A New Framework of the Unsupervised Classification for High-Resolution Remote Sensing Image

Penglin Zhang¹, Zhiyong Lv*¹, Lipeng Gao² and Li Huang¹
¹School of Remote Sensing and Information Engineering, Wuhan University, No.129, LuoYu Road, Wuhan, China
²Key Laboratory for Terrestrial Environment and Geohazards Monitoring of State Bureau of Surveying and Mapping, China University of Mining and Technology, Xuzhou 221116, China
*corresponding author, e-mail: lvzhiyong_fly@163.com

Abstract
Classification plays a significant role in change detection when monitoring the evolution of the Earth’s surface. This paper proposes a novel object-oriented framework for the unsupervised classification of high-resolution remote sensing images based on Jenks’ optimization. The fractal net evolution approach is employed as an image segmental technique, the spectral feature of each image object is extracted, and an algorithm of Jenks’ optimization is adopted as a classifier. Two experiments with different image platforms are conducted to evaluate the performance of the proposed framework and to compare with other traditional unsupervised classification algorithms such as the iterative self-organizing data analysis technique algorithm and k-means clustering algorithms. The proposed approach is found to be feasible and valid.

Keywords: Object-oriented framework, unsupervised classification, high-resolution remote sensing image, Jenks’ optimization.

Copyright © 2012 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction
High-resolution remote sensing image has received much attention because of the detailed information it provides of the earth's surface. The problem of image classification lies in assigning a label to each image pixel[1]. In fact, the nature of classification is categorizing all pixels in a digital image into one of the several land cover classes. However, classification or information availability for high-resolution remote sensing images poses challenges [2]. Such difficulty is a result of this type of data adequately presenting detailed spectra, which may cause different objects to have spectral similarity, or similar objects to have different spectral reflectance. Thus, a “bottleneck” in classification occurs.

Algorithms of unsupervised classification are built to solve the site labeling problem without the need for training samples. For example, the familiar and classical iterative self-organizing data analysis technique algorithm (ISODATA)[3] and k-means algorithm[4] iteratively assign all pixels in a digital image into one of several clusters. Generally, an arbitrary initial cluster vector is first assigned; second, each pixel is classified into the closest cluster; and third, the new cluster mean vectors are calculated based on all the pixels in one cluster. The second and third steps are repeated until a small “change” is achieved between iterations. ISODATA is a more sophisticated version of k-means that allows the splitting and merging of classes [5]. In the iterative procedures, fuzzy sets describe the relationship between a pixel and a cluster, such as in literature [6] that used a fuzzy membership called the “fuzzy k-means” algorithm. Other methods for clustering include the fuzzy c-means [7] that is improved in [8], Markov random fields (MRF) [9], and the Bayesian network [10], another discriminative supervised neighborhood preserving embedding feature extraction for hyperspectral-image classification has been also introduced in[11]. MRF is similar to a Bayesian network in its representation of dependencies; the difference is that Bayesian networks are directed and acyclic, whereas MRF is undirected and possibly cyclic. As of this writing, a few unsupervised classifiers for remote sensing image based on artificial immune networks (aiNet) such as artificial immune systems (AISs) [12] and remote sensing unsupervised artificial immune network (RSUAIN) have been reported [5]. RSUAIN utilizes the aiNet advantages by decreasing the number of user-defined
parameters. Neural net technique is used widely in image processing research, such as an improved technique for image fused which was also base Neural Net [13]. However, the traditional pixel-wise methods based on spectral information cannot provide satisfactory results for unsupervised classification [14].

Based on the aforementioned studies, this paper proposes a novel object-oriented framework of the unsupervised classification for high-resolution images. Numerous object-oriented techniques proved their advantage for high-resolution imagery [15]. However, few of these techniques have been applied in the unsupervised classification of high-resolution images. The novelty of this paper lies in that fact that an object-oriented segmental algorithm called the fractal net evolution approach (FNEA) [16] is integrated with Jenks’ optimization algorithm [17] to construct an unsupervised classification framework. Jenks’ optimization algorithm has been used as a data classifier in geographic information system (GIS) thematic map representations [18], but to the best of our knowledge, the algorithm has not yet been applied to the unsupervised classification of high-resolution images. Therefore, the performance of Jenks’ optimization algorithm in such classification is worth evaluating.

The rest of this paper is organized as follows. Section 2 describes the proposed classification framework. Section 3 discusses the experiments and the comparative analysis. Finally, Section 4 draws the conclusion.

2. The Proposed Framework

Unlike traditional pixel-based methods of unsupervised classification, the basic idea of the proposed framework is to classify high-resolution images in an object unit. Object-oriented approaches for high-resolution images are immune to the “salt-pepper” performance during classification. As shown in Figure 1, the proposed framework contains four processing phases: (1) Generation of the image-object according to the multi-resolution segmentation algorithm—the FNEA; (2) Feature extraction of each image-object using an auxiliary program; (3) Labeling each object using Jenks’ optimization approach; and (4) Post processing and accuracy evaluation. The following section details each step.

2.1 Segmentation of high-resolution image

This proposed unsupervised classification framework of high-resolution remote sensing images is similar to other object-oriented techniques that require image segmentation. The
FNEA is widely applied in high-resolution image segmentation such as in perspectives for environmental applications [19], evaluation of multi-scale landscape structure [20], and in detection of urban vegetation [21]. It is also effectively implemented as a function within the eCognition original object-based image analysis software[22]. Therefore, this paper adopts FNEA to extract objects from images. FNEA is a bottom-up segmentation algorithm based on a pairwise region merging technique, which is started from a single pixel. The process is an optimization that minimizes the average heterogeneity and maximizes the respective homogeneity for a given number of images. Region merging criterion as well as more detailed explanation and description can be found in [14].

In applying FNEA, the weights of the spectra and shapes for the segmental scale parameters are important. The scale parameter is an abstract term that determines the maximum heterogeneity allowed for the resulting image objects. Changing the value of a scale parameter can adjust the size of image objects. For heterogeneous images, the resulting objects for a given scale parameter are smaller than those for homogeneous images. Scale parameter refers to the region-homogeneity defined by the composition of homogeneity criterion, the weights of which are depicted in Equation (1) [23].

\[
H_{\text{homo}} = \omega_{\text{spectra}} h_{\text{spectra}} + (1 - \omega_{\text{spectra}}) h_{\text{shape}},
\]  
(1)

\[
h_{\text{spectra}} = \frac{\sum_{i=0}^{N} p(i,j)}{1 + (i-j)^2},
\]  
(2)

\[
h_{\text{shape}} = \omega_{\text{cmpt}} h_{\text{cmpt}} + (1 - \omega_{\text{cmpt}}) h_{\text{smooth}},
\]  
(3)

Where \( H_{\text{homo}} \) is the region homogeneity that consists of the spectra (\( h_{\text{spectra}} \)) and shape (\( h_{\text{shape}} \)). \( 0 \leq \omega_{\text{spectra}} \leq 1 \) is the user-defined weight for spectra (relative to the shape weight). Emphasis on the spectra provides a larger value of \( \omega_{\text{spectra}} \); conversely, importance on the shape provides a larger \((1 - \omega_{\text{spectra}})\). \( h_{\text{spectra}} \) represents the spectral change criteria in heterogeneity, as given by Equation (2), where \( p(i,j) \) is the spectral feature value of \((i,j)\), and \( N \) represents the total pixel number in the current region polygon. \( h_{\text{shape}} \) is given by Equation (3), where \( 0 \leq \omega_{\text{cmpt}} \leq 1 \) is the weight of the compactness \( h_{\text{cmpt}} \), and \((1 - \omega_{\text{cmpt}})\) is the weight of smoothness \( h_{\text{smooth}} \). The compactness criterion is used to optimize image objects in terms of compactness. The resulting image objects are optimized using the smoothness criterion in terms of smoothing the borders with the shape criterion \( h_{\text{shape}} \). Equations (4) and (5) show the formula for compactness and smoothness criteria, respectively.

\[
h_{\text{cmpt}} = \frac{4 \cdot \pi \cdot s}{l^2},
\]  
(4)

\[
h_{\text{smooth}} = \frac{l}{4 \cdot \sqrt[4]{s}},
\]  
(5)

Where \( s \) is the area of the region polygon, and \( l \) is the perimeter of the region polygon. More details can be found in [24].

In actual applications, these homogeneity criteria may be applied in different ways. Although in most cases, the spectral criterion is the most important for creating meaningful objects, a certain degree of shape homogeneity often improves the quality of object extraction because of the compactness of spatial objects associated with the image shape concept. Therefore, \( h_{\text{shape}} \) is especially helpful in preventing highly fractured image object from resulting in strongly textured data. Selecting appropriate scale parameters and weights for the identified preparation image is a repetitive process that avoids an overall or weak segmentation. An interpreter often decides on scale parameters based on users’ visual inspection of the image.
2.2 Feature extraction of image objects

After segmentation, the image scene is divided into many linked homogeneous regions, which are also called the “image object.” Segmental objects within an image do not overlap. Thus, statistics of pixel value can obtain the image spectral or textural features of each region. Two main steps for feature extraction of image objects are detailed as follows:

(1) The segmental result is converted into a polygon vector layer and is then saved as an Esri shapefile, a popular geospatial vector data format. The Esri shapefile can store non-topological geometry and attribute information for the spatial features in a data set, which minimally consists of a main file, an index file, and a database table[25]. The segmental objects of an image are exported to a vector shapefile, which is a preparatory measure for the subsequent step of storing the corresponding image feature of each image object.

(2) Each object in the shapefile is traversed, the image spectral features are made static, and then the shapefile is saved. Obtaining the spectral feature of an object requires the superposition of the image upon its corresponding image object vector. An image object is a set of pixels, which are constrained in a spectral or shape feature. The group of pixels makes up an image object, which is denoted by a record in the vector shapefile. Equation (6) provides the spectral feature of the \( l \)-th object in the vector shapefile.

\[
\overline{P_l} = \frac{1}{N_l} \sum_{(i,j) \in D_l} p(i, j),
\]

Where \( \overline{P_l} \) represents the mean value of the \( l \)-th image object for the spectra, \( p(i, j) \) is the spectral value of the pixel at \( (i,j) \), \( N_l \) is the total pixel number in the \( l \)-th object, and \( D_l \) is the domain of the \( l \)-th object for the spectral value. In other words, point \((i,j)\) is contained by the geometry of the \( l \)-th object. Thus, \( \overline{P_l} \) is taken as an attribute of the \( l \)-th object stored in the shapefile.

2.3 Classification using Jenks’ optimization algorithm

Jenks’ optimization method, also called the natural breaks classification method, is designed to determine the best arrangement of values into different classes. This method is done by minimizing each class’s average deviation from the class mean while maximizing each class’s deviation from the means of other groups. Thus, the method reduces the variances within classes and maximizes the variances between classes [26].

Similar to other unsupervised methods, Jenks’ optimization method requires an iterative process. That is, calculations are repeated using different breaks in the data set to determine which set of breaks provides the smallest in-class variance. In other words, Jenks’ optimization method is used to minimize the squared deviations of the class means. Optimization is achieved upon maximization of the quantity goodness of variance fit (GVF). The process begins by dividing the ordered data into groups. Initial group divisions can be arbitrary. The entire procedure consists of six steps:

Step 1: The number of classes \( k \) is specified; the spectral feature of object \( (\overline{P_l}) \) is taken as a feature for classification.

Step 2: A set of \( k-1 \) random values is generated in the range \([\min(\overline{P_l}), \max(\overline{P_l})]\), which is used as initial class boundaries.

Step 3: The sum of squared deviations between classes \( (SDBC) \) is calculated.

Step 4: The sum of squared deviations from the array mean \( (SDAM) \) is calculated.

Step 5: The \( SDBC \) is subtracted from the \( SDAM \) \( (SDAM-SDBC) \). The difference is the sum of the squared deviations from the class means \( (SDCM) \).

Step 6: The GVF value given by Equation (7) is calculated.

The above description of Jenks’ classification shows its iterative nature. The method first specifies an arbitrary grouping of the numeric data. \( SDAM \) is a constant that changes only when the data changes. The mean of each class and the \( SDCM \) are calculated. Observations are then moved from one class to another to reduce the sum of \( SDCM \) and thereby increase the GVF statistic. This process (Steps 2–6) continues until the GVF value is maximized and optimization is achieved.
Jenks’ classification method has been applied in thematic map representations[27] made available in Esri’s ArcGIS software. In our study, Jenks’ classification method is employed as an unsupervised classification tool for high-resolution images. We refer the reader to [26] for the definition of the important convergence criteria, GVF, which is given as Equation (7).

\[
GVF = \frac{SDAM - SDCM}{SDAM} = 1 \cdot \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (z_{ij} - \bar{z}_{j})^2} {\sum_{i=1}^{n} (z_{i} - \bar{z})^2},
\]  

(7)

Where \( k \) is the number of classes, \( n \) denotes the number of class-specific object, and \( z_{ij} \) represents the spectral feature value of the \( j \)-th class and the \( i \)-th region, which corresponds to the spectral feature of an image object. \( \bar{z}_{j} \) is the mean of the \( j \)-th class based on the spectral feature, which can be obtained by Equation (8).

\[
\bar{z}_{k} = \frac{1}{M_{k}} \sum_{i=1}^{M_{k}} z_{ik},
\]  

(8)

Where \( M_{k} \) is the total number of objects in the \( k \)-th class, \( z_{ik} = \bar{P}_{ik} \), which is given by Equation (6). Thus, \( \bar{z}_{k} \) is the mean value of the \( k \)-th class based on the region attribute, \( z_{ik} \).

Other parameters in Equation (7), such as \( z_{ij} \), \( \bar{z} \), and \( z_{i} \), can also be similarly deduced.

2.4 Post-processing of classification

In this study, post-processing consists of two steps: pruning small regions through spatial topology relationship and correcting the error of classification by priori knowledge. The first step is achieved by following two rules: (1) The regions labeled as the same class and share edges are merged; (2) Smaller regions contained by other larger regions are eroded. \( s_{1} \) and \( s_{2} \) are the areas of the small and larger regions, respectively, and \( T_{1} \) is the threshold. If \( s_{1} / s_{2} \geq T_{1} \), then \( s_{1} \) is merged with \( s_{2} \). For the other step, we correct the misclassification by using prior knowledge. This study uses four knowledge rules, which are as follows:

R1: If \( A \) is water, \( B \) is a building, and \( A \) shares an edge with \( B \), then \( A \) is a shade.
R2: If \( A \) is a building, \( B \) is a shade, \( A \) does not share an edge with \( B \), \( L \) and \( W \) are the length and width of the boundary rectangle of \( A \), respectively, while \( L / W \geq T_{1} \), then \( A \) is a road.
R3: If \( A \) is water, \( B \) is grass or a road, and \( A \) is contained by \( B \), then \( A \) is a shade.
R4: If \( A \) is a road, \( B \) is a shade, \( A \) shares an edge with \( B \), \( L \) and \( W \) are the length and width of the boundary rectangle of \( A \), respectively, while \( L / W \geq T_{1} \), then \( A \) is a building.

In our study, \( T_{1}, T_{2}, T_{3} \) and \( T_{4} \) are chosen by experiment. \( T_{1} \) and \( T_{2} \) are the ratios of length/width. Larger values indicate that the object envelope is longer and narrower; otherwise, the object envelope is broader. The roof of the building and the road appear similar in the image and are often misclassified. However, they differ in the ratio of length/width. Thus, the misclassification can be reviewed and corrected.

3. Experiment and Accuracy Analysis

In this section, two remote sensing images with high geometrical resolutions validate the effectiveness and the robustness of the proposed unsupervised framework for classification. One of the two images for evaluation is an aerial image with a higher geometrical resolution of the ground and more spectral noise than the satellite image. The other image is a panchromatic SPOT5 image of Tianjin, which is a city in northern China. In both experiments, the proposed framework of classification is compared with ISODATA and k-means.
3.1 Experiment 1: the aerial orthophoto data

The experimental data are acquired by an aerial data service camera, which contains 625 scan lines with 390625 pixels and a resolution of R=0.32 m. Typically, an urban infrastructure consists of buildings, roads, shade, and water. However, these objects usually present ambiguous spectral features in a high-resolution image. At the same time, the same object type can present differing spectral features. Hence, object-oriented segmentation and priori knowledge are needed for object discrimination.

The primary user-defined parameters are the segmental scale S, number of classes k, maximum iterations I, and the thresholds for post-classification T1, T2, and T3. For the proposed framework and the experimental data, the value of the parameters is set by experience at S=50 (shape=0.1 and compactness=0.5). Here, k=5 because the image is expected to fall into five classes. For convenient comparison of the proposed approach with ISODATA and k-means, the parameters for the latter two are set as follows. For ISODATA: maximum iteration is set at I=20.0, k=5.0, change threshold=10%, minimum number of pixels in class=10.0, minimum class distance=5.0, maximum class standard deviation=1.00, and maximum merge pairs=2. For k-means: k=5, change threshold=10%, I=20.0. In addition, the post-classification processing parameters are as follows: T1=25.0, T2=20.0 in R2, and T3=4.0 in R4.

Table 1 Classification accuracies for the aerial image using the proposed framework

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>building</th>
<th>water</th>
<th>road</th>
<th>grass</th>
<th>shade</th>
<th>Sum</th>
<th>PA</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>building</td>
<td>40357</td>
<td>12</td>
<td>2917</td>
<td>5854</td>
<td>294</td>
<td>49434</td>
<td>0.816</td>
<td>0.184</td>
</tr>
<tr>
<td>water</td>
<td>0</td>
<td>52994</td>
<td>157</td>
<td>1062</td>
<td>300</td>
<td>54413</td>
<td>0.972</td>
<td>0.028</td>
</tr>
<tr>
<td>road</td>
<td>4336</td>
<td>0</td>
<td>26752</td>
<td>1801</td>
<td>0</td>
<td>32889</td>
<td>0.813</td>
<td>0.187</td>
</tr>
<tr>
<td>grass</td>
<td>185</td>
<td>1674</td>
<td>1002</td>
<td>39923</td>
<td>4868</td>
<td>47652</td>
<td>0.838</td>
<td>0.162</td>
</tr>
<tr>
<td>shade</td>
<td>542</td>
<td>0</td>
<td>305</td>
<td>725</td>
<td>663</td>
<td>2235</td>
<td>0.297</td>
<td>0.703</td>
</tr>
<tr>
<td>sum</td>
<td>45420</td>
<td>54580</td>
<td>31133</td>
<td>49365</td>
<td>6125</td>
<td>186623</td>
<td>0.8605</td>
<td>0.147</td>
</tr>
<tr>
<td>PA</td>
<td>0.889</td>
<td>0.969</td>
<td>0.859</td>
<td>0.809</td>
<td>0.108</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OE</td>
<td>0.111</td>
<td>0.141</td>
<td>0.141</td>
<td>0.191</td>
<td>0.892</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Comparison of classification accuracies of the three methods for the aerial image

<table>
<thead>
<tr>
<th></th>
<th>ISODATA</th>
<th>k-means</th>
<th>Proposed Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Accuracy (AA)</td>
<td>62.724%</td>
<td>62.59%</td>
<td>72.68%</td>
</tr>
<tr>
<td>Overall Accuracy (OA)</td>
<td>77.1%</td>
<td>76.99%</td>
<td>86.05%</td>
</tr>
<tr>
<td>Kappa coefficient (KC)</td>
<td>0.6965</td>
<td>0.6952</td>
<td>0.8147</td>
</tr>
</tbody>
</table>

Table 1 lists the accuracies of the proposed framework classification. The class-specific accuracies are the producer's accuracy (PA) and the omission error (OE). The Kappa coefficient (KC) and the overall accuracy (OA) are computed based on the confusion matrix. The observed image is expected to fall into five classes: building, water, road, grass, and shade. Table 1 also indicates the classes and the number of their ground truths. For this experiment, the PA of shade has a lower value at PA (shade) =10.8% because the shape of the trees and the grass in the experimental image display similar spectra (Figure 2).

Table 2 Comparison of classification accuracies of the three methods for the aerial image

The proposed unsupervised framework for classification easily avoids the “salt-pepper” noise and is convenient for post-processing according to priori knowledge. We compare the classification accuracies of the proposed approach with those of ISODATA and k-means (Tables 1 and 2). In this experiment, the values of each accuracy (AA, PA, and KC) of ISODATA and k-means are similar, whereas the accuracies of the proposed approach indicate significant improvement. AA is improved by about 10%, OA is improved by about 9.06%, and KC is improved by about 0.1188.
Figures 2(a)–(f) illustrate the unsupervised classification results using ISODATA, k-means, and the proposed approach. The visual comparison indicates varying degrees of accuracy. The pixel-based classification method [Figures 2(c) and (d)] clearly denotes the “salt-pepper” effect and the misclassification between spectrally similar image objects. Compared with Figures 2(c) and (d), Figure 2(f) shows that using the proposed framework results in a reduced “salt-pepper” effect and in an improved visual interpretation.

3.2 Experiment 2: Tianjin Spot5 satellite image data

This experiment is conducted using a Spot5 image with 2.5 m pixels. The image (2162×1683 pixels) of Tianjin City, China, as shown in Figure 3(a), was acquired in May 2007. The survey image is part of a rural area and is different from that in experiment A. The primary objective of the survey district discriminates four classes: pools (including rivers, lakes, and canals), vegetation, street block, and bare land (including paved and unpaved roads).

The parameters for this experimental image are set as follows. For the proposed approach, segmental scale $S=150$ (shape=0.1, compactness=0.8) and $k=4$. ISODATA parameters are set as follows: maximum iteration is $I=20.0$, number of class is $k=4.0$, change threshold=10%, minimum number of pixels in class=5.0, minimum class distance=5.0, maximum class standard deviation=1.00, and maximum merge pairs=2. The k-means parameters are as follows: $k=4.0$, change threshold=10%, and $I=20$. In addition, $T1=25.0$, $T2=10.0$ (R2), and $T3=6.0$ (R4).
Table 3 Classification accuracies for Spot5 satellite image based on the proposed approach

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>pool</th>
<th>vegetation</th>
<th>street block</th>
<th>bare land</th>
<th>sum</th>
<th>PA</th>
<th>OE</th>
</tr>
</thead>
<tbody>
<tr>
<td>The proposed framework</td>
<td>48270</td>
<td>5806</td>
<td>686</td>
<td>31</td>
<td>54793</td>
<td>0.881</td>
<td>0.119</td>
</tr>
<tr>
<td>vegetation</td>
<td>1254</td>
<td>82933</td>
<td>92</td>
<td>0</td>
<td>84279</td>
<td>0.984</td>
<td>0.016</td>
</tr>
<tr>
<td>street block</td>
<td>0</td>
<td>1435</td>
<td>26663</td>
<td>3618</td>
<td>31716</td>
<td>0.841</td>
<td>0.159</td>
</tr>
<tr>
<td>bare land</td>
<td>442</td>
<td>3485</td>
<td>34879</td>
<td>100455</td>
<td>139261</td>
<td>0.721</td>
<td>0.279</td>
</tr>
<tr>
<td>sum</td>
<td>49966</td>
<td>93659</td>
<td>62320</td>
<td>104104</td>
<td>310049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.966</td>
<td>0.885</td>
<td>0.428</td>
<td>0.965</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OE</td>
<td>0.034</td>
<td>0.572</td>
<td>0.572</td>
<td>0.035</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OA=83.32%      KC=0.7676

Based on the above parameter settings, Table 3 reports the accuracies of the proposed framework, where the class-specific accuracies are OA, KC, PA, and OM. In particular, OA and KC are computed based on the confusion matrix. Relatively independent of the number of samples for each class, the average accuracies (AA) are also compared among the proposed approach, ISODATA, and k-means in Table 4. The proposed approach indicates improvements in OA and KC, but a decline in AA, especially for the experimental data with a resolution of 2.5 m. Such decline is attributed to lower pool accuracy at this particular resolution level.

Figure 3 Classification results of the Tianjin Spot5 image: (a) Spot5 image for Tianjin City; (b) ground reference data; (c) ISODATA classification; (d) k-means classification; (e) segmentation result; (f) classification of the proposed approach.

Table 4 Comparison of classification accuracies for the Spot5 satellite image

<table>
<thead>
<tr>
<th></th>
<th>ISODATA</th>
<th>k-means</th>
<th>Proposed Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Accuracy (AA)</td>
<td>86.01%</td>
<td>86.01%</td>
<td>81.1%</td>
</tr>
<tr>
<td>Overall Accuracy (OA)</td>
<td>82.26%</td>
<td>82.26%</td>
<td>83.3%</td>
</tr>
<tr>
<td>Kappa coefficient (KC)</td>
<td>0.74</td>
<td>0.74</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Table 4 lists the comparisons of accuracy, where ISODATA has an AA of 86.01%, an OA of 82.26%, and a KC of 0.74. For the proposed framework, OA is improved by about 1.04% and KC by about 0.03. AA is decreased by about 4.91% because the product accuracy of the street block is lower for the selected ground truth, which is about 42.8%. Clearly, the proposed framework outperforms the other two traditional techniques.

Figure 3 depicts the visual comparison among the proposed approach, ISODATA, and k-means, suggesting varying degrees of classification accuracy. In this study, the experimental image and the given parameters produce similar results for ISODATA and k-means. Comparing Figures 3(c) and (d) with Figure 3(f), the proposed unsupervised classified framework is clearly immune to the “salt-pepper” performance, and OA is improved.

4. Conclusion
In this study, we proposed a novel framework for the unsupervised classification of high-resolution remote sensing images. The novelty of this paper lies in two facts: (1) unlike the traditional pixel-based method, the proposed approach is object-oriented for unsupervised classification; (2) Jenks’ optimization is first introduced for unsupervised classification of high-resolution remote sensing images. Based on the proposed framework, our experiments delineated a more accurate classification, which indicated a significant improvement in “salt-pepper” performance. On the whole, the experiments on the two data sets showed that the proposed framework for unsupervised classification of high-resolution images provides higher accuracies compared with the classical and traditional methods of ISODATA and k-means algorithms. The primary results verified that the proposed framework could be effectively used in the unsupervised classification of high-resolution remote sensing images.

Acknowledgements
The authors would like to thank the anonymous reviewers for the insightful suggestions. This work was supported by the High technology supported plan of China under Grant 2012BAJ15B04.

References

A New Framework of the Unsupervised Classification for High-Resolution … (Penglin Zhang)