A NEW GENERATION IMAGE INTERPRETATION TECHNOLOGY BASED ON OBJECT-ORIENTED SEGMENTATION AND CLASSIFICATION

Yan Li, PhD
International Institute for Earth System Science, Nanjing University
Nanjing, Jiangsu, China, 210093
liyan@nju.edu.cn

Peng Gong, Professor
The State Laboratory of Remote Sensing Science, Institute of Remote Sensing, CAI
Beijing, China, 100101
gong@irma.ac.cn

ABSTRACT

A new generation object-oriented remote sensing image interpretation technique is presented in this paper. There are many classifications in practical use classifying the image at pixel level. Thus the result is sometimes difficult to catch on, and causes noise points due to errors. Besides, the problems of edge uncertainty is usually encountered, even for the cases of multi-scale methods. For most remote sensing applications, the area containing a single or pure ground objects to some spatial scale is more prefer than the single pixels. Such an object level is superior to the pixel level because of its less number of objects and convenience in texture analysis. In our study, the segmentation based on the object is implemented before the further image analysis. The similarity measurement is defined from both the spectral feature and the morphological feature of the neighboring object pair. The classification is carried out to the objects more than the pixels as usual in most remote sensing software. The classification noise can be effectively eliminated. On the other hand, because the edge between the different ground object is preserved well in the segmentation stage even in the large scales. It is convenient to extract the object feature by shape analysis. In the case study, we present the examples of shadow objects extraction and building objects extraction.

INTRODUCTION

Recent period, a new classification concept of object-oriented classification has drawn a great concerning in high resolution remote sensing fields. Recent studies have proven the superiority of the new concept over traditional classifiers(Ursula, 2004) (Darwish, 2003) (Mitri,2002) (Niemeyer,2001) (Esch,2003)(Sande,2003). The basic elements of an object-oriented approach are image objects. Image objects are contiguous regions in an image. Comparison to pixels, image objects carry much more useful information. Thus, they can be characterized by far more properties such as form, texture, neighborhood or context, than pure spectral or spectral derivative information (Baatz,1999). In our study, the segmentation based on the object is implemented before the further image analysis. The similarity measurement is defined from both the spectral feature and the morphological feature of the neighboring object pair. The classification is carried out to the objects more than the pixels as usual in most remote sensing software. The classification noise can be effectively eliminated.

SEGMENTATION

The segmentation is initialized by single pixels. When object scale goes to large, similar objects are merged as a new object. It is a kind of multi scale segmentation, from fine scale with small threshold to rough scale with high threshold, segmentation show itself a multi scale characteristics.
Object features: Spectral features and morphological features

Spectral features. Spectral features are derived from the standard deviation of the object. Suppose $\sigma_{N,a}$ and $\sigma_{N,b}$ refers to the standard deviation of object A and object B in channel N respectively. The merged object of A and B is represented by AB, i.e., $AB = A \cup B$. $\sigma_{N,ab}$ refers to the standard deviation of AB in channel N. It is approximated from $\sigma_{N,a}$, $\sigma_{N,b}$, and the averages $m_{N,a}$ and $m_{N,b}$ of A and B. Difference in spectral heterogeneity $h_p$ for the merged object is defined as

$$\Delta h_p = \sum_N w_N (n_{a,N} \sigma_{N,a} - (n_{a,N} \sigma_{N,a} + n_{b,N} \sigma_{N,b})), \quad 0 < w_N \leq 1, N = 0, 1, ... \sum_N w_N = 1$$ (1)

Morphological features. Difference in shape heterogeneity is defined as

$$\Delta h_s = w_c \Delta h_c + w_s \Delta h_s, \quad 0 < w_c, w_s \leq 1, w_c + w_s = 1$$ (2)

where $\Delta h_c$ and $\Delta h_s$ refers to the difference of compactness and smoothness heterogeneity of the objects respectively. They are estimated by the perimeter and area of the object $A$, $B$, and $AB$ respectively.

Merging Rule

Defined merging parameter $r$ as following,

$$r = w_p \Delta h_p + w_i \Delta h_i, \quad 0 < w_p, w_i \leq 1, w_p + w_i = 1$$ (3)

The neighboring objects with the least $r$ should be merged in an iterative. $r$ is getting larger and larger in this merging procedure. When $r$ reaches a threshold $R$ defined by user, the merging procedure stops. In the implementation, $r$ is assigned the value from 0 to R, and in an iterative all the neighboring objects with merging parameter equal to $r$ should be merged. Figure 1(a) shows a part of image of Songhua River from http://asterweb.jpl.nasa.gov. Figure 1(b) shows the objects edges on different thresholds 10, 20, 30, where rate of morphological feature is 0.2 with compactness rate of 0.1 and smoothness rate of 0.9. The merging tree in Figure 1(c) illustrates the objects merging procedure for the multi scale segmentation.

![Figure 1. Object oriented segmentation](image)

CLASSIFICATION

We employ two classifiers for the object oriented classification. The input of the classifier is not features of pixel, but features of the object derived from the object orientated segmentation. The color of an object is represented by the average intensities of channels of the object, as the input features.
C-Nearest-Neighbor (CNN)

It is a supervised classification. In the procedure of training, the user should choose samples for each class as training sets. Suppose we have K classes. In classification procedure,

1. Computes the Euclidean distance of a sample(object) s to all the training samples.
2. Resort the distances from least to greatest. Decide the first C distances and the associated training samples \( t_c, c=0,1,\ldots,C-1 \).
3. Vote method is used to decide the class belonging of sample s. If training sample \( t_c, c=0,1,\ldots,C-1 \), is a sample of class \( k, k=0,1,\ldots,K-1 \), then the voting number of class \( k \) \( v(k) \) plus 1. From all the K classes, the one with the greatest \( v(k) \) is the class that sample s belongs to.

Figure 2(a) is the segments of a larger region to Figure 1(a). The class number is taken as three, as pool, cropland, and watercourse. The red objects represent the training samples of watercourse, the yellow ones of the cropland, and the green ones of the pool. Figure 2(b) shows the CNN classification result. The color of a class is taken as the average color of the ones of this class. Some cropland objects in the upper left part are taken as pools. Some objects belonging to pool in the middle bottom are taken as the croplands. The watercourse objects basically are segmented correctly. These error can be decreased by taking more classes, since the wrong pool objects have large distance from both cropland and pool. They should be classified furthermore to a new class. So do the wrong cropland objects.

![Image of classification results](image.png)

Figure 2. Object oriented CNN classification

Fuzzy Neural Network

The network structure. Neural networks is commonly used to construct a classifier. Because of the uncertainty of the data, fuzzy logical is introduced to the classification. There are many fuzzy neural classifiers. Here we employ one similar to the classification system of Nauck (Nauck, 1997) (Nauck, 1999).

It is a three-layer fuzzy perceptron. First layer \( U_1 \) has N nodes, where N equals the number of channels of the image. The hidden layer \( U_2 \) is the rule layer, with M nodes, where M equals the number of the rules. It is added dynamically in the training procedure. The last layer \( U_3 \) has K nodes, corresponding to K classes.
The fuzzy nets for the first layer. The input of channel n, n=0,1,...,N-1 equals the intensity or normalized intensity \( x_n \). The output activation for pattern p of input unit n, n=0,1,...,N-1 is noted by \( a_{n}^{(p)} \). It is decided by the degrees to the function of fuzzy sets \( f_{n,k}, n = 0,1,...,N-1; k = 0,1,...,K-1 \), for this channel.

\[
\begin{align*}
\alpha_{n}^{(p)} &= \max \{d_{n,0}^{(p)}, d_{n,k}^{(p)}\} & (4) \\
n_{k} &= \text{the fuzzy set corresponding to the greatest degree.}
\end{align*}
\]

The fuzzy function shape is chosen as a Gaussian. The initial parameters of Gaussian is decided by the training patterns.

\[
\begin{align*}
d_{n,k}^{(p)} &= \max \{d_{n,0}^{(p)}, ..., d_{n,K-1}^{(p)}\} & (5)
\end{align*}
\]

where \( d_{n,k}, k = 0, ..., K - 1 \) is the degree to the function of fuzzy set \( f_{n,k} \) for pattern p.

The Gaussian shape is asymmetrical. Gaussian mean for each class equals the average of the training patterns of the class. \( m_k \) is the resorted average mean of training patterns for class \( k \). \( m_k < m_{k+1} \). \( m_{k+1} - m_k = 4\sigma_k \), where \( \sigma_k \) refers to the variance of the Gaussian. The half side span of Gaussian equals 3 times of \( \sigma_k \). For the least mean class, the Gaussian is left shouldered, and for the greatest mean class, the Gaussian is right shouldered.

The Rule base for the hidden layer. The rule is a semantic set which includes small, middle, large, very large,… corresponding to each fuzzy set \( f_{n,k} \) in channel n. For the beginning, there is no rule in the rule base, in another words, the rule layer is empty. When next training pattern arrives, compute the activation \( \alpha_{n}^{(p)} \) for each
channel. The fuzzy set corresponding to \( a_n^{(p)} \) for channel \( n \) is noted by \( k_n \). Then a semantic set \( R = \{r_0, ..., r_{N-1}\} \) is generated for the pattern. For all the current rules, if no rule \( R_m = \{r_{m,0}, ..., r_{m,N-1}\} \) equals it, then add a new rule node in the hidden layer and a new rule in the rule base.

The output of the rule node is the activation \( a_m^{(p)} \) of the node.

\[
a_m^{(p)} = \min\{a_n^{(p)}\}, n = 0, ..., N-1, m = 0, ..., M-1
\]  

(6)

**The output layer.** \( U_3 \) has the same nodes with the class number \( K \), each corresponds to a class. Node \( c_k \) only connects with the rule nodes \( R, i = 0, ..., I_k - 1 \) that the training patterns generated them belong to class \( k \). \( I_k \) is the number of rules connecting with \( c_k \). The input of \( c_k \) equals the output of \( R \). The activation of \( c_k \) is noted by \( a_k^{(p)} \).

\[
a_k^{(p)} = \max\{a_i^{(p)}\}, i = 0, ..., I_k - 1, k = 0, ..., K - 1
\]  

(7)

The output of \( c_k \) is 1 or 0. If the input training pattern is of class \( k \), the output is 1, otherwise is 0.

**The learning procedure.**

1. Initial the parameters of the fuzzy sets for each channel. The mean of Gaussian is taken as the average of the training samples, as illustrated in Figure 4. The variance of Gaussian is taken as three fourth of the difference of the neighboring means, see Figure 4. The Gaussian corresponding to the minimum and maximum means take left shoulder and right shoulder respectively.

2. Build the rule base and the connection between rule nodes and output nodes using all the training patterns. The method is just introduced in equation (6) and the explanation above it.

3. Adjust parameters. For all the training patterns again, propagate them with the network. Use the rules to decide the activation \( a_m^{(p)} \). The classification of pattern \( p \) is the class with the greatest \( a_m^{(p)} \). Compute the error of node \( c_k \) by \( \delta_k = t_k - a_k^{(p)} \). \( t_k \) refers to the output of \( c_k \), taking 1 or 0. For all the rule nodes, compute the error \( \delta_m = a_m^{(p)}(1-a_m^{(p)})\delta_k \). For each rule with \( a_m^{(p)} > 0 \), find the channel \( n \) and the fuzzy set \( k \) corresponding to it. Adjust the parameters of the fuzzy set \( f_{n,k} \) according to \( \delta_m \) and the learning rate \( \rho \).

The iteration stops when a stop criteria is reached.

Figure 5 shows the fuzzy classification result for the objects same with those of Figure 2(a). The cropland errors in the left top are caused by reason similar with CNN classifier, however there are seldom watercourse errors in the middle bottom like for CNN classifier.
OBJECT ORIENTED FEATURE EXTRACTION

Feature detection for a particular target has been an important research all the time. Except for the techniques in frequency domain, the common methods carry out feature detection in the pixel level. We will present two case studies of object based shadow extraction and building extraction. One uses spectral parameter and the other uses morphological parameters.

Case Study: Shadow Detection

The big shadow of the building is common for the urban images. They can be used to calculate the height of the building. Sometimes we need to remove them for certain applications, such as change detection. Traditional shadow detection based on intensity analysis detects shadows in the pixel level. For example, Figure 6(a) is a block of buildings. Taking a threshold of 44 for an image of 24bit color image, the shadows are shown as black in Figure 6(b). To generate the shadow region, some techniques like open or close operations are needed.

In object oriented segmentation and classification technique, by setting a threshold to the objects, the shadow regions can be easily obtained. For the same image, the result is illustrated in Figure 6(c).
Case Study: Building Extraction

High resolution remote sensing images have found their applications in the field of urban modeling, digital city, change detection, and other GIS problems. The meter to centimeter spatial resolutions make the objects feature in the ground easily observed. The most popular problem is building, road, or tree extraction. They construct the major features in most urbans. Zheltov (2001) adopts the linear extraction method and combines with the rectangular building model to extract building information from aerial images. Sohn and Dowman (2001) extract the polygon of building based on Fourier Transform and Binary Space Partitioning tree and combined with Building Unit Shape knowledge. Stassopoulou (2000) combines multi-scale region segmentation based on canny operator with edge segmentation to extract regional features (geometry shape, radiation characteristic, context information), and extracts building features by Bayesian network. Lin and Nevatia (1998) derives the analytic geometric relationship between building margin line and building shadows according to general illumination model to analyze the relationship of different ground features. Zhang (Zhang, 2006) detect building roof by supervised learning of roof sub-objects.

Here we use object oriented concept and propose an approach of building extraction. The image is firstly segmented into objects using the method introduced above. An object can be the whole or part of the building depending on the parameters to the segmentation. We use the structure elements and constraints to the building to decide an object to be part of a building or not.
**Structure elements.** From the contour of an object, the Hough Transform histogram $h(\theta, \rho)$ is generated. The lines parameters can be extracted by iteratively detecting the peaks, and the line segment can be obtained by fitting.

Parallel and perpendicular: For all the lines detected, decide if the approximate parallel pairs and orthogonal pairs exist by compare the parameters.

**Morphological Constraining.** Because the main structure of a building is the lines which are orthogonal to each other, there will be two peaks $\theta_1, \theta_2$ for the histogram of the angle of the lines for the object of a building. They generally differs about 90 degree. These two angles represent the direction of the building. We fit a square with the minimum area using $\theta_1, \theta_2$ to cover the object. For the object of a building part, the size is limited. As the opposite example, the road objects generally have the long shapes. A constrain of size limitation can be used to distinguish road objects and building objects. Here we don’t employ the compactness as one of the topple constrain, because there may be holes within the outer contour of a building object, such as the windows, chimneys, or other fine structures. They will result in a compactness similar with a long shape object.

**Shadow elimination.** In high resolution image, the shadow of building sometimes has contour of perpendicular lines. A simple threshold method is used to eliminate the shadow region.

Figure 7 shows the building extracted from the image of Figure 6(a). There are some errors which are labeled with different colors. The objects labeled with red color refer to the objects which were not extracted as the building objects because of high shadow threshold. Those labeled with rose color refer to the ones not extracted because of no building structure detected. The green one is the other class object taken as building. Figure 8 shows another example of building extraction.

![Figure 7](image_url)

*Figure 7. The buildings extracted by object oriented method.*
CONCLUSION

Object oriented segmentation and classification is a trend in very high resolution image analysis. In our work, we used this segmentation approach to obtain the basic objects which are the substructure of various classes. They support semantic and meaningful interpretations than pixels. Two classifiers of fuzzy perceptron and C-nearest neighbor were used to implement classification on the objects. We present the case studies of shadow and building extractions. Object shape features are used in the building extraction and spectral features are used in the shadow extraction. The result shows that object based feature extraction gives clear and well edges and the less computational amount. In our future study, we will take the texture features of the objects, such as fractal parameters, for the classification. Before the object concept, texture features for a pixel were computed on a certain window center at it. It is difficult to decide the classes of the pixels at the edge between different ground classes. Errors usually happen at these locations. The object oriented segmentation gives correct edges of the objects, texture classification based on the objects is expected to perform better than pixel based ones. Texture feature will also be used in building extraction to eliminate the trees within buildings.

REFERENCES


