A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information

Miao Li ¹, Shuying Zang¹, Bing Zhang², Shanshan Li² and Changshan Wu ¹³*  

¹Key Laboratory of Remote Sensing Monitoring of Geographic Environment, College of Heilongjiang Province, Harbin Normal University, 150025 - Harbin, China  
²Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, n. 9 Dengzhuang South Road, Haidian District, 100094, Beijing – China  
³Department of Geography, University of Wisconsin-Milwaukee, 53201 - Milwaukee, U.S.A.  
*Corresponding author, e-mail address: cswu@uwm.edu

Abstract

This paper reviewed major remote sensing image classification techniques, including pixel-wise, sub-pixel-wise, and object-based image classification methods, and highlighted the importance of incorporating spatio-contextual information in remote sensing image classification. Further, this paper grouped spatio-contextual analysis techniques into three major categories, including 1) texture extraction, 2) Markov random fields (MRFs) modeling, and 3) image segmentation and object-based image analysis. Finally, this paper argued the necessity of developing geographic information analysis models for spatial-contextual classifications using two case studies.

Keywords: Remote sensing image classification, Spatio-contextual information, Geographic information analysis techniques, Land use land cover classification.

Introduction

Frequently updated land use land cover information is essential to many socio-economic and environmental applications, including urban and regional planning, natural resources conservation and management, etc. [Homer et al., 2007; Lu and Weng, 2007; Jensen, 2009]. Remote sensing imagery, covering a large geographic area with high temporal frequency, offers a unique opportunity for deriving land use and land cover information through the process of image interpretation and classification. For generating updated land use land cover information at different scales, remote sensing image classification techniques have been developed since 1980s. During 1980s and 1990s, most classification techniques employed the image pixel as the basic unit of analysis, with which each pixel is labeled as a single land use land cover class. With the pixel as the basic analysis unit, a series of classification techniques, such as unsupervised (i.e. k-means and ISODATA), supervised (i.e. maximum likelihood, artificial neural network, decision tree, support vector machine,
random forests), and hybrid classification (i.e. semi-supervised and fusion of supervised and unsupervised learning) [Zhang et al., 2005; Alajlan et al., 2012], have been developed. These pixel-wise classification approaches, when applied to heterogeneous regions, however, are with limitations, as the size of an object may be much smaller than the size of a pixel. In particular, a pixel may not only contain a single land use land cover type, but a mixture of several land use land cover types. As a result, fuzzy classification and spectral mixture analysis techniques have been developed in 1990s to address the mixing pixel problem [Adams et al., 1986; Wang, 1990], and such sub-pixel based analyses have been applied in geology, forestry, as well as urban analyses [Adams et al., 1986; Roberts et al., 1998; Wu and Murray, 2003]. With the launch of very-high-resolution (VHR) remote sensing sensors like IKONOS and QuickBird, object-based classification methods have been developed since late 1990s [Blaschke, 2010; Dribault et al., 2012; Wilson and Oreopoulos, 2013]. Object-based methods group a number of pixels with homogeneous properties into an object, and objects, instead of individual pixels, are considered the basic unit for analyses [Myint et al., 2011].

Although a large number of remote sensing classification techniques have been developed in recent decades [Lu and Weng, 2007], most methods only utilize spectral variables, and spatial information is more or less ignored. Spectra-based classification approaches are conceptually simple and easy to be implemented, but they neglect the spatial components, which are inherited in real-world remote sensing imagery [Moser et al., 2013]. This issue becomes severe with the availability of very high resolution (VHR) remote sensing imagery (i.e. IKONOS and QuickBird). With higher spatial resolutions, images are likely to have higher within-class spectral variability. As a result, less than satisfactory results have been reached with spectral classifiers [Myint et al., 2011]. Spatial information, extracted from a particular spectral band, the panchromatic band, or the first principal component of the image, therefore, has been incorporated into image classification [Blaschke, 2010]. In remote sensing literature, such approaches have been generally called “spatio-contextual” image classification, indicating the relationship between a “target” pixel and its neighboring pixels is incorporated into analyses [Tso and Mather, 1999]. These spatio-contextual image classification approaches can be grouped into three categories, including 1) texture extraction, 2) Markov random fields (MRFs) modeling, and 3) image segmentation and object-based image analysis [Stuckens et al., 2000; Blaschke, 2010; Thoonen et al., 2012; Moser et al., 2013]. Although these spatio-contextual approaches have been applied to derive land use land cover information with different degree of success, they are originated in the fields of computer vision and image processing, without taking geographical knowledge into consideration. Spatial dependence, however, is an essential concept in geography, as stated in Tobler’s first law of geography that “Everything is related to everything else, but near things are more related than distant things” [Tobler, 1970]. A series of geographic information analysis techniques, such as spatial autocorrelation analyses, spatial expansion models, spatial regression models, geographically weighted regression models, and geostatistics, have been developed to address spatial dependence issues [Fischer and Getis, 2010]. Although widely accepted in the fields of geography, geology, economics, and regional science, geographic information analysis techniques have rarely been applied in remote sensing image processing, especially for spatio-contextual image classifications. In recent studies, a fourth group of classification methods that incorporates spatio-contextual information, namely geographic information analysis techniques, was emerged.
Image classification methods

Pixel-wise image classification
As the classic remote sensing image classification technique, pixel-wise classification methods assume each pixel is pure and typically labeled as a single land use land cover type [Fisher, 1997; Xu et al., 2005] (see Tab. 1). With this method, remote sensing imagery is considered a collection of pixels with spectral information, and thereby spectral variables and their transformations (e.g. principal components, vegetation indices, etc.) are input to per-pixel classifiers. In general, pixel-wise classification algorithms can be divided into two groups: unsupervised classification and supervised classification. With unsupervised classifiers, a remote sensing image is divided into a number of classes based on the natural groupings of the image values, without the help of training data or prior knowledge of the study area [Lillesand et al., 2004; Puletti et al., 2014]. Two unsupervised classification algorithms, k-means [Rollet et al., 1998; Blanzieri and Melgani, 2008], and its variant, the Iterative Self-Organizing Data Analysis (ISODATA) technique, are the most commonly used classifiers [Dhodhi et al., 1999]. Recently, Self-Organizing Maps (SOM) method and hierarchical clustering methods were also developed for unsupervised classification [Goncalves et al., 2008]. In comparison, with supervised classifiers, an image analyst selects representative sample sites with known class types (i.e. training samples), and compares the spectral properties of each pixel in the image with those of the training samples, then labels the pixel to the class type according to decision rules [Lillesand et al., 2004]. A large number of supervised classification methods have been developed, and they include Maximum Likelihood Classifier (MLC) [Settle and Briggs, 1987; Shalaby and Tateishi, 2007], Minimum Distance-to-Means Classifier [Atkinson and Lewis, 2000; Dwivedi et al., 2004], Mahalanobis Distance Classifier [Deer and Eklund, 2003; Dwivedi et al., 2004], Parallelepiped [Perakis et al., 2000] and K-Nearest Neighbors Classifier [Zhu and Basir, 2005; Zhang et al., 2008], etc. Recently, machine learning techniques have also been developed to refine the knowledge learning process [Mountrakis et al., 2011], and these methods include artificial neural network [Kavzoglu and Mather, 2003], classification tree [Friedl and Brodley, 1997; Melver and Friedl, 2002; Jiang et al., 2012], random forests [Gislason et al., 2006], support vector machine [Gualtieri and Cromp, 1999; Huang et al., 2002; Pal and Mather, 2005; Marconcini et al., 2009], and genetic algorithms [Ishibuchi et al., 1994; Tseng et al., 2008].

Sub-pixel-wise image classification
Pixel-wise remote sensing image classification techniques assume that only one land use land cover type exists in each image pixel. However, such an assumption is often invalid for medium and coarse resolution imagery, majorly due to the heterogeneity of landscapes when compared to the spatial resolution of a remote sensing image [Lu and Weng, 2007]. As a result, the applications of pixel-wise hard classifications decrease the classification accuracy of land use land cover maps [Zhang and Foody, 1998; Pu et al., 2003; Shanmugam et al., 2006]. As a better alternative, sub-pixel classification techniques are considered more appropriate as the areal proportion of each land use land cover type can be accurately estimated [Foody and Cox, 1994; Zhang and Foody, 1998; Woodcock and Gopal, 2000]. Major sub-pixel classification techniques (see Tab. 1), such as fuzzy classification, neural networks [Foody, 1999; Kulkarni and Kamleesh, 1999; Mannan and Ray, 2003], regression modeling [Yang and Liu, 2005; Yuan et al., 2005], regression tree analysis [Yang et al., 2003; Xian and
Crane, 2005] and spectral mixture analysis [Adams et al., 1995; Roberts et al., 1998; Wu and Murray, 2003; Wu, 2004], have been developed to address the mixing pixel problem. Specifically, with the fuzzy representation, each pixel receives partial memberships of all classes, and the corresponding areal proportion of each class can be estimated accordingly [Zhang and Foody, 1998]. Ji and Jensen [1999] and Myint [2006] developed sub-pixel analysis methods to quantify the amount of urban impervious surfaces and urban vegetation. Roberts et al. [1998] developed multiple endmember spectral mixture analysis technique to map chaparral (Larrea tridentate), a shrubland plant community, in the Santa Monica Mountainous areas. Wu and Murray [2003] developed a four-endmember spectral mixture analysis method to estimate sub-pixel percent urban impervious surfaces. Tang et al. [2007] proposed fuzzy-spectral mixture analysis (fuzzy-SMA) model. Different from traditional SMA approaches, fuzzy-SMA obtained fuzzy mean and fuzzy covariance using training samples derived through SMA, and then applied with the conventional fuzzy classification.

Table 1 - Summary of remote sensing image classification techniques.

<table>
<thead>
<tr>
<th>Classification Techniques</th>
<th>Characteristics</th>
<th>Examples of classifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel-based techniques</td>
<td>Each pixel is assumed pure and typically labeled as a single land use land cover type</td>
<td>Unsupervised (e.g. k-means, ISODATA, SOM, hierarchical clustering) Supervised (e.g. Maximum likelihood, Minimum distance-to-means, Mahalanobis distance, Parallelepiped, k-nearest Neighbors) Machine learning (e.g. artificial neural network, classification tree, random forests, support vector machine, genetic algorithms)</td>
</tr>
<tr>
<td>Sub-pixel-based techniques</td>
<td>Each pixel is considered mixed, and the areal proportion of each class is estimated</td>
<td>Fuzzy classification, neural networks, regression modeling, regression tree analysis, spectral mixture analysis, fuzzy-spectral mixture analysis</td>
</tr>
<tr>
<td>Object-based techniques</td>
<td>Geographical objects, instead of individual pixels, are considered the basic unit</td>
<td>Image segmentation and object-based image analysis techniques (e.g. E-cognition, ArcGIS Feature Analyst)</td>
</tr>
</tbody>
</table>

**Object-based image classification**

Compared to traditional per-pixel and sub-pixel classification methods, object-based models provide a new paradigm to classify remote sensing imagery [Blaschke, 2010; Myint et al., 2011] (See Tab. 1). With object-based models, geographical objects, instead of individual pixels, are considered the basic unit for analysis. That is, instead of considering an image as a collection of individual pixels with spectral properties, object-based methods generate image objects through image segmentation [Pal and Bhandari, 1992], and then conduct image classification on objects rather than pixels. With image segmentation techniques, image objects are formed using spectral, spatial, and textural and contextual information. Then these objects are further classified using spectral and other relevant criteria. Object-based approaches are considered more appropriate for VHR remote sensing images since they assume that multiple image pixels form a geographic object. Many studies have proven that significant higher accuracy has been achieved with object-based approaches [Benz et al., 2004; Wang et al., 2004; Myint et al., 2011].
Spatio-contextual analysis techniques for image classification

Although spectral classifiers have the advantages of conceptual simplicity and computational effectiveness, their limitations are also obvious [Myint et al., 2011]. A number of land use land cover types cannot be effectively separated with spectral information, and thereby less than desired accuracy has been reported with spectra-only classifiers [Tso and Mather, 1999; Stuckens et al., 2000]. For example, there has been a consensus that impervious surfaces and bare soil (e.g. bright urban impervious surfaces and dry soil, and dark impervious surfaces and moist soil) cannot be effectively separated only with spectral information. In order to achieve higher classification accuracy, an increasing number of spatio-contextual analysis techniques have been developed recently to complement the spectral classification approaches [Atkinson and Naser, 2010; Moser et al., 2013]. In this review, we divide these spatio-contextual analysis techniques into three methodological approaches, including 1) texture extraction, 2) MRFs modeling, and 3) image segmentation and object-based image analysis [Thoonen et al., 2012; Moser et al., 2013]. A summary of these three methodological approaches is provided as follows (see Tab. 2).

<table>
<thead>
<tr>
<th>Spatio-contextual Classification Techniques</th>
<th>Role of spatio-contextual information</th>
<th>Classifier types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture extraction</td>
<td>Incorporation of texture metrics can improve the classification accuracy through mitigating spectral confusion among spectrally similar classes</td>
<td>Structural texture extraction (e.g. mathematical morphology techniques) Statistical texture extraction (e.g. first-order statistics, second-order statistics, texture spectrum, semivariance) Model-based texture extraction (e.g. fractal models, autoregressive models, MRFs models) Transform texture extraction (e.g. Fourier transform, Gabor transform, Wavelet transforms)</td>
</tr>
<tr>
<td>MRFs</td>
<td>MRFs incorporate spatio-contextual information into a classifier through modifying the discriminant function with an addition of spatial correlation term.</td>
<td>Integrated algorithm of MRFs and SVM Adaptive MRFs</td>
</tr>
<tr>
<td>Image segmentation and object-based image analysis</td>
<td>Spatio-contextual information has been incorporated in the image segmentation process, with each segment contains spatially contiguous and homogenous pixels, and avoids the salt-and-pepper noise.</td>
<td>Image segmentation (e.g. region-growing, Markovian methods, watershed methods, hierarchical algorithms) Object-based image analysis techniques (e.g. SVM, nearest neighbor classifier)</td>
</tr>
</tbody>
</table>

Table 2 - Summary of spatio-contextual remote sensing image classification techniques.
Texture extraction

Texture is a term of computer vision and image analysis that describes the placement and spatial arrangement of repetitions of tones, and is often employed to quantify the variability of pixels in a neighborhood [Jensen, 2009]. The applications of texture extraction in remote sensing image classifications can be traced back to 1970s [Haralick et al., 1973; Haralick, 1979], and numerous studies have proven that the incorporation of texture metrics can improve the classification accuracy through mitigating spectral confusion among spectrally similar classes [Carleer and Wolff, 2006]. Major texture extraction methods can be grouped into four major categories: 1) structural (including mathematical morphology), 2) statistical, 3) model-based, and 4) transform [Tuceryan and Jain, 1990; Materka and Strzelecki, 1998; Coburn and Roberts, 2004]. Structural approaches [Haralick, 1979] attempt to examine image textures through evaluating pre-defined primitives and spatial arrangements of these primitives. The texture of an image can be defined with the primitives and their placement rules. Recently, structural texture extraction approaches have been advanced through developing mathematical morphology techniques based on non-linear operators associated with Minkowski’s set theory [Haralick et al., 1987]. Especially, morphological profiling [Fauvel et al., 2008] and morphological attribute filters [Dalla Mura et al., 2010] have been developed to capture geometrical and multi-scale properties [Moser et al., 2013]. Statistical methods include first-order statistics (i.e. mean, standard deviation) [Collins and Woodcock, 1999] and second-order statistics, especially the grey-level co-occurrence matrix (GLCM) proposed by Haralick et al. [1973] and Haralick [1979]. With all the fourteen metrics developed by Haralick et al. [1973], six of them, including contrast, variance, correlation, energy, entropy, and inverse different moment, have been widely applied and achieved reasonably satisfactory results [Pacifici et al., 2009]. Other statistical texture metrics include those based on texture spectrum [Wang, 1990; Xu et al., 2003] and semivariance [Jensen, 2009]. The third category of texture extraction is the model-based approaches, such as fractal models [Lam, 1990], autoregressive models [De Souza, 1982], and MRFs models [Cross and Jain, 1983]. Finally, transform methods include Fourier, Wavelet transforms [Mallat, 1989], and Gabor [Daugman, 1985]. When compared to Fourier and Gabor, the wavelet transforms perform better as they are based on a multiple spatial resolutions, and a wide range of wavelet functions can be chosen to improve the classification accuracy. Texture information can be incorporated in the processes of image pre-classification (e.g. as an additional variable) and post-classification (e.g. image filtering) [Stuckens et al., 2000]. Several studies have proven that the integration of textural information into remote sensing image classification can generate better classification accuracy [Chen and Gong, 2004; Fauvel et al., 2008]. One limitation of texture extraction, though, is that unreliable classification results may exist, especially near the edges of different land covers [Fauvel et al., 2008].

MRFs models

Another family of spatio-contextual remote sensing image analysis techniques is MRFs models [Moser et al., 2013]. MRFs are originated in the fields of statistical physics, computer vision and image processing [Li, 1995; Li, 2009], and have recently been applied in the field of remote sensing image classification and interpretation [Jia and Richards, 2008; Zhang et al., 2011]. For remote sensing image classifications, MRFs incorporate spatio-contextual information into a classifier through modifying the discriminant function with an addition of spatial correlation term [Fauvel et al., 2013]. MRFs have the ability to examine the global and local properties of a remote
sensing image, and quantify the spatial autocorrelation among pixels through a mathematically rigorous means [Moser et al., 2013]. As a result, MRFs have been applied to solve many remote sensing image analysis problems, including classification, change detection, sub-pixel analysis, and segmentation [Shekarforoush et al., 1996; Jia and Richards, 2008; Fauvel et al., 2013]. Recently, an increasing number of studies have applied MRF-based image classification techniques and reported significantly better results when compared to the conventional non-contextual classification techniques. As an example, Tso and Olsen [2005] reported that the addition of contextual and edge information through the MRF-based methods improved both the visual interpretation and classification accuracy. Further, Zhang et al. [2011] showed that MRF-based classification methods (e.g. an integrated algorithm of support vector machine (SVM) and MRF, and an adaptive MRF algorithm) have significantly improved the classification accuracy (from 77% to 93%) when applied to the AVIRIS hyperspectral imagery of Indian Pines, Indiana, U.S.A.. Although a number of MRF-based classification techniques have been successfully applied in land use land cover classifications, the concepts of MRF are considered difficult to many remote sensing scientists, and their implementations involve challenging computational problems [Moser et al., 2013].

**Image segmentation and object-based image analysis (OBIA)**

The third major group is the image segmentation and object-based image analysis (OBIA) techniques [Blaschke, 2010]. Image segmentation is a term of computer vision, with which a digital image is partitioned into a number of homogeneous segments, each of which often corresponds to an object or a portion of an object [Pal and Bhandari, 1992]. Image segmentation techniques have been applied in content-based image retrieval, medical imaging, object detection, etc. [Pal and Bhandari, 1992]. In the field of remote sensing, an early image segmentation application was developed by Kettig and Landgrebe [1976], who later developed the ECHO classifier [Landgrebe, 2003]. In segmenting remote sensing images, spatio-contextual information has been incorporated in the algorithms, including region-growing [Mannan and Ray, 2003], Markovian methods [Jackson and Landgrebe, 2002], watershed methods [Salembier et al., 1998], and hierarchical algorithms [Dalla Mura et al., 2011]. As an example, with the region-growing method, a region grows through interactively comparing all neighboring pixels’ values to the region’s mean, and the pixels with small differences are allocated to the region [Wang et al., 2004]. As a result, each region contains spatially contiguous and homogenous pixels, and different regions are with a high degree of heterogeneity. With the segmented imagery, an OBIA classification technique (e.g. SVM, nearest neighbor classifier, etc.) can be applied to derive land use land cover maps. Object-based image classification techniques are considered superior when compared to traditional pixel-based techniques as they can incorporate spectral and spatio-contextual information in the classification process [Blaschke, 2010; Ceccarelli et al., 2013]. A number of recent studies have reported highly accurate classification results when applied to deriving high-spatial-resolution land use land cover maps in urban areas [Su et al., 2008; Blaschke, 2010; Gianinnetto et al., 2014].

**Geographic information analysis techniques**

Traditionally, geographic information analysis techniques are not considered as a major group of models that incorporating spatio-contextual information into remote sensing image classification [Thoonen et al., 2012; Fauvel et al., 2013; Moser et al., 2013]. This may be because of the gaps among different research communities, as texture extraction, MRF
models, and image segmentation and object-based classification are originated in the fields of computer vision, pattern recognition, and image analysis, while geographic information analysis models are embedded in the fields of geography, geology, soil science, economics, regional science, agricultural science, etc. Although originated from different research communities, these techniques attempt to address the same problem: spatial dependence. Recently, a number of geographic information analysis approaches have been incorporated in texture extraction for improving remote sensing image classification accuracy [Van der Meer, 2012]. In particular, variogram-based textural metrics have been input to classifiers [Atkinson and Lewis, 2000; Bahria et al., 2011; Adjorlolo and Mutanga, 2013], or served as a filter in post-classification [Atkinson and Naser, 2010]. In addition, spatial autocorrelation metrics, such as Moran’s Index and Getis statistic [Wulder and Boots, 1998; Myint, 2003; Emerson et al., 2005; Ghimire et al., 2010] have been incorporated as textural variables for image classification. To date, geographic information analysis techniques have rarely been directly applied into classifying remote sensing imagery [Atkinson and Naser, 2010]. For geostatistical techniques, one exception is the indicator kriging developed by Van der Meer [1994]. This method was further applied by Das and Singh [2009], who found that statistically more accurate results were obtained when compared to non-contextual techniques. In addition, Atkinson [2004] and Atkinson and Naser [2010] developed geostatistically weighted classifiers applied to a simulated image and an IKONOS image. In the field of spatial statistics, although spatial regression models (e.g. spatial error and spatial lag methods) and geographically weighted regression (GWR) models have been developed to examine spatial dependence and spatial non-stationary, they have rarely been considered in spatio-contextual image classifications. One interesting study was conducted by Shekhar et al. [2002], who compared the spatial autoregressive model and MRF model in terms of spatial data mining. Moreover, GWR models have been recently applied in examining the accuracies of remote sensing image classifications [Foody, 2005; Comber et al., 2012; Comber, 2013], as well as estimating spatially varying variables, such as surface salinity, housing price, etc. [Yu et al., 2007; Xie et al., 2013]. Recently, Zhang et al. [2013] and Deng and Wu [2013] introduced the concept of local (neighborhood) spatial autocorrelation into per-pixel and sub-pixel remote sensing image classifications. Specifically, Zhang et al. [2013] developed a neighborhood constrained metric, pure neighborhood index (PNI), and incorporated this index in a k-means classifier to classify hyperspectral remote sensing imagery. Deng and Wu [2013] estimated sub-pixel urban imperviousness using a spatially adaptive spectral mixture analysis (SASMA) approach, in which endmember candidates were selected within a neighborhood and synthetic endmembers were derived using an inverse distance weighting (IDW) method. In order to demonstrate the effectiveness of these techniques in spatial-contextual image classifications, we summarized the methods and results of these studies.

**Neighborhood-constrained k-means (NC-k-means) approach**

This case study involves the development of a new index, pure neighborhood index (PNI), and incorporates this index in the process of k-means classification approach. This index attempts to quantify the degree of spatial dependence for each class in an image. That is, if there are a large number of pure neighborhoods of a particular class, a neighborhood-based classification approach should be adopted; otherwise a pixel-level classification algorithm should be applied. In this case study, the PNI value is incorporated into the process of
the k-means classification approach to adjust the assigned class values between iterations. For a better explanation, the following sections provide the definition of PNI, steps of incorporating PNI into k-means classification technique, and results and conclusions.

**Pure neighborhood index (PNI)**

*PNI* is an indicator of whether a neighborhood only contains pixels of a particular class. With the classification results, *PNI* of a neighborhood equals to one if all of the pixels in the neighborhood are classified into one particular class, otherwise it is assigned to a value of zero [Eq. 1].

\[
PNI_{wk} = \begin{cases} 
  1 & \text{if } w \text{ is a pure neighborhood of } k \\
  0 & \text{if } w \text{ is a mixed neighborhood of } k 
\end{cases} \quad [1]
\]

where \( w \) represents a neighborhood, and \( k \) is a class.

With the *PNI* values of each non-overlapping neighborhood (\( w \)) for class \( k \), we can calculate the total number of non-overlapping pure neighborhoods (\( \delta_k \)) for class \( k \) in an image [see Eq. 2], and employ this number as the criteria to decide whether a pixel-level or a neighborhood-level k-means classification should be conducted.

\[
\delta_k = \sum_w PNI_{wk} \quad [2]
\]

**Neighborhood-constrained k-means (NC-k-means) classification**

With the values of *PNI* \( w_k \) and \( \delta_k \), NC-k-means algorithm is divided into four steps:

Step 1: perform the traditional k-means algorithm (i.e. assign all pixels to \( k \) clusters according to a distance criterion).

Step 2: given a particular neighborhood size (e.g. 3×3), calculate the values of *PNI* \( w_k \) and \( \delta_k \) according to Equations [1] and [2].

Step 3: if the number of non-overlapping pure neighborhoods for class \( k \) (\( \delta_k \)) is higher than a pre-defined value, a neighborhood-based k-means clustering for class \( k \) is performed; otherwise a pixel-based k-means clustering is conducted.

Step 4: calculate the objective function of k-means. If the change of the objective function between two adjacent iterations is smaller than a pre-set minimal threshold, then the iteration stops. Otherwise, adjust the assignments of pixels to clusters, and go to Step 2.

This NC-k-means approach is an integration of pixel-level and neighborhood-level clustering algorithms. For an image with heterogeneous land use land covers, pixel-wise k-means clustering dominates the algorithm. On the contrary, for an image with homogenous land use land covers, neighborhood-level k-means clustering dominates. For details of this algorithm, readers can refer to Zhang et al. [2013].

**Results and conclusions**

This neighborhood-constrained k-means (NC-k-means) approach was applied to the Chinese Pushbroom Hyperspectral Imager (PHI) image obtained in Minamimaki, Japan. The image is with eighty spectral bands in visible and near infrared spectra and has a spatial resolution of 3 m. Six land use land cover types, plastic film, Chinese cabbage, Japanese cabbage,
bare land, forest, and grass, exist in the image. The results of the NC-k-means method with 2×2, 3×3, and 4×4 neighborhood windows, and a comparative analysis with the traditional k-means approach are reported in Figure 1 and Table 3. It indicates that the NC-k-means with 4×4 neighborhood window significantly improved the classification accuracy, as the kappa coefficient increased from 0.51 to 0.87, and the classification accuracy of each individual land use land cover type also increases significantly, especially for the classes of grass and Chinese cabbage. For the classes of Chinese cabbage and Japanese cabbage, in particular, they could not be separated with the traditional k-means approach, but when textural information was introduced, they can be effectively distinguished, especially with the NC-k-means with the 4×4 neighborhood window (see Tab. 3).

Table 3 - Classification accuracy comparison of k-means and NC-k-means with the real hyperspectral imagery, numbers in bold indicate the highest classification accuracy of product and Kappa coefficient in the four methods. (reprinted from Zhang et al. [2013]).

<table>
<thead>
<tr>
<th>Method</th>
<th>Classification Accuracy (%)</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plastic film</td>
<td>Forest</td>
</tr>
<tr>
<td>k-means</td>
<td>99.1</td>
<td>70.1</td>
</tr>
<tr>
<td>NC-k-means (2×2)</td>
<td>98.1</td>
<td>85.4</td>
</tr>
<tr>
<td>NC-k-means (3×3)</td>
<td>99.1</td>
<td>94.2</td>
</tr>
<tr>
<td>NC-k-means (4×4)</td>
<td>97.4</td>
<td>78.8</td>
</tr>
</tbody>
</table>

**Spatially adaptive spectral mixture analysis (SASMA)**

Another case study is to incorporate geographic information analysis techniques into a spectral mixture analysis model, and develop the spatially adaptive spectral mixture analysis (SASMA) approach. A major issue of spectral mixture analysis is the selection of representative pure land use land cover type, also termed endmember classes. In this case study, endmember class candidates were automatically chosen using a classification tree approach and the classification rules were established by incorporating spectral and spatial information. Furthermore, an inverse-distance-weighting (IDW) technique was applied to derive synthetic spectra as the most “representative” endmembers. Finally, these synthetic spectra were input to a linear SMA to derive areal fractions of land use land covers. For a better explanation, the following sections provide the methods of the SASMA, including the approaches of deriving localized endmember class candidates, generating synthetic endmembers using IDW, and estimating urban impervious surface fractions using the linear SMA model, as well as the results and conclusions.
Spatially adaptive spectral mixture analysis (SASMA) method

With linear SMA methods, a pixel is assumed to contain several land use land cover types, and the spectra of this mixed pixel can be modeled as a linear combination of representative homogeneous land use land covers (e.g. endmembers) weighted by their respective fractions. In SMA, how to extract “representative” endmembers has become a critical issue. This case study incorporated geographic information analysis techniques, namely spatial autocorrelation, into the process of endmember selection. Specifically, this study extracted local endmember candidates using a classification tree model, then derived synthetic endmembers using an IDW approach. Two variables, normalized difference vegetation index (NDVI) and biophysical composition index (BCI) [Deng and Wu, 2012] were input to the classification tree model to extract endmember
candidates of vegetation, high albedo, low albedo, and soil (see Figs. 2B, 2D, 2F, and 2H). With these endmember candidates (Figs. 2C, 2E, 2G, and 2I), an IDW approach was applied to derive the synthetic endmembers of the target pixel accordingly (see Fig. 2J). With the derived local and synthetic endmembers, an SMA model was then applied using the least-square solution [see Eq. 3].

\[
R_b = \sum_{i=1}^{n} f_i R_{i,b} + e_b \\
\text{subject to} \\
\sum_{i=1}^{n} f_i = 1 \text{ and } f_i \geq 0
\]

where \(R_b\) is the reflectance spectra of a pixel in band \(b\); \(R_{i,b}\) is the synthetic endmember of band \(b\) of endmember \(i\); \(f_i\) represents the fraction of endmember \(i\); and \(n\) is the number of endmembers.

**Results and conclusions**

The developed SASMA was applied to Landsat ETM+ imagery acquired in Franklin County, Ohio, United States on September 10, 1999, and the results showed that the performance of this SASMA model was satisfactory (see Fig. 3). In particular, the distributions of urban impervious surfaces generally follow the land use land cover patterns, with relatively high fractions in the central business district (CBD) area (i.e. located in the center of the image), slightly lower fractions in the residential areas (i.e. located around the CBD), and very low fractions in the rural areas (i.e. near to the boundary). In addition, the estimation error, assessed with 200 randomly selected samples, is relatively small, with a root mean square error (RMSE) of 15.25%, mean absolute error of 8.50%, and systematic error (SE) of -0.93%. These results indicate that the performance of SASMA is significantly better than the traditional SMA method reported in Wu [2004] (e.g. RMSE of 18.3% and SE of -10.8%). Therefore, this case study illustrates that the integration of geographic information analysis techniques into the linear SMA model significantly improves the sub-pixel estimation of urban impervious surfaces.

**Discussion and conclusions**

Remote sensing image classification techniques are essential in deriving land use land cover information for socio-economic planning and environmental applications. Currently, spectral classifiers are still the dominant approaches for classifying remote sensing imagery due to their conceptual simplicity and easy implementation. Recently, an increasing number of researchers have realized the importance of spatio-contextual information in complementing spectral classifiers. Through conducting a comprehensive literature review on remote sensing classification methods, especially the spatio-contextual classification techniques, we have obtained several conclusions.

With the availability of very high resolution remote sensing imagery (i.e. IKONOS and QuickBird), the issues associated with traditionally spectra-based classification techniques have been recognized by many remote sensing scientists. With higher spatial resolutions, images are likely to have higher within-class spectral variability. As a result, less than
satisfactory results have been reached with spectral classifiers. Spatio-contextual classifiers, therefore, have potential to address such problems through employing information extracted from the spatial domain. A number of studies have reported significantly higher classification accuracy with image segmentation and object-based image classification [Blaschke, 2010]. In addition, satisfactory results have been reported with texture extraction and MRFs modeling [Fauvel et al., 2013; Moser et al., 2013]. Recently, more and more remote sensing scientists have recognized the importance of spatial information, and a large number of studies have emphasized on developing spatio-contextual image classification methods.

Figure 2 - An illustration of the automated dynamic endmember extraction of SASMA within a search window: (A) the central target pixel (labeled as a green pixel) and its neighboring pixels; the second column is for figures of extracted endmember candidate pixels of (B) vegetation, (D) high albedo, (F) low albedo and (H) soil; the third column is for weight figures of (C) vegetation, (E) high albedo, (G) low albedo, (I) soil; and (J) the resulting endmember spectral signatures for the target pixel. (reprinted from Deng and Wu, [2013]).
Within the developed spatio-contextual classification techniques, texture extraction is the most popular approach, and many textural metrics developed in 1970s [Haralick et al., 1973] are still dominant in the remote sensing literature. In comparison, image segmentation and object-based classification have recently emerged as a major spatio-contextual classification technique, majorly due to the availability of eCognition software [Pal and Bhandari, 1992; Blaschke, 2010]. Comparably, although MRFs models are the most mathematically rigorous approach, they are still at the stage of algorithm developments, and have not been widely accepted by the remote sensing community. This may be majorly due to its theoretical and computational complexity. Fauvel et al. [2013] argued that a uniform framework of MRFs models for remote sensing image classification is highly necessary, such that these algorithms can be implemented by remote sensing scientists who may not have deep understandings of image analysis techniques.

Besides these three groups of techniques, geographic information analysis techniques could make important contributions to spatio-contextual image classification. These techniques are more or less ignored majorly due to the gaps among different research communities, as the three other major approaches are originated in the areas of computer vision and pattern recognition, and geographic information analysis techniques were developed by geographers, geologists, urban planners, and economists [Anselin, 1995; Fotheringham et al., 2003]. Geographic information analysis techniques, however, also intend to address the spatial dependence problem, and have been widely applied to vector data analysis [Wu, 2012]. This paper argued the effectiveness of geographic information analysis techniques in pixel-wise and sub-pixel-wise classifications. These techniques may have significant potential to be applied in spatio-contextual remote image classifications.

**Acknowledgements**

This study was financially supported by the National Natural Science Foundation of China (n. 41030743; n. 41171322; n. 41171412), University of Wisconsin-Milwaukee Graduate School, as well as Open Research Fund of Key Laboratory of Digital Earth, Center for Earth Observation and Digital Earth, Chinese Academy of Sciences. The authors would like to thank anonymous reviewers for their constructive comments on earlier drafts of this manuscript.
References


Ceccarelli T., Smiraglia D., Bajocco S., Rinaldo S., De Angelis A., Salvati L., Perini L.,


Shekarforoush H., Berthod M., Zerubia J., Werman M. (1996) - Sub-pixel Bayesian


