

Object Discovery in High-Resolution Remote Sensing Images: A Semantic Perspective

Dihua Guo, Hui Xiong, Vijayalakshmi Atluri, Nabil R. Adam

Department of Management Science and Information Systems
CIMIC
Rutgers University, USA

Abstract. Given its importance, the problem of object discovery in High-Resolution Remote-Sensing (HRRS) imagery has received a lot of attention in the literature. Despite the vast amount of expert endeavor spent on this problem, more efforts have been expected to discover and utilize hidden semantics of images for object detection. To that end, in this paper, we address this problem from two semantic perspectives. First, we propose a semantic-aware two-stage image segmentation approach, which preserves the semantics of real-world objects during the segmentation process. Second, to better capture semantic features for object discovery, we exploit a hyperclique pattern discovery method to find complex objects that consist of several co-existing individual objects that usually form a unique semantic concept. We consider the identified groups of co-existing objects as new feature sets and feed them into the learning model for better performance of image retrieval. Experiments with real-world datasets show that, with reliable segmentation and new semantic features as starting points, we can improve the performance of object discovery in terms of various external criteria.

1. Introduction

With the advance of remote sensing technology and the increase of the public interest, the remote-sensing imagery has been drawing the attention of people beyond the traditional scientific user community. Large collections of High-Resolution Remote-Sensing (HRRS) images are becoming available to the public, from satellite images to aerial photos. More than 10 different proposed Earth observing satellite systems are orbiting in space, such as IKONOS, QuickBird,

Received Feb 05, 2008

Revised Jul 04, 2008

Accepted Jul 13, 2008

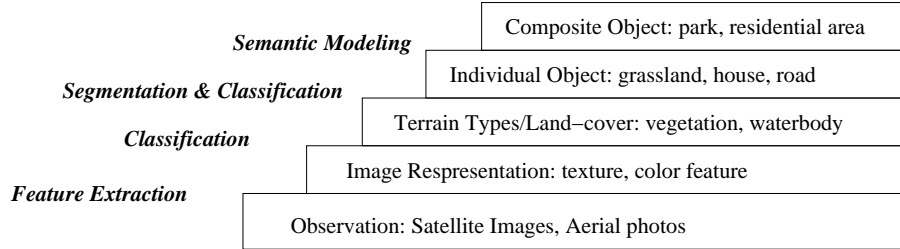


Fig. 1. Information Extracted from Remote-Sensing Images

and AISA, whose spatial resolutions are from 0.8 to 5 meters in the panchromatic bands. Different agencies, such as U.S. Geographic Survey (USGS) and New Jersey Meadowlands Commission (NJMC), provide aerial photos with increasing resolutions from as early as 1980s. The global and repetitive nature of these HRRS systems opens up unlimited opportunities for new products and new information derivable from imagery.

To date, the remote sensing image databases are only for browsing and fetching images according to metadata rather than retrieving images according to users' interest. Users are expected to have extensive domain knowledge. Therefore, the major user communities of remote-sensing images are military and the Earth scientists. To make the HRRS image retrieval more user friendly, we have to allow the users to access image databases through semantic-based queries in addition to metadata-based queries. However, it remains a challenging task to identify semantic-objects in HRRS images. While HRRS images share some common features with traditional images, they possess some special characteristics that make object discovery more complex. This motivates our research work.

Motivating Examples. Users are interested in different types of objects on the Earth as well as groups of objects with various spatial relationships. For example, let us consider Emergency Response Officers who are trying to find shelters to accommodate a large number of people. However, shelters are not distinguishable in Remote Sensing (RS) images. Instead, the officers could search for baseball fields, because most probably, a baseball field is connected to a school and the school could be used as a temporary shelter in emergency. In addition, qualified shelters should not be far away from water sources. Therefore, a query might be “*select all the baseball fields in Newark within 1 mile from any water body*”. Another interesting application domain would be urban planning. With HRRS image retrieval, we may have the task to find out “*the disinvestment area in the industrial area of Hudson county*”. This task indicates that we need to identify the industrial areas with a lot of empty lots. While traditional Content Based Image Retrieval (CBIR) techniques discover objects such as buildings and water bodies, these two examples demonstrate that one needs to discover *semantic* objects such as schools and urban areas from RS or HRRS images.

Based on the above observation, we categorize the target objects that can be recognized in RS or HRRS images into three levels of concepts: (1) Basic Terrain Types; (2) Individual Objects; and (3) Composite Objects. The first concept level is to distinguish the basic terrain type of the area covered by the images. There are several basic ground layouts: bare land, mountain, water, residential area, forest, etc. The second type of objects includes individual objects that

are recognizable in images, such as individual buildings, road segments, road intersections, cars, etc. Objects in the third concept level are composite objects that consist of several individual objects that form a new semantics concept. If we organize and represent all the information, we can retrieve objects from HRRS images, as shown in Figure 1.

While some promising work has been developed by researchers from different aspects (Duygulu, Barnard, de Freitas and Dorsyth, 2002; Fung and Stoeckel, 2007; Gupta, Weymouth and R.Jain, 1991; Mori, Takahashi and Oka, 1999; Feng, Manmatha and Lavrenko, 2004; Wang, Khan, Liu and Wu, 2004), our research focus is on identifying semantic-objects is composed of individual objects that form a new semantic concept. To fulfill our goal, we need to investigate the following two issues that specifically relate to HRRS images:

- How can we segment HRRS images appropriately?
- Can we identify semantic-objects by selecting semantic features?

Along this line, this paper addresses the problem of automatically annotating images using a relevance-based statistical model on HRRS images. Specifically, to answer the questions listed above, we propose a two-stage object-based semantic image segmentation approach that extends the hierarchical HRRS image segmentation algorithm (Guo, Atluri and Adam, 2005), as well as exploits the hyperclique pattern discovery method (Xiong, Tan and Kumar, 2003) to create new semantic features and feeds them into the relevance-based statistical learning model. The extended two-stage semantic segmentation approach avoids a general pitfall for image segmentation: using a single algorithm to solve the universal problem. The proposed approach takes the image semantics into consideration and segments images accordingly. In addition, hyperclique patterns have the ability to capture a strong connection between the overall similarity of a set of objects and can be naturally extended to identify co-existing objects in HRRS images. Traditionally, by using a training set of annotated images, the relevance-model (Jeon, Lavrenko and Manmatha, 2003) can learn the joint distribution of the blobs and words. Here, the blobs are image segments acquired directly from the image segmentation procedure. Our approach extends the meaning of blobs by identifying the co-existing objects/segments as new blobs. The proposed approach has been tested using the USGIS high-resolution orthology aerial images. Our experimental results show that, with reliable segmentation and new semantic features as starting points, the performance of learning model can be improved according to several external criteria.

Overview. The remainder of this paper is organized as follows: In Section 2, we describe the problems and challenges for object discovery in HRRS imagery. Section 3 reviews previous related research work. In Section 4, we introduce a two-stage image segmentation method and propose a semantic feature selection method. Section 5 provides experimental results. Finally, we draw conclusions and suggest future work in Section 6.

2. Domain Challenges

In this section, we identify some domain challenges for object discovery in HRRS images. Foremost, we should investigate HRRS data. We notice that, in addition to the challenges faced by the scenery digital image understanding systems,

HRRS images have some special characteristics, which pose additional challenges to image object discovery. These are discussed below.

- HRRS imagery is big. Here the word “big” has a two-fold meaning. First, every image is bigger than before since it captures more detail and has finer resolutions. Second, for HRRS image archive, the database is always getting bigger as users acquire new images. Thus, it is a challenging task to build a computationally efficient system.
- HRRS images contain tremendous details compared to low-resolution RS images. This increases the computational burden. A more difficult problem is how to distinguish useful details from trivial details.
- HRRS images often contain heavy salt-and-pepper noises. Indeed, existing object discovery algorithms often do not work well in noisy environments as they are initially designed for scenery images. Thus, there is a need to develop a new method that works well for noisy images.
- Spatial relationships among objects are critical to the semantics of HRRS images. Therefore, a semantic-based HRRS image retrieval system should take spatial relationships as image features in addition to regular color, texture, and shape features.

Based on the above characteristics of HRRS images, we summarize the domain challenges for object discovery in HRRS images as follows.

- First, it is nontrivial to do feature selection for image retrieval in HRRS images. In (Wang et al., 2004), researchers developed a mechanism to automatically assign different weights to different features according to the relevance of a feature to clusters in the Corel images. However, unlike Corel Image, HRRS images are severely affected by the noise such as shadow and the surface materials of HRRS images are limited. This makes the primitive features, such as color, texture and shape, not good enough for identifying objects in HRRS images. As a result, in addition to the primitive features, the derivative features, such as geometric features and semantic features, are required for better object discovery in HRRS images. In this research, we add semantic features that capture the spatial relationships among objects to image annotation model.
- Also, HRRS images usually lack salient regions and carry a lot of noise (Guo et al., 2005). This problem has been largely ignored by existing approaches, which makes them not suitable for object discovery in HRRS images. Indeed, existing methods often use segmentation techniques that may not work well in noisy environments. The grid technology (Feng et al., 2004), a substitute for segmentation, often assumes that each grid only contains one salient object. To satisfy this assumption, we have to cut the image into very small grids. However, according to our observation, both traditional segmentation algorithms and grid technology will generate 40-120 segments/grids for a 512×512 1-foot resolution aerial image, which makes the performance of annotation model deteriorate dramatically compared to 10-20 segments/grids per image. Therefore, we propose a two-stage segmentation algorithm to accommodate the uniqueness of HRRS images.
- Finally, another challenge faced by the HRRS image annotation is the importance of measuring spatial relationships among objects. In the HRRS images, individual objects cannot determine the semantics of the entire scene by itself. Rather, the repeated occurrence of certain object in the scene or the

co-occurrence of objects reflect high-level semantic concepts. For instance, if there is a remote sensing image about a city or urban area, instead of roof of an individual house, people may be more interested in identifying a park, which is the composition of grass land, pond, and curvy road. People would not be interested in large building roof alone. Nevertheless, if we identify that large building roofs have large parking lot and major road nearby, this would also be interesting, as we can annotate the image as shopping mall.

3. Preliminary and Related Work

In this section, we describe some basic concepts related to image retrieval and present the related work.

3.1. Three Query Levels

Accessing a desired image from a repository might involve a search for images that depict specific types of objects or scenes, evoke a particular mood, or simply contain a specific texture or pattern. Generally speaking, the user query could be as the following:

- The presence of a particular combination of color, texture or shape features (e.g. red cross)
- The presence or arrangement of a specific type of objects (e.g. a vase on a table)
- The depiction of a particular type of events (e.g. a baseball game)
- The presence of a named individual, location, or event (e.g. the Empire State Building)
- A subjective emotion expressed by an image (e.g. leisure)

According to (Eakins and Graham, 1999), a query can be categorized into three levels with increasing complexity.

The first level comprises of retrieval by primitive features such as color, texture and shape or the spatial location of image elements. Examples of such queries might include "find pictures with round yellow objects on top" and "find images with two objects next to each other". This level of retrieval uses features that are both objective and directly derivable from the images themselves, without the need to refer to any external knowledge bases. The first level has severe limitations, because it is hard to ask the common users to specify the low level features as a specialist.

The second level comprises of retrieval by derived features, involving some degree of semantic inference about the image. To answer queries at this level, some outside knowledge is required.

The third level of queries refers to image retrieval by abstract attributes. This involves a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted. We can divide these queries into two sub-directories. One is retrieval of named events or types of activity, such as "find the image of a Jazz Performance". Another is retrieval of images with emotional or religious significance, such as "find an image depicting anxiety". To

achieve this level of searching, complex reasoning and subjective judgment may be required.

Most of the pioneer content-based image retrieval systems, such as Query by Image Content (QBIC), VisualSEEK and Photobook, do a very good job in retrieving and analyzing the visual features. However, there is a significant gap between the first and the second level, not to mention satisfying the third level queries. To bridge the gap between the low-concept-level features and the semantic meanings of the images is the main purpose of this research.

3.2. High Resolution Remote Sensing Image Retrieval

Three major research fields are related to HRRS image retrieval:

- In photogrammetry community, most of the research has been focusing on individual object detection (Dial, L.Gibson and Poulsen, 2001; G.Sohn and Dowman, 2001; E.P.Baltsavias, 2004; Muller and Segl, 1999; Mayer, 1999). The research is so specific that the proposed building/road detection method cannot be applied for other objects detection.
- Remote-sensing image retrieval approach focuses on terrain type recognition (Datcu, Daschiel, Pelizzari, Quartulli, Galoppo, Colapicchioni, Pastori, Seidel, Marchetti and D’Elia, 2003; Castelli, Bergman, Kontoyiannis, Li, Robinson and Turek, 1998; Datcu, Pelizzari, Daschiel, Quartulli and Seidel, 2002). For example, to determine any given image by semantic labels, such as *water*, *cloud*, or *mountain*.
- There is other research that concentrates on problems of semantic-based image retrieval.
 - > The first is scene recognition. Scene recognition is important, both because the overall type scene depicted by an image is an important filter when used for search, and because it can help in identifying specific objects present. Similar to the systems that support retrieval by terrain types, these research efforts did not consider individual objects.
 - > The second focus of research activity is object recognition, an area of interest to the computer vision community for decades. The best known is human face and human body recognition (Eakins and Graham, 1999). All such techniques are based on the idea of developing a model of each class of object to be recognized, identifying image regions, which might contain examples of the object and building up evidence to confirm or rule out the object’s presence. Nevertheless, most of the models are not designed for regions with rich details that are typical of natural ground truth captured by HRRS images.

3.3. A Multi-level Image Retrieval Model

Most of the image retrieval approaches use multilevel abstraction mechanisms to support content-based image retrieval. As depicted in Figure 2, the top three levels, which correspond to the three user query levels, are the three key functionalities as well: feature extraction, object recognition, and domain-specific spatial reasoning and semantic modeling. This figure provides the essentials of content-based image retrieval.

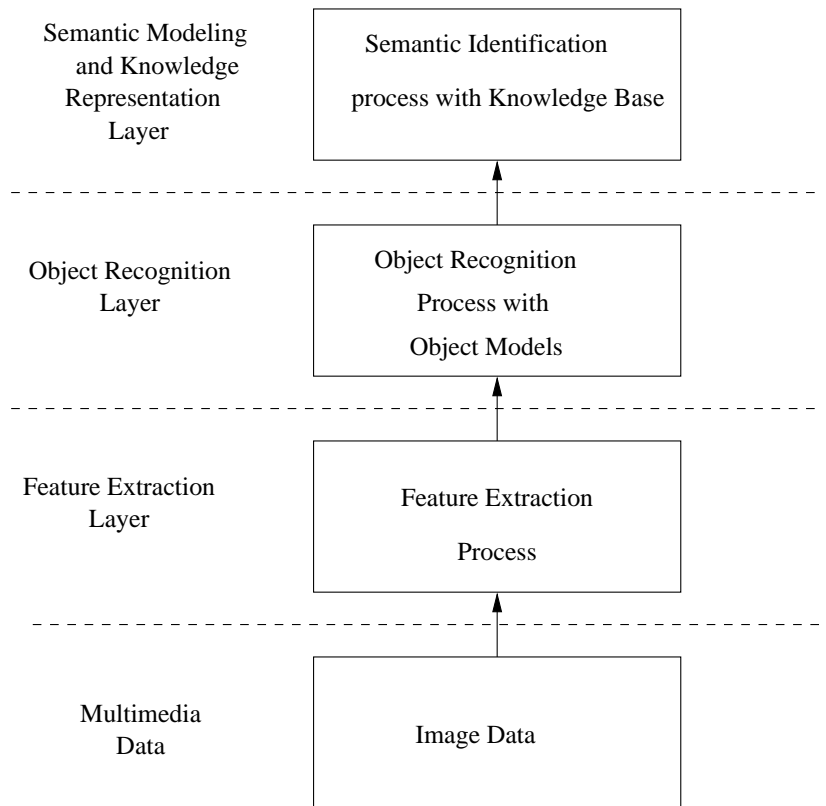


Fig. 2. Multilevel Semantic Data Modeling

Recently, researchers proposed several statistical models (Jeon et al., 2003; Barnard, Duygulu, de Freitas, Forsyth, Blei and Jordan, 2003; Wang et al., 2004; Duygulu et al., 2002; Feng et al., 2004; Mori et al., 1999) for analyzing the statistical relations between visual features and keywords. These methods can discover some hidden semantics of images. However, these methods annotate scenery images according to the presence of individual objects in each image. Spatial relations among objects are not taken into consideration.

The co-occurrence model (Mori et al., 1999) investigates the co-occurrence of words with image regions created using a regular grid. Each divided part inherits all words from its original image. (Mori et al., 1999) makes clusters from all divided images using vector quantization, accumulates the frequencies of words of all partial images in each cluster, and then calculates the likelihood for every word. It tends to require large number of training samples to estimate the correct probability and also tends to map frequent words to every blob. Also, Zhou et al. (Zhou and Zhang, 2006) proposed an elegant MIMLSVM algorithm for image classification. In this work, each image is also represented by a number of blobs which are clustered into representative patterns for images.

Under keyblock model (Zhu, Zhang, Rao and Shihari, 2000), images are partitioned into equal-sized blocks and then each image is represented as a code

matrix in which the elements are the indices of the keyblocks in a codebook. Based on this image representation, information retrieval and database analysis techniques developed in the text domain are transformed to image retrieval. The codebook, using keyblocks as vocabulary, is generated by applying generalized Lloyd algorithm (GLA) and pair wise nearest neighbor algorithm (PNNA) to the segmented grids in a training set. For each image in the database as well as the query, decompose it into block. Then, for each of the blocks, find the closest entry in the codebook and store the index correspondingly. Each image is then a matrix of indices, which can be treated as one-dimensional codes of the keyblocks in the codebook. After converting image features into properties that are similar to text document, they build the n -block model after the n -gram model used for statistical language modeling. This seems to be an application independent approach. However, this approach has an underlying assumption, which assume that objects in the image are usually salient enough so that they occupy more than one blocks. Otherwise, the segmented blocks will not be significant enough to be identified as keyblock. Moreover, all the above approaches do not take the context of image regions into consideration. Also, this approach did not use any textual annotation for images. For each keyblock, the semantic meaning of block is not considered either.

(Duygulu et al., 2002; Feng et al., 2004; Wang et al., 2004) all tried to map between segmented regions or grids and keywords supplied with the images using a method based around statistical algorithm. The variations are the details of the models. For example, Multiple Bernoulli relevance model (Feng et al., 2004) make sure that the annotating keywords are independent and not affected by other words. Also, in (Wang et al., 2004), researchers assign weight for different features based on their histogram distribution.

3.4. Representing Spatial Relationships

As presented in section 1, our goal is to discover composite objects. For example, parks, airports, and baseball fields are all composite objects. In the motivating examples, both shelter and disinvestment area are composite objects. As one can notice, the spatial relationships among objects play a critical role in identifying composite objects and interpreting the semantics of HRRS images.

Those spatial relationships are critical and cannot be ignored in HRRS images. Hence, in HRRS images, users pay more attention on composite objects than on individual objects. This suggests that we have to examine the spatial relationships among objects when we try to identify objects in HRRS images.

General approaches for modeling spatial semantics are based on identifying spatial relationships among objects once they are recognized and marked by the lower layer using bounding boxes or volumes. Several techniques have been proposed to formally represent spatial knowledge at this layer. These include: semantic networks, mathematical logic, constraints, inclusion hierarchies, and frames (A.Yoshitaka, Kishida, Hirakawa and Ichikawa, 1994). *Semantic networks* is a graph-based approach to represent spatial concepts and relationships. *Constraints-based methodology* uses constraints, which are the relationships between two or more objects need to satisfy in knowledge base. *Mathematical logic* projects the spatial relationships in the form of 2D strings to form a partial ordering of object projections in 2D. Another category of spatial representation is *inclusion hierarchies* which group together semantically related objects. *Frames*

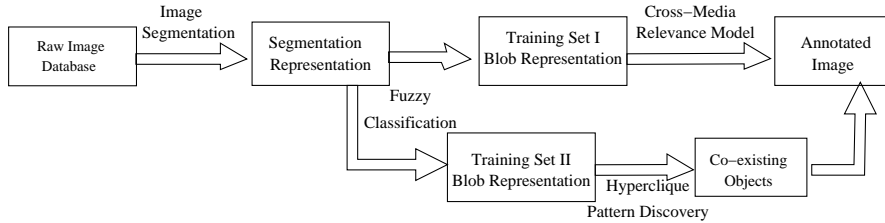


Fig. 3. An Overview of the OCCUE Method.

are used to represent knowledge relevant to particular objects, situations, or concepts. A frame usually consists of a name and a list of attribute-value pairs. However, none of the above approaches can be integrated with CBIR systems. Most of the CBIR systems need quantified features as input. How to integrate these representations with other quantitative input is a research challenge.

In HRRS image retrieval, the observed *domain knowledge* regarding spatial relationships is that the topological relations (Egenhofer, 1997), such as *disjoint*, *meet*, *overlap*, *contains*, *covers*, *inside*, etc., can be simplified as *exist* and *not-exist* in the definition of some composite objects. For example, with the repetitive appearance of single house roof, we can roughly conclude that the image capture the ground truth of the residential area. If we utilize this co-existence pattern, next time when we observe a large number of similar unlabeled rectangular shaped objects in an image, it will have a high probability that the image covers a residential area. Another example is that the co-existence of parking lot and large building roofs implies that it could be a shopping mall. It is not important to know whether the parking lot is left or right to the roof. This observation encouraged us to explore new approach to represent the spatial relationships among individual objects.

4. Object Discovery with Semantic Feature Selection

In this section, we introduce a method for Object disCcovery with semantiC featUre sElection (OCCUE). Figure 3 shows an overview of the OCCUE method. A detailed discussion of each step of OCCUE is given in the following subsections.

4.1. Two-Stage Object-based Semantic Image Segmentation

As the first step of the OCCUE method, image segmentation divides an image into separated regions. In a large-scale HRRS image database, the images naturally belong to different semantic clusters. For example, most of HRRS images can be categorized into four main semantic clusters at the land cover level including grass, water, residence and agriculture (Sheikholeslami, Chang, and Zhang, 2002). These land-cover level semantic clusters can also be divided into semantic subclusters at an object level. For these subclusters, the distinguishing primitive features are different. Moreover, the objects in each land-cover cluster are very different. For example, the objects in urban areas are usually road segments, single house roofs, or small vegetated areas. In contrast, woods and grass are dominant in suburban areas. Likewise, different composite objects also

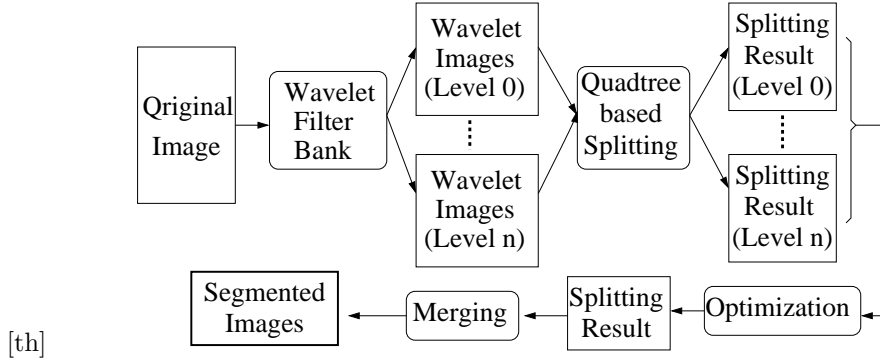


Fig. 4. Unsupervised Texture-based Segmentation Algorithm

appear in different land-cover clusters. For instance, a park is always a large contiguous vegetated area. This different scale distinguishes parks from gardens.

Given the special characteristics of HRSS imagery, in this paper, we propose a two-stage object-based approach to increase segmentation reliability as well as take the image semantics into consideration. In the first stage, we segment the HRSS images resolution according to their land cover. In the second stage, we segment labeled images under different land cover using different parameters. The outcomes of the second step are individual objects or components of an individual object. Our segmentation approach satisfies the uniqueness of RS images by considering and preserving the differences of images with different semantic context. Another major advantage of using proposed approach is that this segmentation approach can reflect the hierarchies that exist in the structure of the real-world objects which we are detecting. By abstracting houses, buildings, roads and other objects, people can identify residential areas and the aggregation of several residential areas yields a town.

Figure 4 illustrates the image process procedure which each image has to go through. To segment image at the land cover level, the segmentation method consists of three steps (Guo et al., 2005):

1. **Hierarchical Splitting.** Algorithm 1 shows the process of hierarchical splitting, which recursively splits the original image into children blocks by comparing texture features of blocks. As can be seen, this is a quad-tree based splitting process. For each image, we divide it into four square sub-blocks recursively until each individual sub-block is uniform in texture, which is determined by comparing the uniformity test metric $\frac{G_{max}}{G_{min}}$ with the pre-defined threshold X . To reduce the computational cost, we choose the enhanced Local Binary Pattern (LBP) approach to process the texture features.
2. **Optimizing.** This step adjusts the splitting result, if the results of the reduced resolution images have dramatically reduced number of segments.
3. **Merging.** Algorithm 2 shows the merging step, in which the adjacent blocks with similar texture are merged until a stopping criterion is satisfied. The input of algorithm 2 is the output of the optimizing step. The merging procedure works from bottom up. For a pair of adjacent segments, if they have the same parent and the ratio of G measure small than the given threshold X , we merge the two segments and update the LBP/C measure. Otherwise, if they have

Algorithm 1 Hierarchical Splitting

```

Require: input image, Level L
{
  Variables
  L: splitting level from original image
  H(x,y,L): Homogeneity indicator, by default, it is one if no need to split
  LBP: Local Binary Pattern as invariant to monotonic transform
  C: a contrast measure
  G: a metric used to measure the distance between two blocks LBP/C distributions
  X: a homogeneous threshold
}
divide input image into four blocks
L ← L + 1
for each subblock do
  compute LBP and C
  construct 2D Histogram
end for
for each subblock i do
  for each subblock j, i <> j do
    compute  $G_{i,j}$ 
  end for
end for
 $G_{max} \leftarrow \max(G_{i,j}), G_{min} \leftarrow \min(G_{i,j})$ 
X ← compute threshold
if  $\frac{G_{max}}{G_{min}} > X$  then
   $H(x, y, L) \leftarrow 0$ 
  its parent  $(X, Y)_{(x,y,L)} \leftarrow \text{null}$ 
else
   $H(x, y, L) \leftarrow 1$ 
  its parent  $(X, Y)_{(x,y,L)} \leftarrow (x, y)$ 
end if
if size(input image) > 16 x 16 then
  for each subblock do
    Quadtree-split(block,L)
  end for
  L ← L - 1
end if

```

different parents, we merge the two segments if their G measure smaller than the adjusted threshold $X\omega$, where ω is a weighted factor.

After the land-cover level segmentation, images are segmented into small regions using eCognition along with different input parameters according to land-cover type (<http://www.definiensimaging.com/>, 2004). For example, if the land-cover of an image is categorized as woods, shape features should have much less weight compare to images of residence land-cover. Each segment is represented by the traditional features, e.g. colors, textures and shapes, as well as the geometric features. eCognition utilizes a bottom-up-region-merging technique starting with one-pixel. In subsequent steps, smaller image segments are merged into bigger ones (<http://www.definiensimaging.com/>, 2004). We believe that this is one of the characteristics of the HRRS images: 1) with salt and pepper noises; 2) affected by the atmosphere and the reflective conditions.

The following extracted features represent major visual properties of each image segment.

– **Layer Values** are features concerning the pixel channel values of an image

Algorithm 2 Merging**Require:** totallevel, split result, X

```

{
Variables
L: splitting level from original image
H(x,y,L): Homogeneity indicator, by default, it is one if no need to split
ω: a weighted factor
G: metric used to measure the distance between two blocks LBP/C distributions
X: a homogeneous threshold
}
1:  $L \leftarrow \text{totallevel}$ 
2: for each level  $l$  do
3:   calculate weight factor  $\omega$ 
4:   for each block(x,y) at level  $l$  do
5:     if  $(X, Y)_{(x,y,l)}$  is null then
6:       for each neighbor(x',y') at level  $l$  do
7:          $G \leftarrow$  histograms distance
8:         if  $H(x, y, l) = 1 \cup H(x', y', l) = 1 \cup G < X\omega$  then
9:           Merge the two blocks
10:          change the histograms
11:         end if
12:       end for
13:     for each neighbor(x',y') at level  $l-1$  do
14:        $G \leftarrow$  histograms distance
15:       if  $H(x, y, l) = 1 \cup H(x', y', l-1) = 1 \cup G < X\omega$  then
16:         Merge the two blocks
17:         change the histograms
18:       end if
19:     end for
20:   end if
21: end for
22: end for

```

segment, mainly the spectral features, including mean, brightness, max difference, standard deviation, the ratio of mean value of an image segment over the mean of entire image, minimum pixel value, maximum pixel value, the mean difference to neighboring segment, the mean difference to brighter neighboring segment, mean difference to darker neighboring object.

- **Shape Features** include area (measured by pixel), length/width ratio which is the eigenvalues of the covariance matrix with the larger eigenvalue being the numerator of the factor, length, width, border length, density expressed by the area covered by the image segment divided by its radius, main direction, asymmetry, compactness (the product of the length m and the width n of the corresponding segment and divided by the number of its inner pixels), elliptic fit and rectangular fit.
- **Texture Features** evaluate the texture of an image segment based on the gray level co-occurrence matrix (GLCM) and the grey level difference vector (GLDV) of the segments pixel (<http://www.definiensimaging.com/>, 2004). The grey level co-occurrence matrix (GLCM) is a tabulation of how often different combinations of pixel grey level occur in an image. A different co-occurrence matrix exists for each spatial relationship. Therefore, we have to consider all four directions (0, 45, 90, 135) are summed before texture calculation. An angle of 0 represents the vertical direction, an angle of 90 the horizontal direction. Every GLCM is normalized, which guarantee the GLCM is symmetrical. The more distant to the diagonal, the greater the difference

between the pixels grey level is. The GLCM matrices can be further broken down to measure the homogeneity, contrast, dissimilarity (contrast increases linearly), entropy (distributed evenly), mean, standard deviation, and correlation. GLDV is the sum of diagonals of GLCM. It counts the occurrence of references to the neighbor pixels. Similarly to GLCM matrices, GLDV can measure the angular second moment (high if some elements are large), entropy (high if all similar), mean, and contrast.

- **Position Features** refer to the positions of segments within an image.

4.2. Fuzzy Classification

After we segment the images into relatively homogeneous regions, the next step is to group similar image segments into a reasonable number of classes, referred as blob tokens in (Wang et al., 2004). Segments in each class are similar even though they are not spatially connected. In the literature (Wang et al., 2004), unsupervised classification algorithms is employed using the primitive features or weighted features. Using the weighted features would successfully reduce the dimensionality compared with using all primitive features as clustering algorithm input. However, we used supervised classification method that is efficient in grouping image segments into semantic meaningful blobs.

Specifically, fuzzy logic based supervised classification is applied to generate blobs. Starting with an empty class hierarchy, we manually insert sample classes and using the features description as definition of a certain class. While nearest neighbor and membership functions are used to translate feature values of arbitrary range into a value between 0 (no membership) and 1 (full membership), logical operators summarize these return values under an overall class evaluation value between 0 and 1. The advantages of fuzzy classification are (<http://www.definiens imaging.com/>, 2004)

- Translating feature values into fuzzy values standardizes features and allows to combine features, even of very different ranges and dimensions.
- It enables the formulation of complex feature descriptions by means of logical operations and hierarchical class descriptions.

Finally, fuzzy classification also helps to merge the neighboring segments that belong to the same class and get a new semantic meaningful image blob which truly represents the feature and not just a part of it.

4.3. Hyperclique patterns

As we argued in the previous section, the next challenge is to find the co-existence objects. In this paper, hyperclique patterns (Xiong et al., 2003) are what we used for capturing co-existence of spatial objects. The concept of hyperclique patterns is based on frequent itemsets. Association rule mining can take two approaches in mining images (Teredesai, Ahmad, Kanodia and Gaborski, 2006; Shyu, Chen and Kashyap, 2001). The first one involves mining image visual features only while the second one involves mining both visual and textual data. In this research, we take the first approach.

In this subsection, we first briefly review the concepts on frequent itemsets, then describe the concept of hyperclique patterns.

Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of items. Each transaction T in database D is a subset of I . We call $X \subseteq I$ an itemset. The support of X $supp(X)$ is the fraction of transactions containing X . If $supp(X)$ is no less than a user-specified minimum support, X is called a frequent itemset. The confidence of association rule $X_1 \rightarrow X_2$ is defined as $conf(X_1 \rightarrow X_2) = supp(X_1 \cup X_2) / supp(X_1)$. It estimates the likelihood that the presence of a subset $X_1 \subseteq X$ implies the presence of the other items $X_2 = X - X_1$.

If the minimum support threshold is low, we may extract too many spurious patterns involving items with substantially different support levels, such as (caviar, milk). If the minimum support threshold is high, we may miss many interesting patterns occurring at low levels of support, such as (caviar, vodka). To measure the overall affinity among items within an itemset, the h-confidence was proposed in (Xiong et al., 2003). Formally, the h-confidence of an itemset $P = \{i_1, i_2, \dots, i_m\}$ is defined as $hconf(P) = \min_k \{conf(i_k \rightarrow P - i_k)\}$. Given a set of items I and a minimum h-confidence threshold h_c , an itemset $P \subseteq I$ is a hyperclique pattern if and only if $hconf(P) \geq h_c$. A hyperclique pattern P can be interpreted as that the presence of any item $i \in P$ in a transaction implies the presence of all other items $P - \{i\}$ in the same transaction with probability at least h_c . This suggests that h-confidence is useful for capturing patterns containing items which are strongly related with each other. A hyperclique pattern is a maximal hyperclique pattern if no superset of this pattern is a hyperclique pattern.

4.4. Converting Spatial Relationship into Feature Representation.

Approaches for modeling spatial relationships can be grouped into three categories: graph-based approaches, rule based approaches, and mathematical logic using 2D strings as the projections of the spatial relationships. However, none of this can be used as input for statistical Cross Relevance Model (CRM). In addition, we concentrate on the presence of the objects in the image rather than the complex geometric or topological spatial relationships. For example, consider a golf course, we are interested in the appearance of the well textured grassland, sand, non-rectangle water-body in a relatively small region. Whether the sand is left or right to the water-body is not important. In OCCUE, we apply hyperclique pattern discovery algorithm (Xiong et al., 2003) to detect co-existing objects.

Example 2.1 After segmentation, images are represented by the blob ID as shown in Table 1, let us consider a pattern $X = \{b_3, b_7, b_{24}\}$, which implies that blob (#3 roof type II, #7 shade type II, #24 grass type IV) usually appears together. We have $supp(b_3) = 82\%$, $supp(b_7) = 91\%$, $supp(b_{24}) = 73\%$, and $supp(b_3, b_7, b_{24}) = 55\%$. Then, $conf(b_3 \rightarrow b_7, b_{24}) = supp(b_3, b_7, b_{24}) / supp(b_3) = 67\%$; $conf(b_7 \rightarrow b_3, b_{24}) = supp(b_3, b_7, b_{24}) / supp(b_7) = 60\%$; $conf(b_{24} \rightarrow b_3, b_7) = supp(b_3, b_7, b_{24}) / supp(b_{24}) = 75\%$. Therefore, $hconf(X) = \min(conf(b_3 \rightarrow b_7, b_{24}), conf(b_7 \rightarrow b_3, b_{24}), conf(b_{24} \rightarrow b_3, b_7)) = 60\%$. According to the definition of **hyperclique pattern**, pattern $\{b_3, b_7, b_{24}\}$ is a hyperclique pattern at the threshold 0.6. Therefore, we treat the set of these three blobs as a new semantic feature. We treated these newly discovered hyperclique pattern as new blobs in addition to the existing blobs. Meanwhile, the original blobs #3, #7, and #24 are deleted from the original table. Table 1 will be converted to Table2. The new

Image	Blobs
in1	3,7,11,12,19,22,23,24,25
in2	3,7,6,12,13,15,18,20,23,24
in3	3,7,6,11,16,18,20,24,26
in5	7,6,10,11,12,20
in6	3,7,6,19,20,23,24,25
in7	3,7,12,19,20,23
in8	3,6,7,10,11,12,19,20,23
in9	3,6,15,11,12,20,24,26
in10	6,7,11,12,23,24
in11	3,6,7,11,12,19,22,23,24
in12	3,7,12,19,20,23,24

Table 1. A Sample Image-Blob Data Set

Image	Blobs
in1	11,12,19,22,23,25, 105
in2	6,12,13,15,18,20,23, 105
in3	11,16,18,20,26, 105
in5	7,6,10,11,12,20
in6	6,19,20,23,25, 105
in7	3,7,12,19,20,23
in8	3,6,7,10,11,12,19,20,23
in9	3,6,15,11,12,20,24,26
in10	6,7,11,12,23,24
in11	6,11,12,19,22,23 105
in12	12,19,20,23, 105

Table 2. A Sample Image Represented by New Blobs

blobs are represented using 3 digits number in order to distinguish from the original blobs. We convert the spatial relationship into a measurable representation, so that we can apply statistical model in the next step.

4.5. A Model of Image Annotation

Suppose we are given an un-annotated image in image collection $I \in \mathcal{C}$. We have the object representation of that image $I = \{o_1 \dots o_m\}$, and want to automatically select a set of words $\{w_1 \dots w_n\}$ that reflect the content of the image.

The general approach is widely accepted by statistical modeling approach. Assume that for each image I there exists some underlying probability distribution $P(\cdot|I)$ for all words. We refer to this distribution as the relevance model of I (Lavrenko and Croft, 2001; Lavrenko, Choquette and Croft, 2002). The relevance model can be thought of as an urn that contains all possible objects that could appear in image I as well as all words that could appear in the annotation of I . We assume that the observed image representation $\{o_1 \dots o_m\}$ is the result of m random samples from $P(\cdot|I)$.

$$P(w|I) \approx P(w|o_1 \dots o_m) \quad (1)$$

Equation (1) explains that the probability of any given word w appears in image I can be represented by the probability of word w appears given the blobs that image I contains.

In order to annotate an image with the top relevance words, we need to know the probability of observing any given word w when sampling from $P(\cdot|I)$.

Therefore, we need to estimate the probability $P(w|I)$ for every word w in the vocabulary. Given that $P(\cdot|I)$ itself is unknown, the probability of drawing the word w can be approximated by training set \mathcal{T} of annotated images, as shown in equation (2), in which J represents every image in training set \mathcal{T} .

$$P(w, o_1, \dots, o_m) = \sum_{J \in \mathcal{T}} P(J)P(w, o_1, \dots, o_m|J) \quad (2)$$

Assuming that observing w and blobs are mutually independent for any given image, and identically distributed according to the underlying distribution $P(\cdot|J)$. This assumption guarantees we can rewrite equation (2) as follows:

$$P(w, o_1, \dots, o_m) = \sum_{J \in \mathcal{T}} P(J)P(w|J) \prod_{i=1}^m P(o_i|J) \quad (3)$$

We assume the prior probability $P(J)$ follows uniform over all images in training set \mathcal{T} . We follow (Jeon et al., 2003) and use smoothed maximum likelihood estimates for the probabilities in equation (3). The estimations of the probabilities of blob and word given image J are obtained by:

$$P(w|J) = (1 - \alpha_J) \frac{Num(w, J)}{|J|} + \alpha_J \frac{Num(w, T)}{|T|} \quad (4)$$

$$P(o|J) = (1 - \beta_J) \frac{Num(o, J)}{|J|} + \beta_J \frac{Num(o, T)}{|T|} \quad (5)$$

Here, $Num(w, J)$ and $Num(o, J)$ represents the actual number of times the word w or blob o occurs in the annotation of image J . $Num(w, T)$ and $Num(o, T)$ is the total number of times w or o occurs in all annotation in the training set \mathcal{T} . $|J|$ denotes for the aggregate count of all words and blobs appearing in image J , and $|T|$ denotes the total size of the training set. The smoothing parameter α_J and β_J determine the interpolation degree between the maximum likelihood estimates and the background probabilities. Due to the different occurrence patterns between words (Zipfian distribution) and blobs (uniform distribution) in images, we separate the two smoothing parameter as α_J and β_J .

Finally, Equation (1) - (5) provide the mechanism for approximating the probability distribution $P(w|I)$ for an underlying image I . We annotate images by first estimating the probability distribution $P(w|I)$ and then select the highest ranking n words for the image. Note that the value of n is generally user-specified.

5. Experimental Evaluation

In this section, we present experiments to evaluate the performance of object discovery with semantic feature selection using real-world HRRS image data sets. Specifically, we show: (1) an example set of identified semantic spatial features, (2) the comparison results of our object-based semantic image segmentation with the segmentation function provided by in commercial tool ENVI, and (3) a performance comparison between our OCCUE model and a state-of-the-art Cross-media Relevance Model (CRM) model (Jeon et al., 2003).

Group	Sample Keywords
Land Cover	Suburban, Grassland, Sand, Pavement, water body
Individual Object	roof, road segment
Composite Object	golf course, industrial building, baseball field

Table 3. Examples of Keywords

5.1. The Experimental Setup

Experimental Data Sets. Since our focus in this paper is on HRRS images rather than regular scenery images, we will not adopt the popular image dataset Corel, which is considered as a benchmark for evaluating the performance of image retrieval algorithms. Instead, we use the high resolution orthoimagery of the major metropolitan areas. This data set is distributed by United States Geological Survey (USGS - <http://www.usgs.gov/>). The imagery is available as Universal Transverse Mercator (UTM) projection and referenced to North American Datum of 1983. For example, the New Jersey orthoimagery is available as New Jersey State Plane NAD83. The file format is Georeferenced Tagged Image File Format (GeoTIFF). Since we are not interested in the GIS information, we treat the GeoTIFF images as 3-bands color images. We choose metropolitan areas, such as New Jersey, and Washington D.C., since we believe that most composite objects users may be interested in are located in urban area. When we build our image data set, we intentionally include different landscapes.

Data Preprocessing. We downloaded the images of 1-foot resolution in the New York metro area, Boston and Springfield MA and Washington DC. Each raw image is about 80MB, which is then be processed using the Remote Sensing Exploitation Platform (ENVI - <http://www.itervis.com/envi/>). Images with blurred scene or with no major interesting objects, such as square miles of woods, are discarded. For images that contain objects we are interested in, we grid the image into small pieces (2048×2048 pixels). Finally, we have 800 images in our experimental data set and there are 32 features: 10 color features, 10 shape features and 12 texture features.

Keywords. The keywords used to annotate the semantics of the HRRS images are also different from the traditional scenery images. First of all, they are not attainable directly from the data set as those of Corel images. Rather, it is manually assigned by domain experts. These keywords can be divided into three groups: keywords regard land cover, individual objects, and composite objects. Table 3 lists some sample keywords.

Validation. In our experiments, we divided the data set into 10 subsets with equal number of images. We performed 10-cross validation. For each experiment, 8 randomly selected sub-dataset are used as training set, a validation set of 80 images and a test set of 80 images. The validation set is used to select the model parameters. Every images in the data set is segmented into comparatively uniform regions. The number of segments in each image, and the size of each segment (measured by the number of pixels) are empirically selected using the training and validating sets.

Blobs. Image Blobs are basic units to form semantic features. A fuzzy classification algorithm is first applied to generate image blobs. In our experiment, we generated 30 image blobs. Table 4 shows some examples of image blobs. Also, Figure 5 shows a small sample image and its blob representation. The more objects an image contains, the more complex it blob representation is. On average, a test image contains approximately 12-15 different blobs.

ID	Description	size	color	shape	texture
1	house I	(0,1200)	(150,180)	rectangle	smooth
2	house II	(1200,3000)	(150, 180)	rectangle	smooth
3	house III	(0,1200)	(180, 255)	rectangle	smooth
4	grass I	(0,2000)	(140, 160)	irregular	smooth
5	grass II	(0,2000)	(140, 180)	irregular	rough
30	sand	(0,5000)	(190,210)	round	rough

Table 4. Examples of Blobs



Blob ID
1, 2
3, 4
11, 12
28

Fig. 5. An Image and Its Blob Representation

Evaluation Metrics. To evaluate the annotation performance, we apply some external metrics including Precision, Recall, and F-measure. Specifically, we judge the relevance of the retrieved images by looking at the manual annotations of the images. A *Recall* measure is defined as the number of the correctly retrieved images divided by the number of relevant images in the test data set. The *Precision* measure is defined as the number of correctly retrieved images divided by the number of retrieved images. In order to make a balance between the recall and precision measures, we also compute the F-measure which is defined as $\frac{2*Recall*Precision}{Recall+Precision}$.

5.2. Semantic Features

All images with identified image blobs are used to identify the co-occurrence of image blobs. Specifically, we exploited a hyperclique pattern discovery method to find complex objects that consist of co-existing image blobs, which usually form a unique high-level semantic concept and are treated as spatial semantic features. For instance, Table 5 shows some example semantic features and composite-objects they represent. In addition, as the extension of our previous work, we also identify the semantic features as a concept hierarchy for better knowledge representation. As shown in figure 6, in both images we identify hyperclique pattern [3,4,24] which is assigned as *industrial building* according to Table 5. If we explore one more step, we find that it also contains composite object lawn [8,18]. We define a new composite object: office building, which provides better environment to employees, to distinguish it from industrial building whose main purpose is large storage and better transportation.

blobID	Comp-Object
17, 8	Golf Course
3, 20	Industrial Building
3, 4, 24	Industrial Building
1, 2, 5	Residential Building
1, 2, 9, 10	Residential Building
2, 12, 22	Baseball Field

Table 5. An Example of Semantic Features

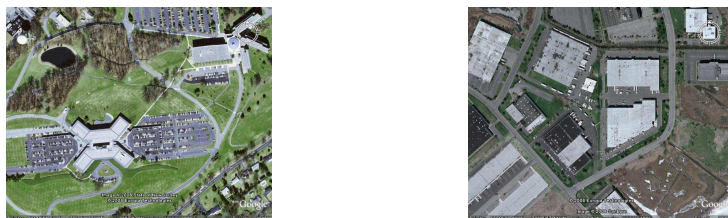


Fig. 6. Image blobs and concept hierarchy. Left: blob [3,4,24] (industrial building) and [8, 18] (lawn) represent new concept - office building; Right: blob [3,4,24](industrial building) and [3,20] (industrial building) represent concept - warehouse

5.3. Results of Parameter Selection

The hyperclique pattern discovery algorithm has two parameters: support and h-confidence. We examine the impact of these two parameters on the performance of object annotation. The minimum support and the h-confidence thresholds would affect object discovery. For example, the set of blobs (1, 2, 5, 9, 10) can be identified as co-existing objects with minimum support 0.05 and h-confidence 0.4, while it could not be identified when we change the minimum support to 0.15. Figure 7 shows the F-measure values with the change of minimum support and h-confidence thresholds. As can be seen, the F-measure values vary at different support and h-confidence thresholds. However, we can observe a general trend is

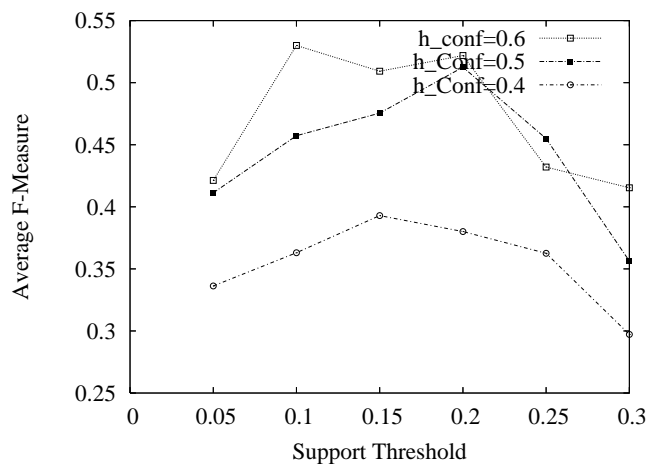


Fig. 7. F-measure Values at Different Parameters.

measures	word class	Avg. Prec.	Avg. Recall	F-Measure
CRM	land use	0.6801	0.5923	0.6332
OCCUE	land use	0.7512	0.7229	0.7368
CRM	object level	0.3013	0.1827	0.2274
OCCUE	object level	0.4682	0.3677	0.4119

Table 6. A Performance Comparison.

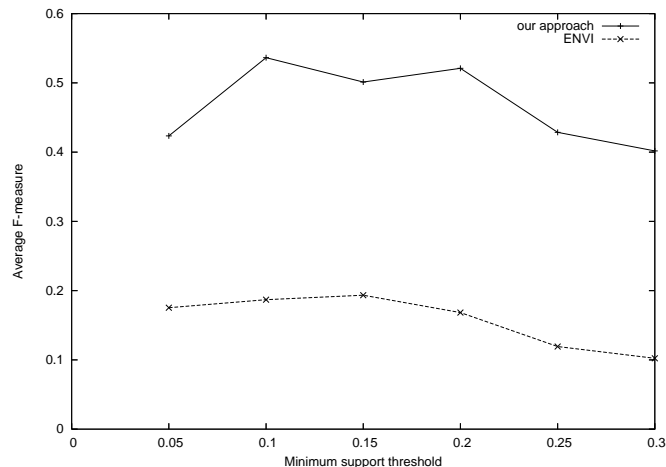


Fig. 8. A Performance Evaluation of Segmentation Methods.

that the F-measure values increase with the increase of H-confidence. Also, the maximum F-measure value is achieved when the support threshold is relatively high. This is reasonable, since a relatively high support threshold can guarantee statistical significance and provide a better coverage of objects. For this reason, in our experiments, we set relatively high support and h-confidence thresholds.

5.4. OCCUE versus CRM

We compared the annotation performance of the two models, the CRM model and the OCCUE model. We annotate each test image with 1 word from the land-cover level, 3 words from the composite object level.

Table 6 shows the comparison results. In the table, we can observe that, for both land-cover level and composite-object level, the performance of OCCUE is much better than that of CRM in terms of Precision, Recall, and F-measure. For instance, for the composite-object level, the F-measure value is improved from 0.2274 (CRM) to 0.4119 (OCCUE). This improvement is quite significant.

5.5. The Effect of the Semantic Segmentation

By exploiting different segmentation algorithms on HRRS images, we would have different image blobs, which can make a huge impact on the performance of the semantic feature selection. Therefore, in this subsection, we evaluate the performance of our segmentation results with those from the commercial system ENVI. We chose ENVI since it is widely used in processing remote sensing im-

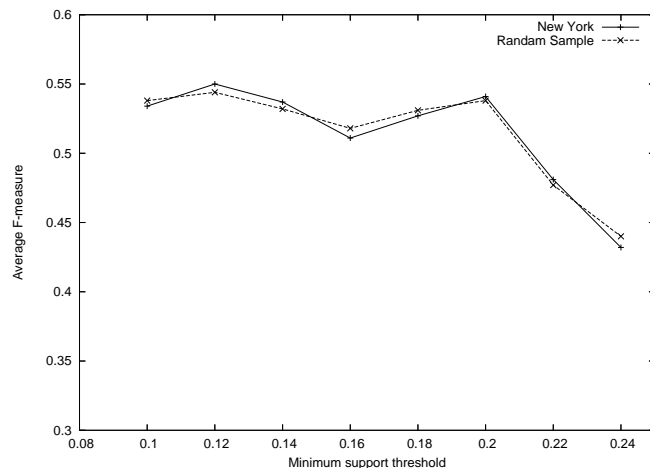


Fig. 9. The Robustness of identified Semantic Features at Different Atmosphere Condition.

ages. With ENVI, each raw image was segmented into smaller regions (blobs) which were fed into our OCCUE model. By using the tools available in ENVI for image segmentation, we can identify fewer semantic-features than we can get from our proposed segmentation approach. Figure 8 shows the final comparison results. As can be seen, our two-stage object-based semantic image segmentation algorithm can lead to much better performance of object discovery than the traditional segmentation method provided in ENVI.

5.6. The Robustness of our Semantic Feature Selection

Here, we evaluate whether our semantic feature method is robust in terms of dealing with different atmosphere conditions. Ideally, our approach should get similar annotation for the images which are taken under different air conditions. Aerial images of different areas usually were taken on different days, we manually selected a group of test images, which cover the New York area and were taken under the same air condition. Then, we compare the result on this selected group with other random selected test groups. The average F-measure values, given h-confidence as 0.6, are shown in Figure 9. As can be seen, the F-measure values from the New York images are consistent with those from the random groups. This result implies that the semantic features extracted by our method are robust enough to tolerant different air conditions.

6. Conclusions and Future Work

In this paper, we studied the problem of object discovery in High-Resolution Remote-Sensing (HRRS) imagery. Specifically, we first proposed a semantic-aware two-stage image segmentation approach, which preserves the semantics of real-world objects during the segmentation process. Then, we exploited a hyperclique pattern discovery technique to capture groups of co-existing individual objects, which usually form high-level semantic concepts. We treated these

groups of co-existing objects as new semantic features and fed them into the learning model. As demonstrated by our experimental results, with more reliable segmentation and new semantic feature sets, the learning performance can be significantly improved.

There are several potential directions for future research on this topic. For example, we plan to adapt Spatial Auto-Regression (SAR) model (Shekhar and Chawla, 2003) for object discovery in HRRS images. The SAR model has the ability in measuring spatial dependency, and thus is expected to have a better prediction accuracy for spatial data.

7. Acknowledgments

We thank the editor and three anonymous reviewers for their very useful comments and suggestions. This work was partially supported by NSF grant # IIS-0306838, the Center for Information Management, Integration and Connectivity (CIMIC) at Rutgers University. Also, this research was supported in part by a Faculty Research Grant from Rutgers Business School- Newark and New Brunswick. The content of this work does not necessarily reflect the position or policy of the government and no official endorsement should be inferred.

References

- A.Yoshitaka, Kishida, S., Hirakawa, M. and Ichikawa, T. (1994), ‘Knowledge-assisted content-based retrieval for multimedia database’, *IEEE Multimedia* **1**(4), 12–21.
- Barnard, K., Duygulu, P., de Freitas, N., Forsyth, D., Blei, D. and Jordan, M. I. (2003), ‘Matching words and pictures’, *Machine learning research* **3**(1), 1107–1135.
- Castelli, V., Bergman, L. D., Kontoyiannis, I., Li, C.-S., Robinson, J. T. and Turek, J. J. (1998), ‘Progressive search and retrieval in large image archives’, *IBM Journal of Research and Development* **42**(2).
- Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P. G. and D’Elia, S. (2003), ‘Information mining in remote sensing image archives: System concepts’, *IEEE Transactions on Geoscience and Remote Sensing* **41**(12), 2923–2936.
- Datcu, M., Pelizzari, A., Daschiel, H., Quartulli, M. and Seidel, K. (2002), Advanced value adding to metric resolution sar data: information mining, in ‘Proceedings of 4th European Conference of Synthetic Aperture Radar’, EUSAE.
- Dial, G., L.Gibson and Poulsen, R. (2001), Ikonos satellite imagery and its use in automated road extraction, in ‘Automatic extraction of man-made objects from aerial and space images (III)’, Vol. 1, ISPRS, pp. 357–367.
- Duygulu, P., Barnard, K., de Freitas, N. and Dorsyth, D. (2002), Object recognition as machine translation: learning a lexicon for a fixed image vocabulary, in ‘Seventh European Conference on Computer vision’, Vol. 4, pp. 97–112.
- Eakins, J. and Graham, M. (1999), Content-based image retrieval, Technical report, JISC Technology Applications Programm.
URL: <http://www.unn.ac.uk/iidr/report.html>
- Egenhofer, M. J. (1997), ‘Query processing in spatial-query-by-sketch’, *Journal of Visual Languages and Computing* **8**(4), 403–424.
- E.P.Baltsavias (2004), ‘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**(1), 129–151.
- Feng, S., Manmatha, R. and Lavrenko, V. (2004), Multiple bernoulli relevance models for image and video annotation, in ‘Computer Vision and Pattern Recognition, CVPR04’, pp. 1002–1009.
- Fung, G. and Stoeckel, J. (2007), ‘Svm feature selection for classification of spect images of alzheimer’s disease using spatial information’, *Knowl. Inf. Syst.* **11**(2), 243–258.

- G.Sohn and Dowman, I. (2001), Extraction of buildings from high-resolution satellite data, in 'Automatic extraction of man-made objects from aerial and space images (III)', Vol. 1, ISPRS, pp. 345–355.
- Guo, D., Atluri, V. and Adam, N. (2005), Texture-based remote-sensing image segmentation, in 'Proceedings Of the 2005 International Conference on Multimedia and Expo', pp. 1472–1475.
- Gupta, A., Weymouth, T. and R.Jain (1991), 'Sematnic queries in image databases', *Visual Database Systems* **2**(1), 204–218.
- <http://www.definiens imaging.com/> (2004), 'ecognition userguide'.
- Jeon, J., Lavrenko, V. and Manmatha, R. (2003), Automatic image annotation and retrieval using cross-media relevance models, in 'SIGIR', pp. 254–261.
- Lavrenko, V., Choquette, M. and Croft, W. (2002), Cross-lingual relevance models, in 'Proceedings of the 25th annual international ACM SIGIR conference', pp. 175–182.
- Lavrenko, V. and Croft, W. (2001), Relevance-based language models, in 'Proceedings of the 24th annual international ACM SIGIR conference', pp. 120–127.
- Mayer, H. (1999), 'Automatic object extraction from aerial imagery - a survey focusing on buildings', *Computer vision and image understanding* **74**(2), 138–149.
- Mori, Y., Takahashi, H. and Oka, R. (1999), Image-to-word transformation based on dividing and vector quantizing images with words., in 'in MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management'.
- Muller, M. and Segl, K. (1999), Object recognition based on high spatial resolution panchromatic satellite imagery, in 'Joint workshop of ISPRS on Sensors and Mapping from Space 1999', ISPRS.
- Sheikhholeslami, G., Chang, W., and Zhang, A. (2002), 'Semquery: Semantic clustering and querying on heterogeneous features for visual data', *IEEE Transactions on Knowledge and Data Engineering* **14**(5), 988–1002.
- Shekhar, S. and Chawla, S. (2003), *Spatial Databases: A Tour*, Prentice Hall.
- Shyu, M.-L., Chen, S.-C. and Kashyap, R. L. (2001), 'Generalized affinity-based association rule mining for multimedia database queries', *Knowledge and Information Systems* **3**(3), 319–337.
- Teredesai, A. M., Ahmad, M. A., Kanodia, J. and Gaborski, R. S. (2006), 'Comma: a framework for integrated multimedia mining using multi-relational associations', *Knowledge and Information Systems* **10**(2), 135–162.
- Wang, L., Khan, L., Liu, L. and Wu, W. (2004), Automatic image annotation and retrieval using weighted feature selection, in 'Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering', Kulwer Publisher.
- Xiong, H., Tan, P. and Kumar, V. (2003), Mining strong affinity association patterns in data sets with skewed support distribution, in 'Proceedings of the 3rd IEEE International Conf. on Data Mining', pp. 387–394.
- Zhou, Z.-H. and Zhang, M.-L. (2006), Multi-instance multi-label learning with application to scene classification, in 'Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems (NIPS), Vancouver, British Columbia, Canada, December 4-7, 2006', pp. 1609–1616.
- Zhu, L., Zhang, A., Rao, A. and Shihari, R. (2000), Keyblock: An approach for content-based image retrieval, in 'Proceedings of ACM Multimedia 2000', pp. 147–156.

Author Biographies



Dihua Guo is currently a member of Research Applied Data Mining, AT&T Labs. She received the B.E. degree in Management Information Systems from Tianjin University, China, and the Ph.D. degree in Information Technology from Rutgers, the State University of New Jersey, USA. Her general area of research is data mining and knowledge discovery, with a focus on innovation in large-scale information mining and effective data analysis techniques for geo-business data.



Hui Xiong is currently an Assistant Professor in the Management Science and Information Systems department at Rutgers, the State University of New Jersey. He received the B.E. degree in Automation from the University of Science and Technology of China, China, the M.S. degree in Computer Science from the National University of Singapore, Singapore, and the Ph.D. degree in Computer Science from the University of Minnesota, USA. His general area of research is data and knowledge engineering, with a focus on developing effective and efficient data analysis techniques for emerging data intensive applications. He has published over 50 technical papers in peer-reviewed journals and conference proceedings. He is a co-editor of *Clustering and Information Retrieval* (Kluwer Academic Publishers, 2003) and a co-Editor-in-Chief of *Encyclopedia of GIS* (Springer, 2008). He is an associate editor of the *Knowledge and Information Systems* journal and has served regularly in the organization committees and the program committees of a number of international conferences and workshops.

He was the recipient of the 2007 Junior Faculty Teaching Excellence Award and the 2008 Junior Faculty Research Award at the Rutgers Business School. He is a senior member of the IEEE, and a member of the ACM, the ACM SIGKDD, and Sigma Xi.



Vijayalakshmi Atluri received her B.Tech. in Electronics and Communications Engineering from Jawaharlal Nehru Technological University, Kakinada, India, M.Tech. in Electronics and Communications Engineering from Indian Institute of Technology, Kharagpur, India, and Ph. D. in Information Technology from George Mason University, USA. She is currently a Professor of Computer Information Systems in the MSIS Department, and research director for the Center for Information Management, Integration and Connectivity (CIMIC) at Rutgers University. Her research interests include Information Systems Security, Databases, Workflow Management, Spatial Databases, Multimedia and Distributed Systems. She is a member of the Steering Committee for ACM Special Interest Group on Security Audit and Control (SIGSAC). She served as a steering committee member for the ACM Symposium on Access Control Models and Architectures (SACMAT), general chair for the 2004 and 2005 ACM Conference on

Computer and Communications Security (CCS), and co-general chair for the 2005 International Conference on Web Information Systems Engineering. She served as the program chair for the 2003 CCS, 2000 ACM Workshop on Role Based Access Control, and the 2008 IFIP WG11.3 Working Conference on Database Security. She is the chair of the IFIP WG11.3. She has recently been elected as the secretary/treasurer for the ACM SIGSAC. In 1996, she was a recipient of the National Science Foundation CAREER Award. In 1999, she received the Rutgers University Research Award for untenured faculty for outstanding research contributions. Dr. Atluri is a senior member of the IEEE, and a member of the ACM.



Nabil R. Adam is serving as a Fellow at the Science & Technology Directorate, US Department of Homeland Security. He is a Professor of Computers and Information Systems; the Founding Director of the Rutgers University Center for Information Management, Integration and Connectivity (CIMIC); past chair of the MSIS Department at Rutgers Business School; past Director of the Meadowlands Environmental Research Institute; and the Director of the recently established information Technology for Emergency mAnageMent (i-TEAM) Research Laboratory at Rutgers. Dr. Adam has published over 100 technical papers covering such topics as information management, information security and privacy, data mining and Web services. His papers appeared in referred journals and conference proceedings including, IEEE Transactions on Software Engineering, IEEE Transactions on Knowledge and Data Engineering, ACM Computing Surveys, Communications of the ACM, Journal of Management Information Systems, and International Journal of Intelligent and Cooperative Information Systems. He has co-authored/co-edited ten books. Dr. Adam is the co-founder and the Executive-Editor-in-Chief of the International Journal on Digital Libraries and serves on the editorial board of a number of journals including Journal of Management Information Systems, and the Journal of Electronic Commerce. He is also the co-founder and past chair of the IEEE Technical Committee on Digital Libraries. Dr. Adam co-holder of a European issued Patent and has two pending patent applications submitted to the U.S. Patent and Trademark Office; all related to Web services. His research work has been supported by over \$15 million from various federal and state agencies, including the National Science Foundation (NSF), the National Security Agency (NSA), National Oceanic and Atmospheric Administration (NOAA), the U.S. Environmental Protection Agency, the Defense Logistics Agency (DLA), the National Library of Medicine, the New Jersey Meadowlands Commission, and NASA.

Correspondence and offprint requests to: Hui Xiong, Management Science and Information Systems Department, Rutgers University, Newark, NJ 07102, USA. Email: hxiong@rutgers.edu