Classification of high resolution remote sensing image based on Geo-ontology and Conditional Random Fields

Liang hong *a,b

aSchool of Geosciences and Info-Physics, Central South University, Changsha, China; bCollege of Tourism and Geography Science, Yunnan Normal University, Kunming, China

ABSTRACT

The availability of high spatial resolution remote sensing data provides new opportunities for urban land-cover classification. More geometric details can be observed in the high resolution remote sensing image, also ground objects in the high resolution remote sensing image have displayed rich texture, structure, shape and hierarchical semantic characters. More landscape elements are represented by a small group of pixels. Recently, the an object-based remote sensing analysis methodology is widely accepted and applied in high resolution remote sensing image processing. The classification method based on Geo-ontology and conditional random fields is presented in this paper. The proposed method is made up of four blocks: (1) the hierarchical ground objects semantic framework is constructed based on geo-ontology; (2) segmentation by mean-shift algorithm, which image objects are generated. And the mean-shift method is to get boundary preserved and spectrally homogeneous over-segmentation regions; (3) the relations between the hierarchical ground objects semantic and over-segmentation regions are defined based on conditional random fields framework; (4) the hierarchical classification results are obtained based on geo-ontology and conditional random fields. Finally, high-resolution remote sensed image -GeoEye, is used to testify the performance of the presented method. And the experimental results have shown the superiority of this method to the eCognition method both on the effectively and accuracy, which implies it is suitable for the classification of high resolution remote sensing image.

Keywords: Classification, high resolution remote sensing image, Geo-ontology, Conditional Random Fields (CRF)

1. INTRODUCTION

An increasing number of optical high resolution (HR) remote sensing satellite systems have become available in the last decade. High spatial resolution remote sensing data provides new opportunities for urban land-cover classification[1]. High-resolution imagery can show detailed object information, such as structure, texture and detail more clearly. However, with the availability of object information, the user community faces new challenges in land-cover classification of these types of data, i.e. increased spectral within-field variability as the spatial resolution increases. Therefore, it translates into the reduction of the classification accuracy by traditional pixel-wise based on spectral information alone. Thus, many classification methods based on pixel-wise have been presented using both spectral and spatial information of a single pixel. However, the traditional classification method based on pixel-wise has been proved to have several drawbacks, such as low classification accuracy, very limited spatial information to be derived and salt-and-pepper effects[2, 3]. To alleviate those problems, an object-based remote sensing analysis methodology is widely accepted and applied. A detail review about object-based remote sensing image analysis can be found in Blaschke’s work[4]. Generally, two steps are needed in object-oriented classification[1]: (1) segmentation, and (2) classification. Segmentation involves partitioning the image into contiguous groups of pixels called objects. Ideally, these objects correspond to real world objects of interest. The second step commences with the classification of these objects based on spectral, textural, size, shape, and contextual features. The use of successfully segmented images may lead to improved classification accuracy when compared to pixel-based classification methods[5]. High-resolution remote sensing image is a high random signal. Therefore, the statistical method is used to study the high-resolution remote sensing image segmentation is an effective way. In statistical analysis method, the method based on Markov random field can effectively describe the spatial information of image and have the best theoretical foundation which received extensive attention. However, for computational tractability, the observed spectral vectors in MRF framework are assumed to be conditional independent. This assumption neglects the contextual information in the observed spectral data of a given class. To capture the contextual information in MRF frameworks, several attempts have been made in the past. The most usual strategy incorporates the statistics of spectral differences between neighboring sites into the Gibbs energies of
MRF (Markov Random Field) frameworks [6, 7]. Another available method develops MRF frameworks with tree structures, which utilize the observed data at a given site and its parents. Although these MRF frameworks have the abilities to capture the contextual information in both the labels and observed data, they all suffer from problems derived mainly from the need of explicit likelihood model. Conditional random field (CRF) [8, 9] has recently gained popularity, which is a sequence modeling framework that has all the advantages of MRF but also solves the label bias problem in a principled way. The critical difference between CRFs and MRF is that a MRF uses per-state exponential model for the conditional probabilities of next states given the current state, while a CRF has a single exponential model for the joint probability of the entire sequence of labels given the observation sequence.

High spatial resolution remote sensing image features can be described from spectral, texture, size, shape, context, and structure, etc. However, High spatial resolution remote sensing image can also be illustrated from the so-called semantic content. The semantic content can show Hierarchically Multi-scale semantic information characteristic. Remote sensing image interpretation is a difficult task and can be defined as the extraction of the remote sensing image semantic. It consists in extracting the high-level semantic content of images from low-level feature, the representation of the semantic content. This problem is known as the semantic gap [10, 11] and is defined as the lack of concordance between low-level information and high-level information. In order to reduce the semantic gap, image analysis methods using object-based approaches with domain knowledge are developed [12].

In this paper we present a framework for high resolution remote sensing image classification based on Geo-ontology and CRF. Firstly, the segmentation algorithm of over-segmented object is simply introduced. The ground objects are obtained by the segmenting. And the each object features are obtained; Secondly, the extraction of the semantic concepts from images and the construction of a Geo-ontology. A key issue is to identify appropriate concepts in terms of hierarchy structure and, in terms of internal definition to describe the thematic objects. Thirdly, the relations between the hierarchical ground objects semantic and over-segmentation regions are defined based on conditional random field framework and the hierarchical classification results are obtained based on geo-ontology and conditional random fields. Fig. 1 illustrates the different steps of the approach. We also present experimental results to highlight the relevance of our method on HR images.

![Figure 1. The Method Workflow](image)

2. IMAGE SEGMENTATION

A key step in object-based image analysis is image segmentation. Image segmentation creates exclusive small parcels and those small parcels (also called superpixels or objects) are over-segmented. Based on superpixels, the number of units to be classified later can be largely reduced. Moreover, a more vivid simulation of spatial structure of land-cover patterns is provided in the parcels based methods. The commonly used approaches for creating superpixels include mean-shift clustering, watershed, and multiresolution segmentation based on Fractal Net Evolution [1, 5].

We chose the mean-shift (MS) method for initial segmentation [13, 14] because its ability to preserve the object boundaries has been proven in both nature images and remote sensing images. Meanwhile, the parameters in the method can be tuned easily according to the user’s purpose, as the tuning parameters in the method have an explicit implication (i.e., the minimum region in these generation result, the spatial range considered for current pixel sample’s density estimation). The method can be regarded as a process of discontinuity preserving smoothing and fusion of small regions.
Detailed information about mean-shift segmentation and its software, the EDISON system can be found in Refs[4, 15]. In this paper, we are only concerned with its role in providing initial over-segmentation for following classification procedures, so only some issues regarding the parameter selection are addressed.

The issue is the optimization of spatial domain bandwidth $h_s$ and range domain bandwidth $h_r$. Ideally, the bandwidth should be adaptively evaluated by the local distribution of the joint features (spatial coordinate and pixel spectral value) in the feature domain, because inappropriate bandwidth can cause modes to be over-merged or under-merged. However, an over-segmentation initial result with small fixed-bandwidth parameters is still acceptable in our classification task, because it can be regarded as a trade-off between the average region size and homogeneity. In the algorithm, the initial segments should be suited to represent the smallest object types to be analyzed in the later image processing step. In other words, if a region is a super-set of several potential ground objects, some classification errors will undoubtedly appear, because one patch will be labeled with only one ground type. To keep the balance between the region area and region homogeneity, we tended to choose a smaller bandwidth parameter. Because the statistic in regions with small pixels is not robust, multiscale regional spectral modes are considered. The minimum region number can be associated with the size of objects in a scene directly, which can be defined as the minimum region across all land types in which an observer is interested. In detail, the parameters for mean-shift clustering are listed as follows: $h_s=7$, $h_r=7$. Over-segmentation image objects are shown in Fig. 2.

![Over-Segmentation Using Mean Shift](image)

Figure 2. The Over-Segmentation Using Mean Shift

3. GEO-ONTOLOGY FRAMEWORK

An ontology models a domain in a formal way. It defines a set of concepts (water, no water, impervious, no impervious, house, road, grass, tree, etc) their characteristics and their relations to each other. In our case, each concept has a label (e.g. house, road, etc) by some attributes (corresponding to low-level descriptors) associated to an interval of accepted values[12]. We propose a model allowing the representation of geographic objects through an ontology and a matching process which allows to compare a region build during a segmentation and the different concepts defined in ontology.

The proposed ontology is composed of a hierarchy of concepts. In this hierarchy each node corresponds to a concept. Each concept has a label (e.g. road, house, etc) and is defined by its attributes (e.g. spectral, texture, size, shape, context, and structure, etc). Each attribute is associated to an interval of accepted values (e.g. [0-255]) and a weight (in [0-1]) representing its importance to class the geographical object corresponding to this concept. The values of these concepts haven filled by geographers experts thanks to their knowledge about the morphology of urban objects. In this paper, the hierarchy ontology framework is composed of 8 concepts (see figure 3), 10 attributes (see table1) and 5 leaf concepts. The depth of the ontology tree is 4[12].

![Ontology Diagram](image)
Mean Shift (MS) Algorithm is employed to generate over-segmentation image objects on high resolution remote sensing image over-segmentation image object are shown in figure 2. The CRF model is denoted as no direct graph $G = (V, E)$. The numbers of image objects are denoted as $V$. The no direct adjacent edges are denoted as $E$. Then $(X, Y)$ is a conditional random field in case, $X$ is denoted by random field as feature field, $x$ is denoted as random field reality, $Y$ is denoted as label field, $y$ is denoted as label field reality. $x$ is considered to be a set of feature vectors (spectral, texture, spatial, etc) $\{x_1, x_2, ..., x_I\}$, where $x_i = [x_{i1}, x_{i2}, ..., x_{id}]^T$ denotes a feature vector associated with an image objects site $i \in S$, $d$ is the number of feature, and $S \in \{1, 2, 3, ..., I\}$ is the set of image objects site. The label set of the whole image is give by $y = \{y_1, y_2, ..., y_I\}$, where $y_i \in \{1, 2, ..., L\}$ and $L$ is the number of classes.

Under the Bayesian framework, image classification generally considers the posterior $P(y|x)$, in the MRF framework, the posterior is formulated as $P(y|x) = P(x|y)P(y)$, where the prior distribution $P(y)$ is usually formulated as a Gibbs distribution, and $y$ is said to be an MRF. For computational tractability, the observed feature values are assumed to be conditional independent. Then, the likelihood $P(x|y)$ has a factored form $\prod_{i \in S}P(x_i|y_i)$ and the posterior is finally expressed as

$$P(y|x) \propto \prod_{c \in C} \Phi_c (y_c, \nu) \prod_{i \in S} P(x_i|y_i) \quad (1)$$

Where $c$ is a clique, $C$ is the set of cliques, $\Phi_c (y_c, \nu)$ is the potential function corresponding to clique $c$ in the Gibbs distribution $P(y)$, and $\nu$ is the set of MRF parameters. We can see from the formulation (1) that the modeling of posterior $P(y|x)$ involves implicit modeling of the likelihood. In contrast, the discriminative CRF framework directly models the posterior as a Gibbs distribution with the following form[16]:

$$P(y|x, \theta) = \frac{1}{Z(x, \theta)} \prod_{c \in C} \Psi_c(y_c, x, \theta) \quad (2)$$

Where $Z(x, \theta) = \sum_y \prod_{c \in C} \Psi_c(y_c, x, \theta)$ is a normalizing constant known as the partition function, and $\Psi_c$ is the potential function corresponding to clique $c$ with parameters $\theta$.

Table 1. attributes of ontology

<table>
<thead>
<tr>
<th>Attribute class</th>
<th>Attribute name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral</td>
<td>Spectral_Bule</td>
</tr>
<tr>
<td></td>
<td>Spectral_Red</td>
</tr>
<tr>
<td></td>
<td>Spectral_Green</td>
</tr>
<tr>
<td></td>
<td>Spectral_NDVI</td>
</tr>
<tr>
<td>texture</td>
<td>GLCM</td>
</tr>
<tr>
<td>Spatial</td>
<td>Diameter</td>
</tr>
<tr>
<td></td>
<td>Area</td>
</tr>
<tr>
<td></td>
<td>Perimeter</td>
</tr>
<tr>
<td></td>
<td>Miller index</td>
</tr>
<tr>
<td></td>
<td>elevation</td>
</tr>
</tbody>
</table>
5. IMAGE CLASSIFICATION BASED ON GEO-ONTOLOGY AND CONDITIONAL RANDOM FIELDS

Traditionally, satellite image processing and classification is based on multispectral supervised classification methods. The Hierarchical semantic information is considered. The binary tree of geo-ontology semantic is constructed (see figure 3) in this paper. The geo-ontology semantic information is considered as prior knowledge. The tree-structured CRF based on geo-ontology is presented to remote sensing image classification [16, 17].

The tree-structured CRF is considered as a binary tree $T$, identified by its nodes (classification number) and their mutual relationships. Expect for the root, each node $i$ has one parent $u(i)$, and each internal node has two children $l(i)$ and $r(i)$. We also define $\tilde{T} = \{ i \in T : l(i) = r(i) = \phi \}$, the set of terminal nodes or leaves, $\tilde{T} = T - \tilde{T}$, the set of internal nodes. With each node $i$ in $T$ we associate the following items: a set of sites $S^i \subseteq S$, corresponding to a segment of the image; a binary CRF $Y^i = \{ Y^i_s : s \in S^i \}$ with neighborhood system $\eta^i$, and realization $y^i$ where $y^i_s \in \{ l(i), r(i) \}$; a set of parameters $\theta^i$ that specify the potentials $V^i_c (\cdot)$ of the Gibbs distribution $P(y^i) = P(y^i; \theta^i)$.

Now we imposed the additional constraint that the set of sites associated with any given node is obtained from the binary classification of the parent set of sites. More formally, for each internal node of the tree $i \in \tilde{T}$

\[
S^i(l(i)) = \{ s : s \in S^i : y^i_s = l(i) \} \tag{3}
\]

\[
S^i(r(i)) = \{ s : s \in S^i : y^i_s = r(i) \} \tag{4}
\]

Given this condition, we define the Tree-structured CRF $Y^T$ associated with the tree $T$, as the set of all binary fields associated with the internal nodes of $T$ with realization $y^T = \bigcup_{i \in \tilde{T}} y^i$ and distribution law $P(y^T) = P(x^i, i \in \tilde{T})$.

As an example, consider the tree $T$ depicted in figures, with the interior nodes $\tilde{T} = \{ 1, 2, 4, 5 \}$ and the leaves $\tilde{T} = \{ 3, 6, 7, 8, 9 \}$, the tree-structured CRF associated with $T$ is $Y^T = \{ Y^1, Y^2, Y^4, Y^5 \}$, and its realization $y^T = \{ y^1, y^2, y^4, y^5 \}$ has probability distribution

\[
P(y^T) = P(y^1, y^2, y^4, y^5) = P(y^1)P(y^2 | y^1)P(y^4 | y^2)P(y^5 | y^2) \tag{5}
\]

$Y^i$ is a binary CRF defined on the set of sites $S^i = S$, with neighborhood system $\eta^i$ and potential $V^1_c (\cdot)$, defined on the cliques $c \in C^i$, where $C^i$ is the set of cliques of interest for $Y^i$. Therefore

\[
P(y^1) = \frac{1}{Z^1} \exp \left[ - \sum_{c \in C^i} V^1_c (y^1) \right] \tag{6}
\]

\[
P(y^2 | y^1) = \frac{1}{Z^2(x^i)} \exp \left[ - \sum_{c \in C^i(x^i)} V^2_c (y^2) \right] \tag{7}
\]

\[
P(y^4 | y^2) = \frac{1}{Z^4(x^i)} \exp \left[ - \sum_{c \in C^i(x^2)} V^4_c (y^4) \right] \tag{8}
\]

\[
P(y^5 | y^2) = \frac{1}{Z^5(x^i)} \exp \left[ - \sum_{c \in C^i(x^2)} V^5_c (y^5) \right] \tag{9}
\]
6. EXPERIMENTATIONS

The remote sensing image used in this experiment refers to the block of Changsha and was acquired in 2009 from GeoEye satellite. The GeoEye image consists of panchromatic image with 0.41m and four Multi-spectral bands with 1.65m, then this experiment takes three bands of fused image with 0.5m. the image comprised 600 lines and 600 columns.

Figure 4 provides some visual results. In part (a), a sample band of the original test image is reported to allow for a visual inspection of results. Then, parts (b)-(e) show the output of the proposed technique at various steps, namely L for going from 2 to 5, with region colors selected so as to simplify visual analysis. Focusing on the case $L = 5$, it can be observed that the algorithm provides reasonable segmentations.

![Figure 4: Classification based on Geo-ontology and Conditional Random Fields](image)

7. CONCLUSION

We have presented a new classification algorithm based on Geo-ontology and Conditional Random Fields. The algorithm has been realized based on orient-object image analysis method. The algorithm can be quite fast because all classifications are binary. The experiments show that the proposed algorithm performs very well on simple, relatively clean. The classification process has Hierarchical semantic information. The different hierarchical classification results are obtained based on hierarchical geo-Geo-ontology. The feature research will concentrate on an improvement of the CRF model and hierarchical classification results fusion.

REFERENCES

ACKNOWLEDGMENTS

This work was supported in part by the national natural science foundation(41201463), the postdoctoral foundation of central south of university, the foundation of department of education of Yunnan Province under Grant 2011Y311, the Chinese 973 program (2012CB719906), and then 863 High Technology Program of China under Grant 2012AA121400.