.00	0.02	0.06	0.05	0.02	0.20	0.04
3	0.00	0.00	**0.16**	0.00	0.34	0.00	0.00	0.00	0.26	0.00	0.00	0.08	0.15
4	0.04	0.08	0.00	**0.27**	0.01	0.03	0.01	0.02	0.15	0.03	0.06	0.25	0.05
5	0.00	0.08	0.04	0.00	**0.56**	0.04	0.00	0.00	0.00	0.13	0.00	0.03	0.12
6	0.00	0.00	0.00	0.00	0.00	**0.67**	0.16	0.00	0.00	0.00	0.00	0.17	0.00
7	0.13	0.00	0.00	0.04	0.10	0.13	**0.33**	0.02	0.06	0.03	0.06	0.06	0.06
8	0.06	0.09	0.01	0.11	0.03	0.09	0.05	**0.14**	0.10	0.03	0.02	0.20	0.08
9	0.01	0.05	0.01	0.07	0.00	0.03	0.03	0.02	**0.69**	0.01	0.01	0.00	0.08
10	0.02	0.16	0.01	0.17	0.08	0.10	0.02	0.04	0.08	**0.02**	0.07	0.18	0.05
11	0.21	0.00	0.00	0.00	0.11	0.00	0.00	0.11	0.00	0.11	**0.22**	0.12	0.11
12	0.00	0.05	0.01	0.08	0.02	0.18	0.01	0.03	0.00	0.03	0.05	**0.54**	0.01
13	0.07	0.02	0.00	0.06	0.01	0.06	0.08	0.04	0.22	0.01	0.06	0.11	**0.27**

Table 4: The confusion matrix of classification using a maximum likelihood classifier. The features of each object class are modelled using a von Mises-Fisher distribution. 1,500-fold cross validation is performed using a leave-one-out approach. The value at row i, column j indicates the average classification rate that a sample of object class i is classified as class j. The entries on the diagonal are emphasized for readability. These results confirm that we are learning discriminative appearance models for most object classes.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	**0.63**	0.04	0.00	0.00	0.04	0.00	0.02	0.22	0.04	0.01	0.00	0.00	0.00
2	0.00	**0.45**	0.00	0.00	0.14	0.00	0.01	0.21	0.05	0.13	0.00	0.00	0.00
3	0.00	0.17	**0.01**	0.00	0.36	0.00	0.00	0.08	0.33	0.05	0.00	0.00	0.00
4	0.04	0.12	0.00	**0.00**	0.01	0.00	0.03	0.50	0.16	0.13	0.00	0.00	0.00
5	0.00	0.13	0.06	0.00	**0.49**	0.00	0.00	0.18	0.05	0.09	0.00	0.00	0.00
6	0.21	0.00	0.00	0.00	0.00	**0.00**	0.29	0.34	0.00	0.16	0.00	0.00	0.00
7	0.30	0.02	0.02	0.00	0.07	0.00	**0.35**	0.11	0.07	0.06	0.00	0.00	0.00
8	0.05	0.13	0.02	0.00	0.01	0.00	0.04	**0.52**	0.17	0.06	0.00	0.00	0.00
9	0.01	0.06	0.00	0.00	0.00	0.00	0.09	0.06	**0.76**	0.02	0.00	0.00	0.00
10	0.02	0.19	0.01	0.00	0.05	0.00	0.07	0.45	0.10	**0.10**	0.00	0.00	0.00
11	0.23	0.19	0.00	0.00	0.01	0.00	0.00	0.51	0.00	0.05	**0.00**	0.00	0.00
12	0.00	0.09	0.00	0.00	0.01	0.00	0.07	0.68	0.00	0.15	0.00	**0.00**	0.00
13	0.08	0.03	0.00	0.00	0.00	0.00	0.08	0.44	0.33	0.03	0.00	0.00	**0.00**

5. GAZETTEER MAINTENANCE

This section describes our earlier work [21] on using appearance models to improve the spatial extent associated with objects cataloged by gazetteers. Again, most gazetteers, including the ADL gazetteer, only contain single point locations so improving this even to a bounding box, which would still an approximation for most instances of an object, would be a significant improvement. The proposed technique is another example of integrating gazetteers and remote sensed imagery, in this case to provide the means for automatically maintaining, both constructing and updating, a gazetteer.

A summary of the approach is as follows. We learn the appearance models for different object classes from a set of training images. This could be done in an unsupervised fashion using the framework described in the previous section although we here use a set of training images in which the object boundaries have been manually delineated. We then use these models to estimate the bounding box for object instances whose boundaries are not known. Once again, gazetteer information is used to improve the image analysis since the gazetteer point location is used to seed an iterative process which grows the bounding box optimally with respect to the learned appearance model until a stopping criterion is met. Thus, we are not attempting the more difficult problem of detecting and localizing object instances; instead, we focus on the simpler but still important task of localizing an instance given an internal point.

5.1 Appearance Models

The appearance models used here are different from the ones described in the previous section. They are based on image texture characterized by the outputs of orientation and scale selective spatial filters. High-dimensional texture feature vectors formed from the first- and second-order statistics of the filter outputs are quantized into texture motifs using a codebook based on the clustering of features from known object instances using a Gaussian mixture model expectation maximization framework (GMMEM). The spatial arrangement of the texture motifs for a particular object type is modelled as a multi-level logistic model (MLL) Markov random field (MRF) whose parameters are learned using Markov chain Monte Carlo (MCMC) techniques again using known object instances.

The motivation behind this hierarchical model is as follows. Object classes that we wish to discriminate between might share the same low-level texture motifs albeit in different spatial arrangements. For example, both golf courses and parks share grass and trees as their motifs but these two classes can be distinguished based on the the spatial arrangement of these motifs. Our appearance model captures the spatial arrangement of an object's constituent motifs using a Markov random field.

5.2 Bounding Box Estimation

We use the appearance model to estimate the bounding box of a known object instance as follows. First, the box is initialized as a single pixel using the seed location provided by the gazetteer. We then grow the box in an iterative fashion by extending the side abutting the region most like the appearance model. Specifically, at each step, we use MCMC to estimate the MLL parameters of the rectangular regions immediately above, below, to the left, and to the right of the current bounding box. We then add a row or column of pixels from the region whose parameters are most similar to those of the appearance model for the object class being considered. Similarity is measured using the L1 distance. A stopping criterion is used to halt the growing process. We here stop the process when the distances between the MLL parameters of all four adjacent regions and the model are above a pre-specified threshold.

Figure 2 shows the growth of bounding boxes for a mobile home park and a golf course. The image regions outside the object masks are dimmed to indicate the manually specified spatial footprints that serve as the ground truth. The bounding boxes are shown at 75-iteration intervals from random starting points until the stopping criterion is met. These examples demonstrate that the proposed technique effectively estimates the spatial extent of the objects even when the starting points are near the boundaries.

Receiver operating characteristic (ROC) curves can be used to quantitatively evaluate the expansion of the bounding box. Using a manually delineated boundary for the test image, the true-positive ratio can be computed as the percentage of the object that is contained in the bounding box, and the false-positive ratio as the percentage of the bounding box that is not part of the object. Figures 3(a) and 3(b) show the ROC curves averaged over five mobile home park images and five golf course images, respectively. The proposed technique is shown to maintain a high true-positive to false-positive ratio and outperforms a naive approach that begins at the same random location but does not consider the MRF model and uniformly expands the sides of the bounding box.

5.3 Discussion

This section described a method for estimating the bounding box of geospatial objects using high-resolution remote sensed imagery. Visual appearance models were used to optimally grow a seed location internal to the object. Future work on this problem includes investigating other appearance models such as the one described in section 4 above and dealing with the case when the seed location does not fall within the object due to inaccuracies in the georeferencing and/or the location source (such as a gazetteer).

6. CONCLUSION

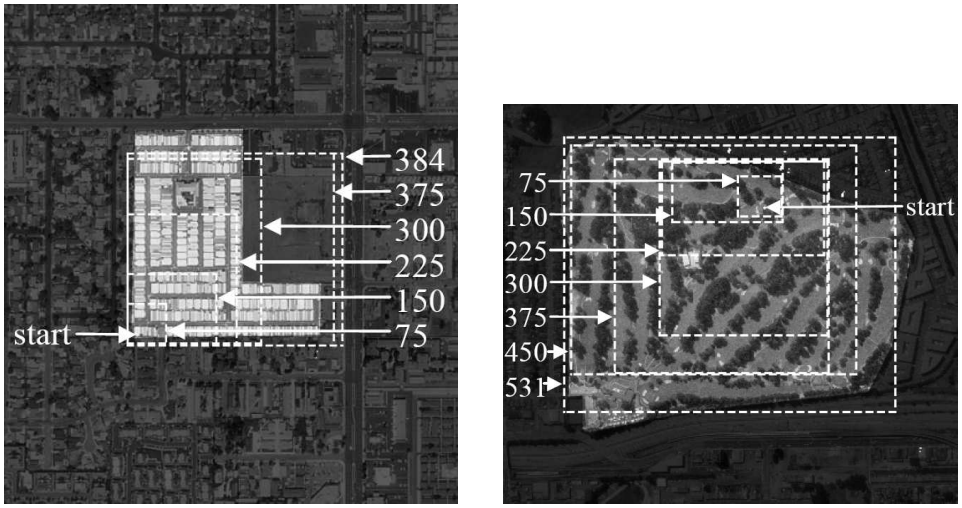
This work explored the potential for increased synergy between gazetteers and remote-sensed imagery. We described two specific ways in which these complementary data sources can be integrated to more fully automate geographic data management. We first presented a novel approach to using gazetteers as a source of semi-supervised training data for appearance based modelling of geospatial objects. We then demonstrated how high-resolution imagery can be used to refine the spatial extents of geospatial objects in gazetters.

7. ACKNOWLEDGEMENTS

This work was funded in part by a Department of Energy Office of Defense Programs Early Career Scientist and Engineer Award.

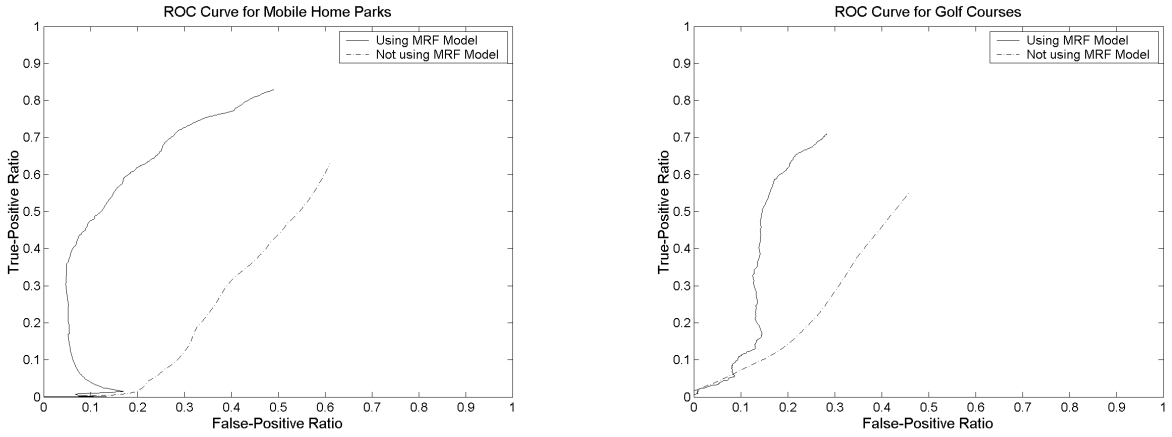
8. REFERENCES

[1] Alexandria Digital Library Gazetteer. 1999- . Santa Barbara CA: Map and Imagery Lab, Davidson Library, University of California, Santa Barbara.



(a) (b)

Figure 2: Bounding box every 75 iterations until stopping for (a) mobile home park, and (b) golf course.



(a) (b)

Figure 3: ROC curves for (a) mobile home parks, and (b) golf courses.

Copyright UC Regents. http://www.alexandria.ucsb.edu/gazetteer.

[2] Flickr photo sharing. http://www.flickr.com.

[3] Zonetag research prototype. http://zonetag.research.yahoo.com.

[4] 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.

[5] 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.

[6] T. Bailloeul, J. Duan, V. Prinet, and B. Serra. Urban digital map updating from satellite high resolution images using GIS data as a priori knowledge. In *Proceedings of the GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pages 283–287, 2003.

[7] E. P. Baltsavias. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4):129–151, 2004.

[8] K. Barnard, P. Duygulu, and D. Forsyth. Clustering art. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 434–441, 2001.

[9] K. Barnard and D. Forsyth. Learning the semantics of words and pictures. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2, pages 408–415, 2001.

[10] 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.

[11] 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 *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 848–854, 2004.

[12] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. Temporal event clustering for digital photo collections. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 364–373, 2003.

[13] 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*, pages 20–33, 1999.

[14] D. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. The MIT Press, 2001.

[15] 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.

[16] L.-J. Li, G. Wang, and L. Fei-Fei. Optimol: automatic Online Picture collecTion via Incremental MOdel Learning. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

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

[18] M. Michalowski, C. A. Knoblock, K. Bayer, and B. Y. Choueiry. Exploiting automatically inferred constraint-models for building identification in satellite imagery. In *Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, pages 35–42, 2007.

[19] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[20] M. Naaman. Eyes on the world. *Computer*, 39(10):108–111, Oct. 2006.

[21] S. Newsam, S. Bhagavathy, and B. S. Manjunath. Object localization using texture motifs and markov random fields. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 1049–1052, 2003.

[22] S. Newsam, B. Sumengen, and B. Manjunath. Category-based image retrieval. In *Proceedings of the IEEE International Conference on Image Processing*, volume 3, pages 596–599, 2001.

[23] S. Newsam and Y. Yang. Comparing global and interest point descriptors for similarity retrieval in remote sensed imagery. In *Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, 2007.

[24] S. Newsam and Y. Yang. Geographic image retrieval using interest point descriptors. In *Proceedings of the International Symposium on Visual Computing (ISVC)*, 2007.

[25] T. Quack, U. Mönich, L. Thiele, and B. S. Manjunath. Cortina: a system for large-scale, content-based web image retrieval. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 508–511, 2004.

[26] V. Walter and D. Fritsch. Automatic verification of GIS data using high resolution multispectral data. *International Archives of Photogrammetry and Remote Sensing*, 32(III/1):485–490, 1998.

[27] Y. Yang and S. Newsam. Comparing SIFT descriptors and Gabor texture features for classfication of remote sensed imagery. In *Proceedings of the IEEE International Conference on Image Processing*, 2008 (accepted for publication).

[28] 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.

Table 5: Object classes indexed by the ADL gazetteer along with counts.

Class	Count	Class	Count	Class	Count
administrative areas	2,126,610	cemeteries	64,535	dunes	5,270
military areas	813	disposal sites	247	flats	4,722
parks	20,408	fisheries	45	gaps	15,762
political areas	32,623	fortifications	2,471	isthmuses	79
countries	165	historical sites	66,228	karst areas	113
countries, 1st order div	3,328	archaeological sites	1,654	ledges	644
countries, 2nd order div	14,602	hydrographic structures	123,991	mesas	5,529
countries, 3rd order div	13,115	breakwaters	101	mineral deposit areas	667
countries, 4th order div	1,330	canals	21,482	moraines	39
populated places	2,000,821	dam sites	45,828	mountains	362,194
cities	273	harbors	5,250	mountain ranges	526
capitals	271	levees	726	mountain summits	21,401
reference locations	42,597	marinas	45	ridges	21,916
reserves	4,704	offshore platforms	58	natural rock formations	2,154
tribal areas	4,183	piers	403	arches (natural formation)	446
hydrographic features	636,5 64	reservoirs	49,976	plains	7,171
bays	34,898	waterworks	82	plateaus	1,165
fjords	1,656	landmarks	545	playas	4
channels	13,874	mine sites	24,070	reefs	8,270
deltas	66	monuments	8,560	seafloor features	2,112
drainage basins	208	oil fields	4,834	continental margins	140
estuaries	424	recreational facilities	7,526	ocean trenches	252
floodplains	4	amusement parks	18	seamounts	1,025
gulfs	420	camps	3,744	submarine canyons	590
guts	932	performance sites	312	tectonic features	476
ice masses	3,569	sports facilities	1,948	earthquake features	345
glacier features	3,492	storage structures	11,969	faults	127
lakes	94,758	telecom features	12,240	fracture zones	123
seas	273	towers	14,423	folds (geologic)	4
oceans	42	transportation features	77,933	anticlines	4
ocean currents	24	airport features	24,814	valleys	36,815
streams	480,921	heliports	3,802	canyons	18,706
rivers	15,907	seaplane bases	451	volcanic features	2,262
bends (river)	1,637	aqueducts	101	lava fields	391
rapids	3,095	bridges	4,709	volcanoes	1,819
waterfalls	4,945	locks	259	regions	120,132
springs (hydrographic)	3,016	parking sites	5	biogeographic regions	44,782
thermal features	407	pipelines	186	barren lands	11
land parcels	12,424	railroad features	34,029	deserts	588
manmade features	858,145	roadways	578	forests	19,476
agricultural sites	174,912	trails	7,582	petrified forests	2
buildings	243,448	tunnels	761	woods	871
capitol buildings	29	wells	71,680	grasslands	4,419
commercial sites	2,711	windmills	171	habitats	56
industrial sites	1,465	physiographic features	575,964	oases	362
power generation sites	26	alluvial fans	66	shrublands	1,900
court houses	664	arroyos	45,359	snow regions	77
institutional sites	189,02	badlands	17	tundras	6
correctional facilities	32	banks (hydrographic)	2,014	wetlands	17,887
educational facilities	14,885	bars (physiographic)	12,167	coastal zones	165
medical facilities	5,558	basins	8,957	economic regions	8
religious facilities	78,26 3	storage basins	20	firebreaks	8
library buildings	2,134	beaches	3,584	land regions	70,094
museum buildings	886	bights	590	continents	3
post office buildings	11,16	capes	19,707	islands	70,088
research facilities	339	caves	966	map regions	2,311
data collection facilities	14	cirques	196	map quadrangle regions	2,311
residential sites	22,719	cliffs	5,557	research areas	195
housing areas	21,756	craters	435	uncategorized	20,719