Remote Sensing Image Classification Using the Spectral-Spatial Distance Based on Information Content

Siya Chen 1, Hongyan Zhang 1,*, Tieli Sun 1,2,*, Jianjun Zhao 1 and Xiaoyi Guo 1

1 School of Geographical Science & School of Information Science and Technology, Northeast Normal University, Changchun 130024, China; chensy413@nenu.edu.cn (S.C.); zhaojj662@nenu.edu.cn (J.Z.); guoxy914@nenu.edu.cn (X.G.)

2 Key Laboratory of Intelligent Information Processing of Jilin Universities, Northeast Normal University, Changchun 130024, China

* Correspondence: zhy@nenu.edu.cn (H.Z.); suntl@nenu.edu.cn (T.S.); Tel.: +86-431-8509-9147 (H.Z.); +86-431-8453-6353 (T.S.)

Received: 12 August 2018; Accepted: 10 October 2018; Published: 12 October 2018

Abstract: Among many types of efforts to improve the accuracy of remote sensing image classification, using spatial information is an effective strategy. The classification method integrates spatial information into spectral information, which is called the spectral-spatial classification approach, has better performance than traditional classification methods. Construct spectral-spatial distance used for classification is a common method to combine the spatial and spectral information. In order to improve the performance of spectral-spatial classification based on spectral-spatial distance, we introduce the information content (IC) in which two pixels are shared to measure spatial relation between them and propose a novel spectral-spatial distance measure method. The IC of two pixels shared was computed from the hierarchical tree constructed by the statistical region merging (SRM) segmentation. The distance we proposed was applied in two distance-based contextual classifiers, the k-nearest neighbors-statistical region merging (k-NN-SRM) and optimum-path forest-statistical region merging (OPF-SRM), to obtain two new contextual classifiers, the k-NN-SRM-IC and OPF-SRM-IC. The classifiers with the novel distance were implemented in four land cover images. The classification results of the classifier based on our spectral-spatial distance outperformed all the other competitive contextual classifiers, which demonstrated the validity of the proposed distance measure method.

Keywords: remote sensing image classification; statistical region merging; information content; contextual classifier

1. Introduction

Classification is an important technology to extract useful information from remote sensing images, so improving the accuracy of classification is always a popular research topic. The newest improvement direction focuses on exploiting spectral-spatial features to classify remote sensing images [1], namely spectral-spatial classification. Spectral-spatial classification approaches can also be called contextual classification approaches. Spectral-spatial classifiers are more accurate and work better compared with the traditional classifiers because they take into account the contextual spatial characteristics to enhance the performance of the classification results [2]. The techniques commonly used in contextual classifiers to extract spatial features include the random field (RF) [3], mathematical morphology [4], segmentation and clustering [5], and deep learning [6].
The random field (RF) is a useful approach to capture the prior knowledge of an image. It models the classification task as a RF model to obtain the spatial correlation between pixels and realize the spectral-spatial classification. The Markov random field (MRF) was first introduced into spectral-spatial classification and significantly improved the traditional classifiers. Numerous works can be found that combine MRF with various traditional classifiers, including Bayesian Network [7], the SVM [8,9] and the optimum-path forest (OPF) [10–12]. Later, the conditional random field (CRF) model as an improved MRF model was successfully applied in spectral-spatial classification [13]. CRF models the posterior distribution as more flexible in classification. Different from the MRF considers spatial features at the label level, the CRF incorporates the spatial information on both the pixel data level and the class label level [14]. With the appearance of new, powerful optimization algorithms that can solve CRF inference problems, the application of CRF is more extensive [15].

Classifying pixels that belong to one region as homogeneous is another typical way to consider spatial information. The region can be obtained by implementing segmentation and clustering on an image. One strategy is to perform segmentation (e.g., watershed (WH), the squared-error clustering method (ISODATA) and expectation maximization (EM)) and pixel-wise classification independently, and the final classification result is determined by the majority voting (MV) rule within the segmented region [16,17]. Another way is to segment the image using fractional-order Darwinian particle swarm optimization (FODPSO) and mean shift segmentation (MSS) first, and then the segmented image is used as the input of the pixel wise classification [18]. The third type is to construct the spectral-spatial similarity or spectral-spatial distance measure, such as Huo et al., who computed both the spatial and spectral similarity by hierarchical segmentation (HSEG) to construct the spatial kernel for the SVM [19].

The original mathematical morphology-based contextual classifier exploits morphological profiles (MPs) to represent the spectral-spatial information. The MP is built from the first principal components of the image and constructs an extended morphological profile (EMP) used for classification [20]. However, there are still some limitations in MPs that were overcome by attribute profiles (APs) [21]. APs are more flexible and efficient than MPs but suffer from the threshold values. The extended attribute profile (EAP) should be built for classification [22]. The extended multi attribute profile (EMAP) is obtained when several constructed EAPs considering different attributes are stacked together [23]. Recently, an advanced approach was proposed, named the extinction profile (EP), which is fully automatic and independent of the dataset’s distribution [24].

The capability of deep learning to extract joint spatial-spectral features independently has been proven in recent years and become a popular way to improve classification results [25]. Lee et al. [26] and Yang et al. [27] proposed deep convolutional neural network (CNN)-based contextual remote sensing image classification. By considering the size of the spatial window in the CNN as limited to 3*3, Chen et al. proposed a novel method that uses a larger spatial window to integrate features in the CNN [28]. Erchan et al. explored the collaboration of two powerful methods, APs and the CNN, to achieve a more effective strategy for classification [29].

In addition to the above methods, jointly taking into account the spatial and spectral information on dimensionality reduction of hyperspectral images (HSI) is also an effective way to achieve spectral-spatial classification. Yuan et al. [30] propose a spectral-spatial linear discriminant analysis (SSLDA) and its kernel version spectral-spatial kernel discriminant analysis (SSKDA). Sellami and Farah [31] incorporated the tensor locality preserving projections method (TLPP) and constrained band selection method (CBS) to preserve as much as possible the spectro-spatial information during spectro-spatial dimensionality reduction in HSI.

In this paper, we extend two segmentation-based spectral-spatial classifiers in [32] by propose a novel spectral-spatial distance measure method. Chen et al. proposed a spectral-spatial distance using statistical region merging (SRM) segmentation for two distance-based classifiers, the k-NN and OPF, and obtained two contextual classifiers, the OPF-SRM and k-NN-SRM. However, they only simply used the height information to compute the spectral-spatial distance ignored other useful
information within segmentation results. In our novel spectral-spatial distance measure method, additional IC is also considered. The novel spectral-spatial distance measure method takes full advantage of spatial information in the segmentation results which makes the distance computation more accurate. Exploiting the proposed spectral-spatial distance in classification could reduce the effect of segmentation results on classification to some extent; make up for the shortcoming that the results of spectral-spatial classification method based on segmentation is highly dependent on segmentation. As far as we know, this is the first time to introduction IC in remote sensing image classification. Our improved spectral-spatial distance measure is applied to the OPF-SRM and k-NN-SRM to obtain two extended contextual classifiers, OPF-SRM-IC and k-NN-SRM-IC. The classification results using the improved measure on four real land cover remote sensing images outperformed the original classifiers and other competitive contextual classifiers, which illustrated the validity of the proposed spectral-spatial distance function. This paper is organized in five sections. Section 2 described the proposed spectral-spatial distance measure that considers the detailed IC. Section 3 presented the datasets used for the experiments and the parameters used in the experiments. Section 4 provided a discussion and Section 5 is the conclusion of this paper.

2. Materials and Methods

2.1. Principle of SRM Segmentation

The SRM [33] is a commonly used hierarchical multiscale segmentation method. SRM is also a region-based segmentation method. The main idea of the SRM is to start with a pixel of an image and then determine whether a neighborhood should be added according to a particular order and merging criteria. The regional growth is stopped and image segmentation is completed when a certain condition is satisfied. The key steps in the SRM include the calculation of the merging prediction and merging sequence.

Given $I^*$ is the real scene, the statistical regions within $I^*$ represent the theoretical true objects. Image $I$ denotes the observation of $I^*$. Defined $|·|$ is the cardinality of image, so there are $|I|$ pixels contained in $I$ in total. Each pixel of an image contains red-green-blue (RGB) values, each of the three belonging to the set $\{1, 2, \ldots, g\}$. In 8 bit images, $g = 256$. Then each color channel of the pixels in $I^*$ could be described by a set of $Q$ independent random variables within the range of $[0, g/Q]$, which ensure the sum of the $Q$ independent random variables belongs to $\{1, 2, \ldots, g\}$. The merging criterion is that the pixels inside a region are homogeneous if they have the same expected value for any given color channel, and the pixels that belong to different regions are separate if the expected values are different for at least one color channel. The merging predicate can be written as follows:

$$P(R, R') = \begin{cases} 
\text{true if } \forall a \in \{R, G, B\}, \mid R_a - R'_a \mid \leq \sqrt{b^2(R) + b^2(R')} \\
\text{false otherwise}
\end{cases},$$

where $R$ and $R'$ are two adjacent candidate regions to be merged. $R_a$ is the mean value of region $R$ in channel $a$. $b(R)$ is calculated as follows:

$$b(R) = g \left[ \frac{1}{2Q|R|} \left( \ln \frac{|S|R|}{\delta} \right) \right],$$

where $S|R|$ denotes the set of regions with $|R|$ pixels and $\delta = 1/(6|I|^2)$. The regions $R$ and $R'$ that are homogeneous will be merged if the $P(R, R')$ is true.

A pixel $P$ and its 4 connecting neighbors (up, down, left and right), $P'$, constitute a pixel pair $(P, P')$. There are a total of $N < 2|I|$ pixel pairs. The SRM defines the merging sequence as the ascending order of dissimilarity of these pixel pairs. The pixel pair with the smallest dissimilarity
is the first to be predicated. The dissimilarity is measured by \( (P, P') = \max |P'_\alpha - P_\alpha| \), in which \( P_\alpha \) and \( P'_\alpha \) are the values of pixels \( P \) and \( P' \) in channel \( \alpha \in \{R, G, B\} \), respectively. So we can see, \( Q \) is the only parameter in SRM used to determine the statistical complexity of \( I' \). One can achieve different segmentation granularity by adjusting \( Q \) values.

2.2. Spectral-Spatial Distance Based on SRM

As mentioned above, the merging prediction of the SRM segmentation method is based on the adjacent regions, and thus, the segmentation results of the SRM method reflect the spatial relation among the samples of an image. Moreover, the segmentation results of SRM could be used to measure the spatial distance between samples of an image.

In the segmentation results of SRM, the segmented regions decrease, and the number of regions becomes larger when \( Q \) increases and vice versa. We can obtain different scales of segmentation results of an image by tuning the \( Q \) values and put all these results together can build a hierarchical tree with the hierarchical level (HL) \( h \). Such as in Figure 1, the hierarchical tree is constructed with the segmentation results of four different \( Q \) values for one image. Each level in the hierarchical tree is a different scale of segmentation. With the increasing values of \( Q \), the hierarchical value levels decrease, the number of regions increases, and the size of regions decreases. The highest hierarchical level is \( H \), in which is the initial undivided image. The structure of the hierarchical tree can be used to measure the spatial relation between the pixels. Huo et al. exploited the levels, namely the heights of segments to compute the spatial similarity between pixels. The segment with the minimum level among all common segments of two pixels calls their first common segment. The hierarchical level of the first common segment of pixels \( x_i \) and \( x_j \) is computed as their hierarchical level distance \( d_{HL} \). Then the similarity formula is defined as follows:

\[
\text{Sim}(x_i, x_j) = 0.9^{d_{HL}(x_i, x_j)}, \quad (a > 0),
\]

\[
d_{HL}(x_i, x_j) = \arg\min_{1 \leq h \leq H} \left\{ \text{seg}^h_{x_i} = \text{seg}^h_{x_j} \right\},
\]

where \( \text{seg}^h_{x_i} \) denotes that pixel \( x_i \) belongs to a segment at level \( h \). According to the above formula, the hierarchical level distance \( d_{HL} \) of the green and yellow objects in Figure 1 is 3. The lower the hierarchical level \( h \) of the first common segment of two pixels is, the more spatially similar they are. The spatial similarity is minimal when the first common segment of two pixels is the original undivided image at the highest level \( H \).

The traditional distance function is computed based on the spectral feature space when applied in remote sensing image classification, which ignore the spatial characteristics useful for classification and
lead to a lower classification accuracy. Inspired by Huo, Chen et al. proposed a spectral-spatial distance based on SRM, which considered both the spatial and the spectral characteristics of remote sensing image during classification. They used $1/0.9^{d_H(x_i,x_j)}$ to denote the spatial distance $d_{\text{spatial}}$ between pixels $x_i$ and $x_j$ and combined it with the traditional spectral feature to get a hybrid spectral-spatial distance function as follows:

$$d_{\text{spatial} - \text{spectral}}(x_i,x_j) = f_{\text{spatial}}d_{\text{spectral}}(x_i,x_j),$$  \hspace{1cm} (5)

where $d_{\text{spectral}}$ is the distance between pixels based on the spectral feature space, such as the Euclidean distance that was used.

### 2.3. Novel Spectral-Spatial Distance Based SRM Considering IC

The accuracy of classification based on spectral-spatial distance proposed by Chen et al. is highly dependent on the performance of the SRM segmentation results, so we should make full use of the available information within the segmentation results and the corresponding hierarchical tree. However, Chen et al. only simply used the height information of segments in the hierarchical tree, and there is still a lot of other information of hierarchical tree can used to capture richer spatial characteristics and more accurate spatial relation. In this section, we extend the works of Chen et al. by proposing an improved spectral-spatial distance measure. The proposed distance measure method introduced the IC of two pixels shared.

The notion of IC was widely applied in the concept similarity measure method in data mining. Following the standard argumentation of information theory, the IC of a concept can be quantified by the probability. As the probability increases, the more abstract a concept becomes and the lower its IC. This same argument can apply to segments in a segmented image. The more pixels a segment contains increases its probability and lowers its IC. In the hierarchical tree obtained by SRM, from the low hierarchical level to the high hierarchical level, the size of segments goes from small to large, the segments go from specific to abstract, and the IC in segments goes from large to small. According to Resnik [34], the similarity between concepts depends on the information that two concepts shared, which could be computed by the IC of their common ancestor. Inspired by Resnik, we can obtain a new way to measure pixel distances through the quantitative characterization of information. The distance between the two pixels depends on the information that the two pixels shared. The less information that two pixels share, the more dissimilar they are, and the greater the distance.

The distance measure-based IC between pixels is defined as follows:

$$d_{\text{IC}} = \log(\text{count}_{\text{FCS}}(x_i,x_j))/\log(N),$$  \hspace{1cm} (6)

where $\text{count}_{\text{FCS}}(x_i,x_j)$ is number of pixels in the first common segment of $x_i$ and $x_j$. $N$ is the total number of pixels the image contains. This fraction normalizes the $d_{\text{IC}}$ value between 0 and 1. According to function (6), the distance of a pixel from itself is 0. The distance is at the maximum of 1 when the first common segment of two pixels is the initial undivided image at the maximum hierarchical level H. When the first common segment of the two pixel points contains more pixels, these two pixels share less information, and the distance between them is greater. Conversely, if the first common segment of two pixel points contains few pixels, the two pixels share more information, and the distance value is smaller. For example, the first common segment of the red and blue pixels and the first common segment of the yellow and green pixels are both at hierarchical level 3. However, the region sizes of these two first common segments are different. The first common segment of the red and blue pixels contains four pixels, and the first common segment of the yellow and green pixels contains three pixels. Therefore, the spatial similarity between them is different. We combine our distance-based IC with the $d_{\text{spatial} - \text{spectral}}$ proposed by Chen et al. to obtain a comprehensive distance formula:

$$D = d_{\text{spatial} - \text{spectral}} \times d_{\text{IC}},$$  \hspace{1cm} (7)
Figure 2 is the flowchart of the proposed spectral-spatial distance-based SRM considering IC. The proposed comprehensive distance formula measures distance between sample both in spatial level and spectral level. Both the height information and IC are taken into account during spatial distance measurement which could increase the accuracy of the spatial distance and the final distance.

2.4. Application of the Proposed Comprehensive Distance Measure

The advantage of the distance-based remote sensing image classification algorithm is that it is simple and easy to execute, but the accuracy of the algorithm needs to be improved. Especially, the traditional distance function in this type classifier only use the spectral characteristic and only depend on the spectral distance between samples during the classification, which ignore the spatial characteristic of remote sensing data. This can be avoiding by applying the hybrid spatial-spectral distance.

The k-NN and OPF algorithms are two commonly used methods based on distance due to their good performance. And they are closely related to each other. The OPF is equivalent to the 1-NN when all training samples are used as prototypes in the OPF [35]. Chen et al. applied their spectral-spatial distance to these two classifiers obtained two contextual classifiers, the OPF-SRM and k-NN-SRM. The comprehensive distance considers the IC proposed in this paper can also be applied to these two distance-based classifiers. We applied our comprehensive distance measures to the k-NN and OPF to obtained two contextual classifiers named as the k-NN-SRM-IC and OPF-SRM-IC, respectively. The comprehensive distance formula proposed in this paper is used to measure distance between samples in k-NN-SRM-IC and OPF-SRM-IC.

2.4.1. k-NN-SRM-IC

The idea of k-NN is each test sample can be represented and classified according to its k-nearest neighbors in the training set. The classifier assigns a test sample to the class that its k-nearest neighbors most belong to. The novel comprehensive distance formula considers IC is used to find the k-nearest neighbors in k-NN-SRM-IC.
2.4.2. OPF-SRM-IC

OPF construct a complete graph \( A \) that each pair of nodes is connected by a single edge in the training stage, the nodes of \( A \) is the samples of training set \( Z_1 \), and the weight of edge is the distance between the nodes connected by this edge. The adjacent nodes with different labels in the minimum spanning tree (MST) of graph \( A \) were chosen as prototypes set \( S \); each prototype represents one class. \( \tau_t \) denotes a path consist of a sequence of adjacent samples start from a root \( R(t) \) and end with sample \( t \). The maximum distance between adjacent samples along \( \tau_t \) defined as the path-cost \( f(\tau_t) \), which reflects connection strength of nodes in path. The connection strength between \( R(t) \) and \( t \) is maximum and path \( \tau_t \) is optimum when \( f(\tau_t) \leq f(\tau_1) \), where \( \tau_1 \) is any other path containing node \( t \). All the other samples in the training set are connected to a prototype with the minimum path-cost \( f(\tau_t) \). And the minimization of \( f(\tau_t) \) is assigned to sample \( t \) as the minimum cost \( C_t \) used for the testing stage. A prototype with all the samples connected constitute an optimum tree, all the optimum trees constitute the optimum forest which is used to classified the testing set.

For a given testing sample \( v \) in the testing set \( Z_2 \), OPF connect it to all the samples \( t \in Z_1 \). Then we find the optimum path \( P(v) \) according to the minimum cost \( C_v \):

\[
C_v = \min \{ \max \{ C_t, d(v, t) \} \}, \quad \forall t \in Z_1, \tag{8}
\]

The prototype included in \( P(v) \) is connected most closely with \( v \) and they are most likely within the same class, so label \( v \) with the class of \( R(v) \). The novel comprehensive distance formula considers IC is used to compute the weight of edge and the cost function in OPF-SRM-IC.

3. Results

3.1. Datasets

The datasets used in this paper are land cover images obtained from the CBERS-2B and Landsat-5 TM covering the Itatinga area of Brazil and land cover images obtained from Ikonos-2 MS and Geoeye covering the Duque de Caxias area of Brazil. The Green, Red and Near IR bands for CBERS-2B, Ikonos-2 MS and Geoeye were used in the experiment. Near IR, Red and SW IR bands for Landsat-5 TM were used in the experiment. These four images are shown in Figure 3. The RGB values were used as features to describe each pixel. The ground truth images of the datasets were annotated by Rodrigo José Pisani [12] and shown in Figure 4. Tables 1 and 2 describe the detailed information of land cover classes included in the datasets.

| Class       | grass lands | reforesting cultures | roads | dams | bushes |
|-------------|-------------|----------------------|-------|------|--------|
| Color       | light green | dark gray            | salmon| gray | red    | green |

| Class       | roads | clear tonal signatures | bare moist soil | tree cover |
|-------------|-------|------------------------|-----------------|-----------|
| Color       | gray  | white                  | salmon          | dark green |

| Class       | shadows | dark tonal signatures | bare soil clear | grasslands |
|-------------|---------|-----------------------|-----------------|-----------|
| Color       | black   | dark tonal signatures | yellow          | light green |
Sensors 2018, 18, x FOR PEER REVIEW 8 of 14

Figure 3. Images obtained from the (a) CBERS-2B CCD (20 m) sensor, band 2 (520–590 nm) = red, band 3 (630–690 nm) = green, band 4 (770–890 nm) = blue, (b) Landsat-5 TM (30 m) sensor, band 4 (760–900 nm) = red, band 3 (630–690 nm) = green, band 5 (1550–1750 nm) = blue, (c) Ikonos-2 MS sensor, band 4 (760–900 nm) = red, band 3 (630–690 nm) = green, band 2 (520–600 nm) = blue, and (d) Geoeye, band 5 (780–920 nm) = red, band 4 (655–690 nm) = green, band 3 (510–580 nm) = blue. The Ikonos-2 MS and Geoeye images were obtained through a fusion process between MS (4 m) and PAN (1 m), resulting in a final spatial resolution of 1 m.

Figure 4. Ground truth images for (a) CBERS-2B, (b) Landsat-5, (c) Ikonos-2 MS and (d) Geoeye.

3.2. Parameter Settings

In addition to comparisons with the original k-NN-SRM and OPF-SRM, the contextual classifiers using the proposed comprehensive distance (k-NN-SRM-IC and OPF-SRM-IC) were also compared against with k-NN, OPF, SVM, the spectral-spatial classifiers based on WH segmentation as well as the spectral-spatial classifiers based on MRF. For the MRF-based classifiers, we used the OPF-MRF proposed by Nakamura [10] and the SVM-MRF proposed by Osaku [12]. The original
spectral-spatial classifier based on WH and MV is the SVM-WH-MV [17], which uses the SVM as the pixel-wise classifier performed on the original hyperspectral image first. Then, for every watershed region, all the pixels are assigned to the most frequent class within this region by MV approach. Other classifiers can also be used as the pixel-wise classifier. Thus, we built another two WH-MV-based spectral-spatial classifiers, the k-NN-WH-MV and the OPF-WH-MV, with the k-NN and OPF pixel-wise classifiers, respectively.

We used the accuracy measure proposed by Papa to assess the classification results [36]. For the OPF, we used LibOPF [37]. For the MRF-based classifiers, we used the code obtained from [12], and for the k-NN and WH-MV-based methods we used our own implementation. The training and testing sets were constructed using the holdout method with 5% and 95%, respectively. Considering that the randomness of the holdout method may affect the results, the classification for each image was executed five times with different partitions. The final accuracy is the mean value of five iterations’ results. The number of iterations for the MRF-based methods is ten. For the SVM-based methods, the kernel is the radial basis function (RBF), and the relevant parameters have been optimized. k was chosen from the interval [1, 10] according to the result having the maximum accuracy. We set fourteen different Q values for each image using the segmentation. They are presented in Table 3 with the corresponding hierarchical level. When Q increases to a certain extent, the result of the segmentation tends to stabilize, and we can determine the maximum value of Q.

### Table 3. The Q values and the corresponding hierarchical level values of h used in the SRM, and $H = 15$.

| Q  | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 |
|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| h  | 14 | 13 | 12 | 11 | 10 | 9  | 8  | 7  | 6  | 5  | 4  | 3  | 2  | 1  |

### 3.3. Quantitative Results

The accuracy of the classification methods using the proposed distance measure and comparison methods for all images is shown in Table 4. The values in brackets are the parameter values for the relevant classifiers, including β used in two MRF-based classifiers and k with the best results in the k-NN-based classifiers. The spectral-spatial classifiers using the proposed comprehensive distance that considers the IC have an obvious improvement. The accuracy of SRM-based methods that consider IC is higher than SRM-based methods without IC. Among all the methods, the k-NN-SRM-IC has the best performance and obtained the highest accuracy. For each pixel-wise classifier, the SRM-based methods performed better than WH-MV-based methods, which indicate that SRM is more effective than WH in capturing spatial information. The k-NN-SRM outperformed the SVM-MRF in the case that the k-NN has similar performance to the SVM, and the OPF-SRM outperformed OPF-MRF. This illustrates that the spatial information obtained by the SRM is much more accurate than the spatial information obtained by the MRF.

### Table 4. The mean accuracy obtained by all methods on four land cover images.

| Method/Image | CBERS-2B | Landsat-5 | Ikonos-2MS | Geoeye |
|--------------|----------|-----------|------------|--------|
| k-NN         | 69.95%(10) | 71.43%(7) | 68.84%(1) | 75.32%(9) |
| k-NN-WH-MV   | 72.17%(1) | 72.72%(1) | 73.13%(1) | 77.36%(1) |
| k-NN-SRM     | 76.90%(7) | 76.69%(3) | 79.49%(10) | 82.45%(9) |
| k-NN-SRM-IC  | 77.91%(7) | 77.75%(3) | 80.45%(10) | 83.62%(9) |
| OPF          | 66.04% | 68.04% | 68.00% | 69.31% |
| OPF-WH-MV    | 70.74% | 71.19% | 72.27% | 73.72% |
| OPF-MRF      | 68.90%(0.6) | 71.11%(0.8) | 69.42%(1.2) | 72.97%(1.0) |
| OPF-SRM      | 73.65% | 75.36% | 78.84% | 78.13% |
| OPF-SRM-IC   | 73.86% | 76.83% | 80.06% | 79.02% |
| SVM          | 70.26% | 71.21% | 66.73% | 76% |
| SVM-WH-MV    | 70.35% | 71.40% | 67.81% | 76.47% |
| SVM-MRF      | 73.73%(1.2) | 74.96%(1.3) | 72.16%(1.2) | 76.81%(1.3) |
3.4. Visual Results

Figure 5 shows the classification results for the CBERS-2B image. The k-NN-SRM-IC and k-NN-SRM obtained similar results that classified all the classes correctly, and the k-NN-SRM-IC is more accurate than the k-NN-SRM in culture. For the OPF-SRM and OPF-SRM-IC classifiers, reforesting and bushes were confused. The OPF failed to identify culture, reforesting and bushes. All the other methods cannot recognize culture completely. There are many discontinuous fragments in the results of methods using the WH and MV mechanisms.

As shown in Figure 6, the k-NN-SRM-IC and k-NN-SRM recognized all the classes correctly for the Landsat-5 image, and the k-NN-SRM-IC is more accurate than the k-NN-SRM in culture. The OPF, OPF-SRM, OPF-SRM-IC, and OPF-WH-MV confused a part of grasslands and reforesting. Other methods’ results are similar in that they misclassified a portion of reforesting and culture.

![Classification results on the CBERS-2B image](image1.png)

Figure 5. Classification results on the CBERS-2B image obtained by using the (a) k-NN, (b) k-NN-WH-MV, (c) k-NN-SRM, (d) k-NN-SRM-IC, (e) OPF, (f) OPF-WH-MV, (g) OPF-MRF, (h) OPF-SRM, (i) OPF-SRM-IC, (j) SVM, (k) SVM-WH-MV, (l) SVM-MRF.

![Classification results on the Landsat-5 image](image2.png)

Figure 6. Classification results on the Landsat-5 image obtained by using the (a) k-NN, (b) k-NN-WH-MV, (c) k-NN-SRM, (d) k-NN-SRM-IC, (e) OPF, (f) OPF-WH-MV, (g) OPF-MRF, (h) OPF-SRM, (i) OPF-SRM-IC, (j) SVM, (k) SVM-WH-MV, (l) SVM-MRF.

As shown in Figure 7, the result of the k-NN-SRM-IC is still the best for the Ikonos-2 MS image, and the k-NN-SRM is second. The OPF-SRM and OPF-SRM-IC misclassified bare soil as grassland. All the other methods misidentified tree coverage, clear tonal signatures and the covering of dark tonal signatures as road.
As shown in Figure 8, the k-NN-SRM-IC classified all the classes correctly and achieved the best result on the Geoeye image once again. The k-NN-SRM can also distinguish all the classes, but the performance is poorer than the k-NN-SRM-IC in clear tonal signatures and bare, clear soil. The OPF-SRM and OPF-SRM-IC misclassified the clear tonal signatures. The k-NN, k-NN-WH-MV and OPF-MRF misclassified tree coverage and average tonal signatures. The OPF and OPF-WH-MV misclassified the clear tonal signatures, tree coverage and average tonal signatures. The SVM, SVM-MRF and SVM-WH-MV misclassified dark tonal signatures, tree coverage and average tonal signatures.

Figure 7. Classification results on the Ikonos-2 image obtained by using the (a) k-NN, (b) k-NN-WH-MV, (c) k-NN-SRM, (d) k-NN-SRM-IC, (e) OPF, (f) OPF-WH-MV, (g) OPF-MRF, (h) OPF-SRM, (i) OPF-SRM-IC, (j) SVM, (k) SVM-WH-MV, (l) SVM-MRF.

Figure 8. Classification results on the Geoeye image obtained by using the (a) k-NN, (b) k-NN-WH-MV, (c) k-NN-SRM, (e) k-NN-SRM-IC, (e) OPF, (f) OPF-WH-MV, (g) OPF-MRF, (h) OPF-SRM, (i) OPF-SRM-IC, (j) SVM, (k) SVM-WH-MV, (l) SVM-MRF.

4. Discussion

The drawback of segmentation-based spectral-spatial classification methods is that they are highly dependent on the performance of the segmentation methods used, which includes two aspects: one is the accuracy of the segmentation results; the other is the information expression/capture capability of the segmentation results. The strategy based on MV belongs to the first case, the segmentation results are used directly to determine the final classification of samples. Once there are mistakes in the segmentation results, the final classification results will be affected. Such as the methods of combining...
WH and MV can improve the original classification to a certain extent, however, there are still many discontinuous fragments without practical significance. This is due to the WH segmentation results are over segmented.

In the methods of construct the spectral-spatial distance depending on segmentation results, the quality of spatial information captured from segmentation results affect the final classification results. Simply and insufficient spatial information express/capture from segmentation results may reduce the classification accuracy. In our proposed method, the spatial information capture capability of the segmentation results is enhanced. Our proposed method consider both the height information and the IC of samples to measure the spatial distance, which take full advantage of segmentation results and capture more complex spatial information. The performance of the k-NN-SRM-IC is better than the k-NN-SRM, and the OPF-SRM-IC outperformed the OPF-SRM in both quantitative and visual assessments, which illustrate the proposed comprehensive spectral-spatial distance considers IC is more effective than the spectral-spatial distance that only used the hierarchical level to capture spatial information. The results also indicate the feasibility and effectiveness that using the notion of the IC that two pixels share to measure the spatial distance between them is correct.

The classification results based on the MRF are complementary with the classification results based on the SRM. The classes that are correctly recognized by the MRF-based methods cannot be identified by SRM-based methods. Conversely, those classes that are identified correctly by the SRM-based methods are misclassified by the MRF-based methods. Thus, we can consider combining the MRF and SRM together to improve the classification in the future.

5. Conclusions

In this paper, we introduce IC to measure the spectral-spatial distance. We use the number of pixels contained in a segment to measure its IC. The more pixels that a segment includes, the more abstract that it is, and the less information the segment has. The fewer pixels that a segment includes, the more specific that it is, and the more information the segment has. The spatial distance between pixels can be computed according to the IC of their first common segment in a hierarchical tree constructed by the results of SRM multiscale segmentation. Integrating the spatial distance computed by the IC with the spectral-spatial distance that utilizes the hierarchical level and spectral features, we can obtain a novel spectral-spatial distance measure that can capture more spatial information. The novel spectral-spatial distance can be applied to the traditional distance-based classification method to improve the performance, such as the k-NN and OPF. The novel distance-based contextual classifiers using our novel spectral-spatial distance function, named the k-NN-SRM-IC and OPF-SRM-IC, are implemented in four land cover images. The k-NN-SRM-IC obtains the highest accuracy among all methods. The classification results proved the accuracy of the IC-based distance measure function and proved that the novel spectral-spatial distance function has the capability to capture more spatially related information between pixels. In the future, we will explore applying the proposed spectral-spatial distance in other contextual classifiers to improve the classification performance.

Author Contributions: Conceptualization, S.C.; Data curation, J.Z. and X.G.; Methodology, S.C.; Software, S.C.; Supervision, H.Z. and T.S.; Validation, J.Z. and X.G.; Visualization, J.Z. and X.G.; Writing-original draft, S.C.; Writing-review & editing, H.Z. and T.S.

Funding: This paper was supported by the National Natural Science Foundation of China (Grant No. 41771450, 41601438, 41501449, 41871330), the Fundamental Research Funds for the Central Universities (Grant No. 2412017FZ021) and the Foundation of the Education Department of Jilin Province in the 13th Five-Year project (Grant No. JJKH20170916K).

Acknowledgments: The authors would like to thank Rodrigo José Pisani for providing the datasets.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* 2007, 28, 823–870. [CrossRef]

2. Stuckens, J.; Coppin, P.R.; Bauer, M.E. Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sens. Environ.* 2000, 71, 282–296. [CrossRef]

3. Magnussen, S.; Boudewyn, P.; Walder, M. Contextual classification of Landsat TM images to forest inventory cover types. *Int. J. Remote Sens.* 2004, 25, 2421–2440. [CrossRef]

4. Ghamisi, P.; Souza, R.; Benediktsson, J.A.; Rittner, L.; Lotufo, R.; Zhu, X.X. Hyperspectral Data Classification Using Extended Extinction Profiles. *IEEE Geosci. Remote Sens. Lett.* 2016, 13, 1641–1645. [CrossRef]

5. Kartikeyan, B.; Sarker, A.; Majumder, K.L. A segmentation approach to classification of remote sensing imagery. *Int. J. Remote Sens.* 1999, 19, 1695–1709. [CrossRef]

6. Chen, Y.; Lin, Z.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2014, 7, 2094–2107. [CrossRef]

7. Shekhar, S.; Schrater, P.; Vatsavai, R.R.; Wu, W.; Chawla, S. Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. *IEEE Trans. Multimed.* 2002, 4, 174–188. [CrossRef]

8. Tarabalka, Y.; Fauvel, M.; Chanussot, J.; Benediktsson, J.A. SVM-and MRF-based Method for Accurate Classification of Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* 2010, 7, 736–740. [CrossRef]

9. Moser, G.; Seprpico, S.B. Combining Support Vector Machines and Markov Random Fields in an Integrated Framework for Contextual Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2013, 51, 2734–2752. [CrossRef]

10. Nakamura, R.Y.M.; Osaku, D.; Levada, A.L.M.; Cappabianco, F.; Falcão, A. OPF-MRF: Optimum-path Forest and Markov Random Fields for Contextual-based Image Classification. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, York, UK, 27–29 August 2013; pp. 233–240.

11. Osaku, D.; Levada, A.L.M.; Papa, J.P. On the Influence of Markovian Models for Contextual-Based Optimum-Path Forest Classification. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 3946–3952. [CrossRef]

12. Nakamura, R.Y.M.; Pereira, L.A.M.; Pisani, R.J.; Levada, A.L.M.; Cappabianco, F.A.M.; Falcão, A.X.; Papa, J.P. Improving Land Cover Classification through Contextual-based Optimum-path Forest. *Int. J. Remote Sens.* 2015, 36, 6509–6532. [CrossRef]

13. Zhang, G.; Jia, X. Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification. *IEEE Geosci. Remote Sens. Lett.* 2012, 9, 856–860. [CrossRef]

14. Zhao, J.; Zhong, Y.; Shu, H.; Zhang, L. High-resolution image classification integrating spectral-spatial-location cues by conditional random fields. *IEEE Trans. Image Process.* 2016, 25, 4033–4045. [CrossRef] [PubMed]

15. Zhang, G.; Jia, X.; Wang, X.; Tuia, D.; Semisupervised Classification of Remote Sensing Images with Hierarchical Spatial Similarity. *IEEE Geosci. Remote Sens. Lett.* 2015, 12, 150–154. [CrossRef]

16. Fauvel, M.; Benediktsson, J.A.; Chanussot, J.; Sveinsson, J.R. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. *IEEE Trans. Geosci. Remote Sens.* 2008, 46, 3804–3814. [CrossRef]

17. Dalla Mura, M.; Benediktsson, J.A.; Waske, B.; Bruzzone, L. Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Trans. Geosci. Remote Sens.* 2010, 48, 3747–3762. [CrossRef]

18. Dalla Mura, M.; Benediktsson, J.A.; Waske, B.; Bruzzone, L. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *Int. J. Remote Sens.* 2010, 31, 5975–5991. [CrossRef]
23. Ghamisi, P.; Dalla Mura, M.; Benediktsson, J.A. A survey on spectral–spatial classification techniques based on attribute profiles. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 2335–2353. [CrossRef]
24. Ghamisi, P.; Souza, R.; Benediktsson, J.A.; Zhu, X.X.; Rittner, L.; Lotufo, R.A. Extinction Profiles for the Classification of Remote Sensing Data. *IEEE Trans. Geosci. Remote Sens.* 2016, 54, 5631–5645. [CrossRef]
25. Chen, Y.; Zhao, X.; Jia, X. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, 8, 2381–2392. [CrossRef]
26. Lee, H.; Kwon, H. Contextual deep CNN based hyperspectral classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 3322–3325.
27. Yang, J.; Zhao, Y.; Chan, J.C.W.; Yi, C. Hyperspectral image classification using two-channel deep convolutional neural network. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium, Beijing, China, 10–15 July 2016; pp. 5079–5082.
28. Chen, C.; Jiang, F.; Yang, C.; Rho, S.; Shen, W.; Liu, S.; Liu, Z. Hyperspectral classification based on spectral–spatial convolutional neural networks. *Eng. Appl. Artif. Intell.* 2018, 68, 165–171. [CrossRef]
29. Aptoula, E.; Ozdemir, M.C.; Yanikoglu, B. Deep Learning With Attribute Profiles for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* 2016, 13, 1970–1974. [CrossRef]
30. Yuan, H.; Tang, Y.; Lu, Y.; Yang, L.; Luo, H. Spectral-Spatial Classification of Hyperspectral Image Based on Discriminant Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2014, 7, 2035–2043. [CrossRef]
31. Sellami, A.; Farah, I.R. High-level hyperspectral image classification based on spectro-spatial dimensionality reduction. *Spat. Stat.* 2016, 16, 103–117. [CrossRef]
32. Chen, S.Y.; Sun, T.L.; Yang, F.Q.; Sun, H.G.; Guan, Y. Remote sensing image classification using the spatial information obtained by statistical region merging. *Remote Sens. Lett.* 2017, 8, 567–575. [CrossRef]
33. Nock, R.; Nielsen, F. Statistical Region Merging. *IEEE Trans. Pattern Anal. Mach. Intell.* 2004, 26, 1452–1458. [CrossRef] [PubMed]
34. Resnik, P. Using information content to evaluate semantic similarity in a taxonomy. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI 95), Montreal, QC, Canada, 20–25 August 1995; pp. 448–453.
35. Souza, R.; Rittner, L.; Lotufo, R. A Comparison between Optimum Path Forest and k-Nearest Neighbors Classifiers. In Proceedings of the 25th SIBGRAPI Conference on Graphics, Patterns and Images, Ouro Preto, Brazil, 22–25 August 2012; pp. 260–267.
36. Papa, J.P.; Falcao, A.X.; Suzuki, C.T.N. Supervised Pattern Classification Based on Optimum-Path Forest. *Int. J. Imaging Syst. Technol.* 2009, 19, 120–131. [CrossRef]
37. LibOPF: A library for the design of optimum-path forest classifiers 2009. Software version 2.0. Available online: http://www.ic.unicamp.br/~afalcao/libopf/ (accessed on 15 March 2018).