Semantic Feature Selection for Object Discovery in High-Resolution Remote Sensing Imagery

Dihua Guo, Hui Xiong, Vijay Atluri, and Nabil Adam

MSIS Department, Rutgers University, USA
{devaguo, hui, atluri, adam}@cimic.rutgers.edu

Abstract. Given its importance, the problem of object discovery in High-Resolution Remote-Sensing (HRRS) imagery has been given a lot of attention by image retrieval researchers. Despite the vast amount of expert endeavor spent on this problem, more effort has been expected to discover and utilize hidden semantics of images for image retrieval. To this end, in this paper, 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 new semantic features as starting points, we can improve the performance of object discovery in terms of various external criteria.

1 Introduction

With the advances of remote sensing technology and the increases 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. However, it remains a challenging task to identify objects in HRRS images. While HRRS images share some common features with traditional images, they possess some special characteristics which make the object discovery more complex and motivate our research work.

Motivating Examples. Users are interested in different types of objects on Earth as well as groups of objects with various spatial relationships. For example, 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 shelter should not be far away from water source. Therefore, the 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*

Z.-H. Zhou, H. Li, and Q. Yang (Eds.): PAKDD 2007, LNAI 4426, pp. 71–83, 2007.

Hudson county industrial area". 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 need 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 concept levels: (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, forests, etc. The second type of objects are 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. Composite objects are objects that consist of several individual objects that form a new semantics concept. 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.

Despite the vast amount of expert effort, it is well known that the performance of CBIR is limited by the gap between low-level features and high-level semantic concepts. Recently, researchers proposed several statistical models [6,1,12,2,3,9] 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 individual objects' presence in each image. Spatial relations among objects are not taken into consideration. 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.

In this paper, we investigate the problem of automatically annotating images using relevance-based statistical model on HRRS images. Specifically, we exploit a hyperclique pattern discovery method [13] to create new semantic features and feed them into the relevance-based statistical learning model. 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 can learn the joint distribution of the blobs and words. Here, the blobs are image segments acquired directly from 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 new semantic features as starting points, the performance of learning model can be improved according to several external criteria.

2 Domain Challenges

In this section, we describe some domain challenges for object discovery in HRRS images as follows.

- First, it is nontrivial to perform feature selection for image retrieval in HRRS images. In [12], 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 [4]. This data problem has been largely ignored by existing approaches, thus not suitable for object discovery in HRRS images. Indeed, existing methods often use segmentation techniques which may not work well in noisy environments. Moreover, the grid technology [3], a substitute of segmentation, often assume that each grid only contains one salient object. To satisfy the 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 an remote sensing image about a city or urban area, instead of roof of individual house, people maybe 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 Object Discovery with Semantic Feature Selection

In this section, we introduce a method for **O**bject dis**C**overy with semanti**C** feat**U**re s**E**lection (OCCUE). Figure 1 shows an overview of the OCCUE method. A detailed discussion of each step of OCCUE is given in the following subsections.

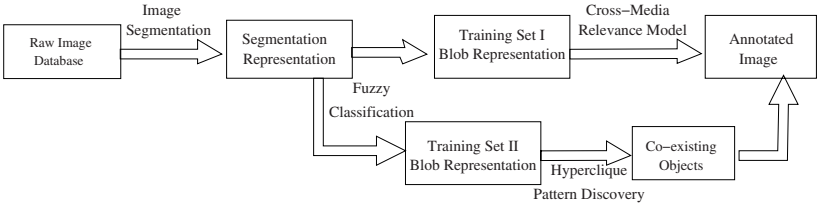

Fig. 1. A Overview of the OCCUE Method

3.1 Image Segmentation

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 [10]. 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 appear in different land-cover clusters. For instance, a park is always a large contiguous vegetated area. This different scale distinguishes parks from gardens. In OCCUE, we exploit a two-step approach to increase segmentation reliability. Our two-step segmentation approach satisfies the uniqueness of RS images by segmenting images at the land-cover level first and then dividing images further into individual objects or components of an individual object.

Another major advantage of using two-step image segmentation 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. This hierarchy is obviously determined by scale.

In OCCUE, we apply the texture-based algorithms proposed by [4] to segment image at the land cover level. This segmenting method consists of three major steps: (i) hierarchical splitting that recursively splits the original image into children blocks by comparing texture features of blocks, (ii) optimizing, which adjusts the splitting result, if the results of the reduced resolution images have dramatically reduced segments, (iii) merging, in which the adjacent regions with similar texture are merged until a stopping criterion is satisfied.

After the land-cover level segmentation, images are segmented into small regions using eCognition along with different input parameters according to land-cover type [5]. 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 [5]. We believe that this is one of the easy-to-use and reliable segmentation tools for HRRS images, given 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 segment, mainly the spectral features, including mean, brightness, max difference, standard deviation, the ratio of layer mean value of an image segment over the all 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 [5]. 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.

3.2 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 [12]. Segments in each class are similar even though they are not spatially connected. In the literature [12], 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 [5]

- 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.

3.3 Hyperclique Patterns

In this paper, hyperclique patterns [13,14] are what we used for capturing co-existence of spatial objects. The concept of hyperclique patterns is based on frequent itemsets. In this subsection, we first briefly review the concepts on frequent itemsets, then describe the concept of hyperclique patterns.

Let $I = \{i_1, i_2, ..., i_m\}$ 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 [13]. Formally, the h-confidence of an itemset $P = \{i_1, i_2, ...i_m\}$ 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.

3.4 Converting Spatial Relationship into Feature Representation

Approaches for modelling 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 [13] to detect co-existing objects.

Table 1. A sample image-blob data set

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

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 additional 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 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.*

Table 2. A sample image represented in new blob

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**

3.5 A Model of Image Annotation

Suppose we are given an un-annotated image in image collection $\mathcal{I} \in \mathcal{C}$. We have the object representation of that image $\mathcal{I} = \{o_1 \dots o_m\}$, 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 modelling approach. Assume that for each image $\mathcal{I}$ there exists some underlying probability distribution P(·|I). We refer to this distribution as the relevance model of I [8,7]. The relevance model can be thought of as an urn that contains all possible objects that could appear in image $\mathcal{I}$ as well as all words that could appear in the annotation of $\mathcal{I}$. We assume that the observed image representation $\{o_1 \dots o_m\}$ is the result of m random samples from P(·|I).

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

$$P(w|I) \approx P(w|o_1 \dots o_m) \tag{1}$$

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

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

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

We assume the prior probability $P(J)$ follows uniform over all images in training set $mathcalT$. We follow [6] 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:

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

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

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 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.

4 Experimental Evaluation

In this section, we present experiments on real-world data sets to evaluate the performance of object discovery with semantic feature selection. Specifically, we show: (1) an example set of identified semantic spatial features, (2) a performance comparison between the OCCUE model and a state-of-the-art Cross-media Relevance Model (CRM) model [6].

4.1 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).

Data Preprocessing. We downloaded the images of 1-foot resolution in the New York metro area and Springfield MA. Each raw image is about 80MB, which is then be processed using the Remote Sensing Exploitation Platform (ENVI - http://www.ittvis.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 $\times$ 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 seta as those of Corel images. Rather, it is manually assigned by domain experts. These keywords can be divided into three groups: keywords regard landcover, individual objects, and composite objects.

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. A fuzzy classification algorithm is applied to generate image blobs. In our experiment, we generated 30 image blobs. Table 3 shows some examples of image blobs. Also, Figure 2 shows a sample image and its blob representation.

Table 3. Examples of 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

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

Fig. 2. An Image and Its Blob Representation

Spatial 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 4 shows some example semantic features.

4.2 Results of Composite Object Annotation

To evaluate the annotation performance, we apply some external metrics including Precision, Recall, and F-measure. Specifically, we judge the relevance of the

Table 4. An Example of Semantic Features

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

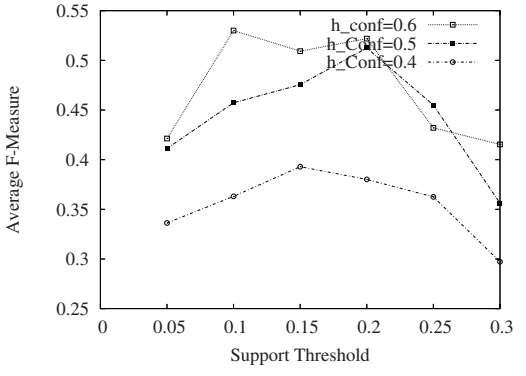

Fig. 3. F-measure Values at Different Parameters

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 $\frac{2*Recall*Precision}{Recall+Precsion}$.

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 3 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 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.

Table 5. A Performance Comparison

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

A Model Comparison. 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 5 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 Conclusions and Future Work

In this paper, we proposed a semantic feature selection method for improving the performance of object discovery in High-Resolution Remote-Sensing (HRRS) images. Specifically, 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 feed them into the learning model. As demonstrated by our experimental results, with new semantic feature sets, the learning performance can be significantly improved.

There are several potential directions for future research. First, we propose to adapt Spatial Auto-Regression (SAR) model [11] 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. Second, we plan to organize the identified semantic features as a concept hierarchy for the better understanding of new discovered high-level objects.

References

1. K. Barnard, P. Duygulu N. de Freitas, D. Forsyth, D. Blei, and M. I. Jordan. Matching words and pictures. *Machine learning research*, 3(1):1107–1135, 2003.
2. P. Duygulu, K. Barnard, N. de Freitas, and D. Dorsyth. Object recognition as machine translation: learning a lexcicon for a fixed image vocabulary. In *ECCV*, volume 4, pages 97–112, 2002.
3. SL Feng, R Manmatha, and V Lavrenko. Multiple bernoulli relevance models for image and video annotation. In *CVPR*, pages 1002–1009, 2004.
4. D. Guo, V. Atluri, and N. Adam. Texture-based remote-sensing image segmentation. In *ICME*, pages 1472–1475, 2005.
5. http://www.definiens imaging.com/. ecognition userguide, 2004.

6. J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *SIGIR*, pages 254–261, 2003.
7. V. Lavrenko, M. Choquette, and W. Croft. Cross-lingual relevance models. In *SIGIR*, pages 175–182, 2002.
8. V. Lavrenko and W. Croft. Relevance-based language models. In *SIGIR*, pages 120–127, 2001.
9. Y. Mori, H. Takahashi, and R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In *MISRM*, 1999.
10. G. Sheikholeslami, W. Chang, and A. Zhang. Semquery: Semantic clustering and querying on heterogeneous features for visual data. *TKDE*, 14(5):988–1002, 2002.
11. S. Shekhar and S. Chawla. *Spatial Databases: A Tour*. Prentice Hall, 2003.
12. L. Wang, L. Khan, L. Liu, , and W. Wu. Automatic image annotation and retrieval using weighted feature selection. In *IEEE-MSE*. Kulwer Publisher, 2004.
13. H. Xiong, P. Tan, and V. Kumar. Mining strong affinity association patterns in data sets with skewed support distribution. In *ICDM*, pages 387–394, 2003.
14. H. Xiong, P. Tan, and V. Kumar. Hyperclique pattern discovery. *Data Mining and Knowledge Discovery Journal*, 13(2):219–242, 2006.