A neighbourhood-constrained $k$-means approach to classify very high spatial resolution hyperspectral imagery

BING ZHANG*†, SHANSHAN LI†‡, CHANGSHAN WU‡, LIANRU GAO†, WENJUAN ZHANG† and MAN PENG§

†Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
‡Department of Geography, University of Wisconsin–Milwaukee, Milwaukee, WI, 53201, USA
§Institute of Remote sensing Applications, Chinese Academy of Sciences, Beijing 100101, China

(Received 24 April 2012; in final form 12 July 2012)

In classifying very high spatial resolution (VHR) hyperspectral imagery, intra-class variation often adversely affects classification accuracy, mainly due to a low signal-to-noise ratio (SNR) and high spatial heterogeneity. To address this problem, this article develops a neighbourhood-constrained $k$-means (NC-$k$-means) algorithm by incorporating the pure neighbourhood index into the traditional $k$-means algorithm. The performance of the NC-$k$-means algorithm was assessed through a series of simulated images and a real hyperspectral image. The results indicate that the classification accuracy of NC-$k$-means algorithm is consistently better than that of the traditional $k$-means algorithm, in particular for the images with significant spatial autocorrelations among neighbouring pixels.

1. Introduction

With the advance of sensor technologies, hyperspectral remote-sensing images with a very high spatial resolution (VHR) have shown significant promise in the applications of agriculture, forestry and urban analyses (Weng et al. 2008, Souci et al. 2009). While VHR hyperspectral imagery provides detailed information both spectrally and spatially, a higher spatial resolution may lead to lower signal-to-noise ratio (SNR) and/or significantly higher intra-class variation (Kramer 1996, Cocks et al. 1998). Therefore, when the traditional spectral-based classification techniques, such as spectral angle mapping (SAM), $k$-means clustering (Bandyopadhyay and Maulik 2002, Wulder et al. 2004) and maximum likelihood classification, are applied to such VHR hyperspectral images, the resultant classification accuracy is usually poor, in particular in regions with complex landscapes (Tadjudin and Landgrebe 1998).

To address this problem, scientists have attempted to incorporate spatial contextual information into classification algorithms (Richards and Jia 2006, Plaza et al. 2009). For unsupervised classification, Luo et al. (2003) proposed a spatially constrained $k$-means (NC-$k$-means) approach for image segmentation. Further, Mignotte (2011) integrated a de-texturing procedure and the spatially constrained $k$-means

*Corresponding author. Email: zb@ceode.ac.cn
approach for better image segmentation. These methods, however, do not handle high-dimension, remote-sensing imagery well. Therefore, it is necessary to develop an algorithm for classifying VHR hyperspectral imagery by incorporating spatial contextual information into the classification algorithms (Jimenez et al. 2005).

In this article, we present a new approach to incorporate spatial contextual information into the \( k \)-means algorithm. We first investigate the spectral and spatial contextual features of pixel groups in a neighbourhood and propose the pure neighbourhood index (PNI). Then, a novel NC-\( k \)-means algorithm is developed by incorporating the PNI into the traditional \( k \)-means clustering method. The performance of this new NC-\( k \)-means algorithm is assessed by comparing its results to the traditional \( k \)-means algorithm applied to both simulated data and real hyperspectral imagery.

2. NC-\( k \)-means classification scheme

2.1 Neighbourhood properties

In this study, a neighbourhood is defined as a set of contiguous pixels in a certain window (e.g. \( 2 \times 2 \) pixels, \( 3 \times 3 \) pixels, etc.) of an image. As an example, figure 1(a) shows an area in a hyperspectral image, in which each small square represents a single pixel, denoted as \( x_1, x_2, x_3 \) and \( x_4 \). The size of a neighbourhood is defined as \( d \) pixels \( \times \) \( d \) pixels (e.g. a neighbourhood with a size of \( 2 \times 2 \) pixels is shown in figure 1). If all the pixels within a neighbourhood belong to the same class, then this neighbourhood is defined as a pure neighbourhood, otherwise it is considered as a mixed neighbourhood.

Let \( X = \{x_1, x_2, \ldots, x_n\} \) be a set of spectral vectors for all pixels in an image. Each \( x_i \) represents the spectral vector of a particular pixel \( i \) (\( 1 \leq i \leq n \)), and can be expressed as \( x_i = \{x_{ij}\} \), \( j = 1, 2, \ldots, m \), where \( x_{ij} \) represents the value (either reflectance or digital number) of pixel \( i \) at spectral band \( j \), and \( m \) is the number of spectral bands. According to the linear mixed model (Keshava and Mustard 2002), the spectrum of this neighbourhood can be calculated as the average spectra of the pixels within the neighbourhood:

\[
N_G = \frac{1}{L} \sum_{i=1}^{L} x_i
\]

where \( N_G \) represents the spectral vector of neighbourhood \( G \) in the image, \( L \) is the number of pixels in neighbourhood \( G \) and \( L = d^2 \).

Figure 1. (a) An example of \( 2 \times 2 \) neighbourhood. (b) Pixel 1 in neighbourhood \( G \) with all neighbouring pixels labelled \( \omega_I \). (c) Pixel 2 in neighbourhood \( Q \) with neighbouring pixels labelled \( \omega_{II}, \omega_{III} \) and \( \omega_{IV} \), respectively.
Further, let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$ be a set of $K$ classes, $K$ represents number of classes ($1 \leq k \leq K$) and $\omega_k$ denotes a particular pixel belongs to class $k$. Figures 1(b) and (c) illustrate a neighbourhood with $2 \times 2$ pixels. In this example, pixels 1 and 2 are assumed to have the same spectral information ($x_1 = x_2$), and they are located in neighbourhoods $G$ and $Q$, respectively. Within the neighbourhood $G$, all pixels around pixel 1 are labelled as $\omega_1$, whereas within the neighbourhood $Q$, pixels around pixel 2 are labelled as $\omega_{11}$, $\omega_{111}$ and $\omega_{1111}$. If only spectral information is considered, pixels 1 and 2 have an equal probability of being classified as class I, e.g. $P(\omega_1|x_1) = P(\omega_1|x_2)$. If the spatial organizations of the neighbourhoods $G$ and $Q$ are taken into consideration, however, the probability of pixel 1 classified as class I should be higher. Therefore, in order to examine the influence of the neighbourhood on the classification of a particular pixel, it is necessary to develop an index to quantify the homogeneity of a neighbourhood.

2.2 Pure Neighbourhood Index

In order to examine the effect of neighbourhood properties on the probability of a pixel assigned to a particular class, we propose an index to represent the homogeneity of the neighbourhood. The probability of a pixel being assigned to class $k$ should be associated with its neighbourhood properties, and the probability of a pixel belonging to class $k$ in a pure neighbourhood of class $k$ should be higher than that in a mixed neighbourhood. Let $w$ be a neighbourhood and $1 \leq w \leq W = \lfloor n/L \rfloor$, where $\lfloor n/L \rfloor$ represents the largest integer smaller than or equal to $n/L$. Therefore, we propose the PNI to represent the posterior probability of classifying a pixel into class $k$. It is a binary function denoted by $\delta_{kw}$, which describes the weighting coefficient of neighbourhood $w$ belonging to class $k$:

$$
\begin{cases}
\delta_{kw} = 1 & \text{if } w \text{ is a pure neighborhood of class } k \\
\delta_{kw} = 0 & \text{if } w \text{ is a mixed neighborhood of class } k
\end{cases}
$$ (2)

So, we can obtain

$$
P(\omega_k|x_i, i \in w) = \delta_{kw}P(\omega_k|x_i)
$$ (3)

2.3 Incorporating PNI into $k$-means clustering algorithm

In order to incorporate the spatial property in the $k$-means clustering, we develop an NC-$k$-means algorithm by incorporating the PNI into the iteration process. NC-$k$-means is a fast, unsupervised, classification algorithm using coupled parallel iterations: pixel-level iteration and neighbourhood-level iteration. Details of this algorithm are described as follows:

Step 1: Classify the images into $k$ clusters and initialize a set of centroids $z_k$ ($1 \leq k \leq K$). In this step, a typical method is to randomly select $k$ vectors from all pixels as the initial centroids of the clusters. The initial optimization algorithm for clustering can be found in Maitra (2009).

Step 2: Compute $n \times k$ partition matrix $U$ to classify all pixels to initial cluster $k$ according to equation (4):
\[
\begin{cases}
    u_{ki} = 1, & \text{if } D(x_i, z_k) \leq D(x_i, z_t) \text{ for } 1 \leq t \leq K \\
    u_{ki} = 0, & \text{otherwise}
\end{cases}
\]

(4)

where \( u_{ki} \) represents whether pixel \( i \) is assigned to cluster \( k \), and \( D(x_i, z_k) = (x_i - z_k)^T(x_i - z_k) \) is the squared Euclidean distance from \( x_i \) to \( z_k \).

Step 3: Set the neighbourhood window scale \( L \), calculate the spectral vector of each neighbourhood \( w \) by applying equation (1) and determine the value of the neighbourhood label according to equation (5):

\[
\begin{cases}
    u'_{kw} = 1, & \text{if } D(N_w, z_k) \leq D(N_w, z_t) \text{ for } 1 \leq t \leq K \\
    u'_{kw} = 0, & \text{otherwise}
\end{cases}
\]

(5)

where \( u'_{kw} \) indicates whether neighbourhood \( w \) is assigned to cluster \( k \), and \( D(N_w, z_k) \) represents the squared Euclidean distance from \( N_w \) to \( z_k \).

Step 4: Determine the PNI (\( \delta_{kw} \)) for all neighbourhoods. Calculate the sum of the PNIs of a particular cluster \( k \) in the image (\( \delta_k = \sum_{w=1}^{W} \delta_{kw} \)), and the sum of PNIs of all neighbourhoods with all clusters is \( \delta = \sum_{k=1}^{K} \delta_k = \sum_{k=1}^{K} \sum_{w=1}^{W} \delta_{kw} \). Finally, determine the total number of pixels in pure neighbourhoods as \( \delta \times L \).

Step 5: Calculate the \( k \)-means objective function

\[
J(U, Z) = \sum_{k=1}^{K} \sum_{i=1}^{n} u_{ki} D(x_i, z_k)
\]

(6)

This algorithm is iterative, and if the difference of objective function \( J \) in adjacent iterations is smaller than a preset minimal threshold, then the iteration stops, and the optimal result is reached, otherwise go to step 6 for clustering centroid adjustment.

Step 6: Adjust the centroid of each cluster \( k \) (\( z_k \)), cluster type of each pixel \( i \) (\( u_{ki} \)) and cluster type of each neighbourhood \( w \) (\( u'_{kw} \)), then go to step 4. Step 6 can be divided into three sub-steps:

Step 6.1: Adjust the set of centroid for each cluster \( k \) with neighbourhood-level and pixel-level adjustments (\( T_k \) is a pre-defined criterion)

\[
\begin{align*}
    z_k &= \frac{\sum_{w=1}^{W} \delta_{kw} u'_{kw} N_w}{\sum_{w=1}^{W} \delta_{kw} u'_{kw}} & \text{for } 1 \leq k \leq K, \quad \delta_k \geq T_k \\
    z_k &= \frac{\sum_{i=1}^{n} u_{ki} x_i}{\sum_{i=1}^{n} u_{ki}} & \text{for } 1 \leq k \leq K, \quad \delta_k < T_k
\end{align*}
\]

(7)

Step 6.2: Adjust \( u_{ki} \) of each pixel according to equation (4).

Step 6.3: Adjust \( u'_{kw} \) of each neighbourhood according to equation (5).

In steps 5 and 6, if all pixels with the same cluster are separated by pixels of other classes, the PNI of each neighbourhood in the image equals to zero. In such a case,
all the neighbourhoods are mixed \((\delta = 0)\), and this NC-\(k\)-means clustering method will be degraded to the \(k\)-means algorithm. On the contrary, if the pixels within a hyperspectral image are spatially correlated, there are a larger number of pure neighbourhoods, and the incorporation of PNI may potentially improve the classification accuracy.

3. Experimental results and discussions

3.1 Simulated data experiments

With simulated experimental data, we attempted to address two issues: (1) comparing the proposed NC-\(k\)-means with the standard \(k\)-means algorithm and evaluating whether the NC-\(k\)-means algorithm produces better results under different scenarios and (2) determining the parameter thresholds \(T_k\) and \(\delta_k\) based on a series of experiments. In particular, experimental data were derived from the spectra of seven classes of land cover, including three vegetation types (pasture, yam and lettuce), forest, plastic, dry soil and wet soil. For each land cover, totally 64 bands with the wavelength ranging from 411 to 850 nm were selected from the Operational Modular Imaging Spectrometer (OMIS) hyperspectral image. OMIS is a hyperspectral sensor that collects data with a spatial resolution of 3.0 m. For each land-cover type, 400 samples were taken, and a total of 2800 samples were employed for all seven classes.

The simulated hyperspectral image is shown in figure 2(a), in which the specified seven land-cover types are arranged in seven adjacent squares (20 lines \(\times\) 20 columns). Therefore, the entire image has 140 lines and 20 columns. Land-cover types from top to bottom are plastic, forest, pasture, yam, lettuce, dry soil and wet soil. Figure 2(b) shows the reference image of the seven land-cover types.

The simulated image was classified using the \(k\)-means clustering and NC-\(k\)-means algorithms, in which the same initial centroids were employed, and the number of clusters was set to be 7. Figures 2(c) and (d) show the classification results of \(k\)-means and NC-\(k\)-means, respectively. Figures 2(e), (f) and (g) illustrate pure neighbourhood maps at the initial, middle and final stages of the iterations with the NC-\(k\)-means. The total number of pure neighbourhoods \((\delta)\) is the lowest at the initial stage of clustering,
and during the iterations, the number of pure neighbourhoods increases (the value of $\delta$ increases). Finally, the iteration stops when the PNIs of all neighbourhoods keep unchanged.

As shown in figure 2(c), significant variation in forest and vegetation classes is observed when the $k$-means clustering approach was applied. Moreover, compared to the reference images, salt and pepper noise is common. However, in the clustering results of NC-$k$-means (see figure 2(d)), much less noise was found in the results of forest and vegetation classifications. As shown in table 1, the overall classification accuracy of NC-$k$-means (95.4%) is much higher than that of the traditional $k$-means method (88.5%). For individual land-cover type, the classification accuracies of forest, yam and dry soil were much higher with the NC-$k$-means algorithm. For other land-cover types, such as plastic, lettuce and wet soil, NC-$k$-means has similar classification accuracy compared to $k$-means.

In order to determine the parameter thresholds $T_k$ and $\delta_k$ in the NC-$k$-means algorithm, we designed 10 scenarios with different spatial arrangements using the seven land-cover types with 400 samples each. Figure 3 shows the false colour composite simulated images with different arrangements and reference images of the corresponding classes. The 10 scenarios of arrangements are sorted according to the

<table>
<thead>
<tr>
<th>Methods</th>
<th>Plastic</th>
<th>Forest</th>
<th>Pasture</th>
<th>Yam</th>
<th>Lettuce</th>
<th>Dry soil</th>
<th>Wet soil</th>
<th>Overall accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$-means</td>
<td>100.0</td>
<td>68.3</td>
<td>67.8</td>
<td>86.5</td>
<td>100.0</td>
<td>97.3</td>
<td>100.0</td>
<td>88.5</td>
</tr>
<tr>
<td>NC-$k$-means</td>
<td>100.0</td>
<td>100.0</td>
<td>67.8</td>
<td>100.0</td>
<td>100.0</td>
<td>99.8</td>
<td>100.0</td>
<td>95.4</td>
</tr>
</tbody>
</table>

Note: Bold values indicate highest classification accuracies.

![Figure 3](https://example.com/f3.jpg)

Figure 3. Simulated hyperspectral data with different number of contiguous pixels: (a) 100, (b) 50, (c) 35, (d) 30, (e) 28, (f) 26, (g) 24, (h) 22, (i) 15 and (j) 1.
A neighbourhood-constrained k-means approach

complexity of spatial structures. The corresponding sequences are named as (a), (b), (c), (d), (e), (f), (g), (h), (i) and (j), as shown in figure 3.

To ensure the comparability of images among all arrangements, there are still 400 pixels of each land-cover type, and each image has 140 rows and 20 columns. In each scenario, the pixels with the same land-cover type are arranged contiguously, and then alternately arranged with pixels belonging to a different class. From scenarios (a) to (j), the spatial complexity gradually increases, and the number of contiguously arranged pixels decreases, and they change progressively from 100 (a) to 1 (j). With the decrease of contiguous pixels, the number of pure neighbourhoods $\delta$ decreases accordingly. With the simulated scenarios with different spatial contiguity, we are able to examine the relationship between threshold values and classification accuracy:

$$\sigma = \frac{L \times \delta}{n}$$

where $\sigma$ represents the ratio between the total number of pixels in pure neighbourhoods and the number of the pixels in the image. Figure 4 shows the relationship between the ratio $\sigma_{2 \times 2}$ ($2 \times 2$ neighbourhood) and overall classification accuracy.

Based on the curve-fitting relationship shown in figure 4, if $\sigma_{2 \times 2}$ is around 40%, NC-k-means has higher classification accuracy than the k-means method. Considering various affecting factors in real hyperspectral images, $\sigma_{2 \times 2}$ is greater than or equal to 0.4, i.e. $\delta_{2 \times 2} \geq \frac{0.4n}{4}$. In this experiment, the number of samples of various classes is the same, and the initial numbers of pixels belonging to cluster $k$ in the neighbourhood have to satisfy the formula $\delta_k(2 \times 2) \geq \frac{0.4n}{4K}$, $k = 1, \ldots, K$ for each cluster. Therefore, the optimal threshold for $2 \times 2$ neighbourhood window size is $T_{k(2 \times 2)} = \frac{0.4n}{4K}$. Similarly, we estimated the optimal thresholds for $3 \times 3$ and $4 \times 4$ neighbourhood window sizes as $T_{k(3 \times 3)} = \frac{0.46n}{9 \times K}$ and $T_{k(4 \times 4)} = \frac{0.50n}{16 \times K}$, respectively.

![Figure 4. Relationship between $\sigma_{2 \times 2}$ and overall classification accuracy. The horizontal line represents the values of $\sigma_{2 \times 2}$, and the vertical line indicates overall accuracy corresponding to experiments (a)--(j) with simulated hyperspectral data.](image-url)
3.2 Chinese Pushbroom Hyperspectral Imager data experiments

An airborne hyperspectral remote-sensing image acquired by the Chinese Pushbroom Hyperspectral Imager (PHI) was employed in this experiment (see figure 5). The image has 80 spectral bands covering the spectral region from 400 to 850 nm with 3 m spatial resolution (Hu et al. 2005). The study area has 200 lines and 200 columns, covering a typical agricultural site in Minamimaki, Japan. Six land-cover types, including plastic film, forest, grass, Japanese cabbage, Chinese cabbage and bare soil, are presented in the image (see figures 5(a) and (b)). With this PHI hyperspectral image, a $k$-means clustering algorithm was applied with six clusters. Results (see figure 5(c)) indicate that one cluster (with yellow colour) identified by the $k$-means method represents salt and pepper noise. Moreover, Chinese cabbage and Japanese cabbage were classified into a single class, as they share similar spectral information. In addition to

Figure 5. Experiments with real hyperspectral image. The centre latitude/longitude of image is 36.02°N, 138.49°E. (a) False colour composite image ($R = 832$ nm, $G = 650$ nm and $B = 553$ nm), (b) reference data, (c) classification result with $k$-means clustering, (d) classification result with the NC-$k$-means with $2 \times 2$ neighbourhood window, (e) classification result with the NC-$k$-means with $3 \times 3$ neighbourhood window and (f) classification result with the NC-$k$-means with $4 \times 4$ neighbourhood window.
neighbourhood-constrained \( k \)-means approach

<table>
<thead>
<tr>
<th>Method</th>
<th>Plastic film</th>
<th>Forest</th>
<th>Grass</th>
<th>Chinese cabbage</th>
<th>Japanese cabbage</th>
<th>Bare soil</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )-means</td>
<td>99.1</td>
<td>70.1</td>
<td>14.7</td>
<td>0.0</td>
<td>92.5</td>
<td>95.5</td>
<td>0.51</td>
</tr>
<tr>
<td>NC-( k )-means (2 × 2)</td>
<td>98.1</td>
<td>85.4</td>
<td>65.6</td>
<td>84.6</td>
<td>46.9</td>
<td>94.2</td>
<td>0.70</td>
</tr>
<tr>
<td>NC-( k )-means (3 × 3)</td>
<td>99.1</td>
<td>94.2</td>
<td>72.2</td>
<td>85.3</td>
<td>57.8</td>
<td>94.2</td>
<td>0.75</td>
</tr>
<tr>
<td>NC-( k )-means (4 × 4)</td>
<td>97.4</td>
<td>78.8</td>
<td>73.8</td>
<td>95.4</td>
<td>98.3</td>
<td>94.1</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Note: Bold values indicate highest classification accuracies and kappa coefficient.

4. Conclusions

To address the issues of low SNR and high spatial heterogeneity associated with VHR hyperspectral imagery, in this article, we developed an NC-\( k \)-means algorithm by incorporating the PNI into the \( k \)-means algorithm. With the NC-\( k \)-means algorithm, centroids of clustering were optimized by reducing the interference of noise and the heterogeneity caused by high spectral and spatial variation. The performance of NC-\( k \)-means is better compared to that of the \( k \)-means algorithm. Further analysis shows that the classification accuracy of the NC-\( k \)-means varies with different neighbourhood window size. Results indicate that a larger neighbourhood window can improve the computing efficiency because of the reduced number of neighbourhoods. Especially, for areas with low spatial variation (e.g. agricultural fields), a larger neighbourhood window size is particularly suitable.
Acknowledgements

The authors are grateful to Prof. Tim Warner from West Virginia University and two anonymous referees for their constructive comments. This work was supported by Director Foundation of Center for Earth Observation and Digital Earth, Chinese Academy of Sciences and National Key Technology R&D Program of China under Grant No. 2011BAH12B04.

References


TADJUDIN, S. and LANDGEBE, D., 1998, Classification of high dimensional data with limited training samples. PhD thesis, Purdue University, USA.
