ISPRS Journal of Photogrammetry and Remote Sensing 169 (2020) 320–336

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Integrating spectral variability and spatial distribution for object-based image analysis using curve matching approaches

Yunwei Tang [a], Fang Qiu [b,*], Linhai Jing [a], Fan Shi [b], Xiao Li [b]

[a] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, No. 9 South Road, Beijing 100094, China
[b] Geospatial Information Sciences, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA

ARTICLE INFO

Keywords:
Object-based image analysis
Curve matching
His-Cov
Spatial covariogram
Land cover/land use classification
High resolution

ABSTRACT

Object-based image analysis (OBIA) has been widely used to classify high spatial resolution (HSR) imagery. In a traditional OBIA, object-level statistical summaries such as mean values are usually used for classification. This implies that the spectral values within objects follow a Gaussian distribution. However, the pixel values in an object do not necessarily conform to a Gaussian distribution because of within object spectral heterogeneity. Consequently, these statistical summaries may misrepresent the features of the object. This shortcoming is addressed in this paper by integrating both the spectral variability and the spatial distribution of the pixels within objects to improve the traditional object-based image classification. The spectral variability is represented by histograms of the pixel values in the object, and the spatial distribution is characterized by the binary spatial covariogram of these pixels. To construct a binary spatial covariogram, a principal component analysis (PCA) is first applied to compress multiple bands into one, and the Otsu thresholding is then performed to generate a binary map reflecting the spatial configuration of the pixels. Spatial covariance is then computed for this binary map and plotted with different lag distances to derive the binary spatial covariogram. Our proposed model utilizing curves composed of the spectral histograms and binary spatial covariogram (referred to as the His-Cov model) are then used for classification based on curve matching approaches. The integration of spectral variability and spatial distribution of the pixels in the object produced superior results to curve matching approaches based on spectral variability alone and to traditional OBIA based on spectral and spatial features of the objects when classifying complex land use types in urban environments.

1. Introduction

With advances in remote sensing technologies and growing demand for detailed spatial information, high spatial resolution (HSR) imagery is becoming one of the most used remote sensing data types for a variety of applications (Belward and Skøien, 2015). The higher spatial resolution of the data has posed great challenges to pixel-based land cover and land use classification, which fails to achieve desirable accuracy despite HSR imagery's promise of more detailed geospatial information (Wang et al., 2013). Object-based image analysis (OBIA) has now been widely accepted to be superior to pixel-based classification of HSR remotely sensed data, primarily because OBIA avoids salt-and-pepper noise, which commonly exists in the pixel-based classification (Blaschke and Strobl, 2001; Ma et al., 2017; Hossain and Chen, 2019; El-naggar, 2018; Su, 2019; Shen et al., 2019). A traditional object-based classification begins with image segmentation. This partitions the image scene into multiple image segments (set of pixels, also known as objects) according to some pre-defined homogeneity criteria (Blaschke, 2010; Newman et al., 2011). Classification is then undertaken at the object level instead of the individual pixel level. Most commonly, inputs into OBIA classifiers are based on statistical summaries of all the pixel values in an object, such as the mean and standard deviation. To further improve classification, object-level spatial features have also been incorporated based on summaries such as shape complexity (Russ, 2002; Jiao et al., 2012) or a compactness index (Hay et al., 1996; Zhong et al., 2020). Most of these object-based classifiers use standard machine learning algorithms such as the support vector machine (SVM) (Geiß and Taubenböck, 2015), decision trees (DT) (Laliberte, et al., 2007), random forest (RF) (Melville et al., 2018), multi-layer perceptron (MLP) (Zhang et al., 2018), K-nearest neighbor (KNN) (Tang et al., 2016) and sometimes boosting classifiers (Georganos, et al., 2018).

In traditional OBIA, each object is treated as if it is a pixel and the object-level statistical summaries are treated as their input pixel values. These object-level summaries are single-valued since only one value is

* Corresponding author.

https://doi.org/10.1016/j.isprsjprs.2020.09.023
Received 3 May 2020; Received in revised form 29 August 2020; Accepted 30 September 2020
Available online 9 October 2020

used to characterize one type of feature for an object. As a result, not all pixels of an image object participate in the object-based image classification, but only their object-level summaries. Since only single-valued statistical summaries are used to characterize each object, the rich information provided by the many pixels constituting an object in HSR imagery is lost (Arroyo et al., 2006). Consequently, the utilization of traditional classifiers based on object-level summaries does not take full advantage of the potential that HSR images could provide. Additionally, these statistical summaries are representative of the object's characteristics only when the pixel values follow a Gaussian distribution (Pedley and Curran, 1991; Shackelford and Davis, 2003; Berger et al., 2013). Unfortunately, a non-Gaussian frequency distribution is common for these pixel values (Stow et al., 2012; Sridharan and Qiu, 2013; Toure et al., 2013) in HSR imagery. At sub-meter level resolution, a real-world object, which is often composed of multiple elements, may not exhibit the spectral homogeneity that is often observed in an image at medium or coarse spatial resolution (Costa et al., 2017). For example, a building object can be an integration of various components with different materials (e.g., glass and brick) and structures (e.g., chimney and flat roof) (Zhou and Troy, 2008). Since within-object spectral heterogeneity increases with the spatial resolution of imagery (Blaschke et al., 2004), statistical summaries may misrepresent the features of the objects and mislead the subsequent classification (Sridharan and Qiu, 2013).

To overcome this problem, novel classifiers based on curve matching approaches have been proposed for object-based image classification (Stow et al., 2012; Sridharan and Qiu, 2013). These approaches were originally designed for hyperspectral image classification (van der Meer, 2000, 2006; Liu and Han, 2017) and have been used for time-series analysis (Vorobiova and Chernov, 2017; Gao et al., 2020). The curve matching approaches measure spectral similarity/divergence between a spectrum of an unknown object and prior spectra from a spectral library (i.e., samples) (Kruse et al., 1993). Stow et al. (2012) suggested using the pixel frequency distribution of an object (i.e., its histogram signature) via a curve matching approach named the histogram matching root sum squared differential area (RSSDA) for classification. It was later compared with using a curve angle mapper (CAM) to classify an airborne multispectral image (Toure et al. 2013). Independently, Sridharan and Qiu (2013) proposed a curve matching approach based on Kolmogorov-Smirnov (KS) distance using the empirical cumulative distribution function (i.e., cumulative histogram) of the pixels in an object as input for object-based classification. Building on this approach, Zhou and Qiu (2015) developed a method to synthesize object-based pseudo-waveform from discrete-returns LiDAR data by utilizing the count or intensity based histogram of the returns for an object. The pseudo-waveforms were then fused with the object-level spectral histograms from HSR imagery to improve object-based classification using a Kullback-Leibler (KL) divergence-based curve matching approach. Compared to object-level statistical summaries, an object-level histogram/pseudo-waveform provides far richer information about the spectral/structural components of an object since it characterizes their variability within the object. These studies have demonstrated that curve matching approaches based on object-level histogram/pseudo-waveform can achieve better performances than traditional OBIA based on statistical summaries.

Although the advantages of curve matching approaches to OBIA have been demonstrated, histogram features can only indicate the frequency distribution of pixel values. They are unable to reflect their spatial distribution or arrangement within the objects. Therefore, current curve matching approaches cannot distinguish two objects having similar frequency distributions but a different spatial distribution of the pixels. In previous studies, (semi)variogram and covariogram measures have proven to be effective to extract spatial features and can be incorporated with spectral features to improve classification (Chica-Olmo and Abarca-Hernández, 2000; Berberoglu et al., 2000; Chen and Gong, 2004; Wu et al., 2006; Berberoglu et al., 2007; Balaguer et al., 2010; Balaguer-Beser et al., 2013; Wu et al., 2015). However, previous studies were all based on traditional classifiers and the spatial features have not been incorporated in curve matching approaches. To address this issue, a novel model is proposed which integrates both spectral and spatial features into a curve matching based classifier. We designate the previous curve matching model, which used an object-level histogram or cumulative histogram, as the His model. The newly developed model, which incorporates object-level spatial distribution using the binary spatial covariogram, is designated as the His-Cov model. This study tests whether the His-Cov model can improve upon a curve matching based classification using only spectral features (i.e., the His model). The ability of the His-Cov model is also compared with some state-of-the-art classifiers based on machine learning algorithms which also incorporate both spectral and spatial features.

2. Methods

As with most object-based classification methods, the His-Cov model includes three general steps: image segmentation, feature extraction, and classification. After segmentation, the object-level spectral and spatial features are extracted from image objects. For the spectral feature, the histogram for each multispectral band is generated by using all the pixels in an object to represent the spectral variety of its pixels. For the spatial feature, principal component analysis (PCA) is utilized to compress all the bands of an image into the first principle component. Otsu thresholding (Otsu, 1979) is then applied to this single-band image to derive a binary map to reflect the general spatial configuration of the pixels in the objects. Binary spatial covariograms (or simply "spatial covariograms" for short) are then produced based on the spatial covariance in east–west and north–south directions to represent the spatial distribution of the pixels in an object. The spectral histogram and spatial covariagram are then fed into a curve matching approach based on the curve divergence between two objects for classification. During this process, each testing object is compared with all the training objects and is labeled with the class of the training object that has the minimum

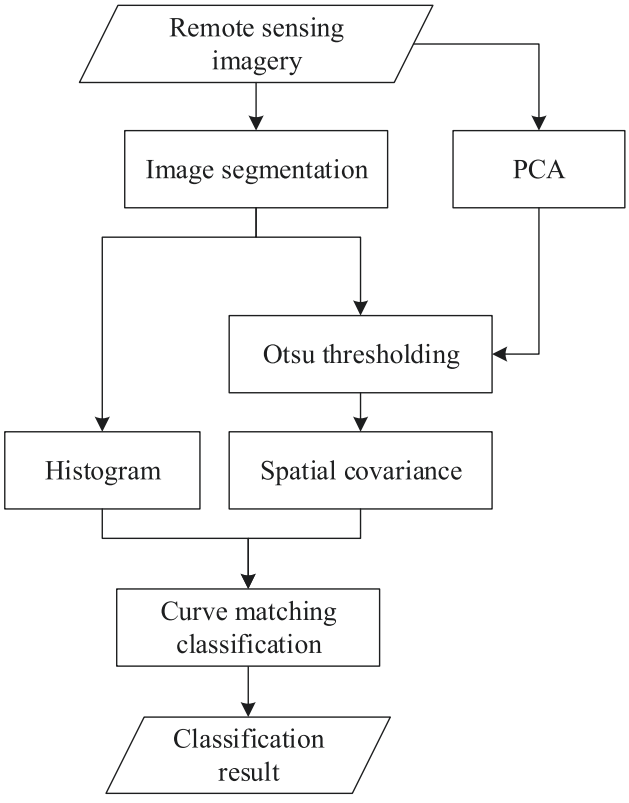

Fig. 1. Flowchart of the His-Cov model.

divergence with the testing object. The flowchart of the His-Cov model is shown in Fig. 1.

2.1. Image segmentation

Small objects can be problematic when exploring the spectral variability and spatial distribution of pixels in an object since they may not provide sufficient data to be representative of a given land use type (Stow et al., 2012). Therefore, the expectation for image segmentation in this study is that a segment should include one dominant class. For example, a road object may include a crosswalk and cars, and a building object may include roof parts of different materials. Image segmentation based on the simple linear iterative clustering (SLIC) (Radhakrishna et al., 2012) and fractal net evolution algorithms (FNEA) (Comaniciu and Meer, 2002) are adopted in this study. The SLIC algorithm outputs a desired number of regular and compact objects with a low computational load. The algorithm is first adopted to ensure that the segmented objects do not have extremely strange shapes, such as centroids outside their boundaries. A regular shape is beneficial to the exploration of spatial distribution. Since the SLIC may suffer from a severe under-segmentation problem for small objects with rich spatial details, the multi-resolution FNEA method is then applied to the SLIC output at a fine scale to produce the final segmentation result. The FNEA is a bottom-up region-merging approach based on local homogeneity criteria, with small image objects iteratively merged into larger ones (Comaniciu and Meer, 2002).

2.2. Feature extraction

The within-object spectral feature at a given multispectral band is represented using relative frequency (i.e., a histogram) or cumulative - frequency (i.e., a cumulative histogram) of all the pixels within this object. Relative means that the values are normalized between zero and one. The former is often referred to as the empirical probability distribution function (PDF), and the latter as the empirical cumulative distribution function (CDF) of the objects. The PDF and CDF curves of an object reflect the empirical distributions of its pixel's spectral variability and can better distinguish different land cover classes than traditional OBIA based on single-valued statistical summaries (Sridharan and Qiu, 2013).

To characterize the pixel spatial distribution of an object, we first transform the pixel values in a sample object to categorical variables to capture the general spatial configuration of the pixels. As in many other spectral-spatial classification methods (Dell'Acqua et al., 2004; Benediktsson et al., 2003), the first spectral principal component (PC) is used. The first PC maximizes information from all the original multispectral bands. The Otsu thresholding method is then applied to the first PC for each object to derive the spatial configuration of the pixels in that object. The Otsu thresholding algorithm returns a single intensity threshold that can be used to separate a group of pixels into two classes, foreground and background. This threshold is iteratively determined by minimizing intra-class intensity variance, or equivalently by maximizing inter-class variance (Otsu, 1979). After applying Otsu thresholding for each sample object, a binary map is derived in which pixel values less than or equal to the threshold are assigned a one (i.e., the foreground) and all other pixels are assigned a zero (i.e. the background).

The covariogram and variogram are two common geostatistical measures for modeling spatial distributions. The former measures similarity (auto-covariance) between pixels and their neighbors at various distances, whereas the latter measures dissimilarity at various distances. The covariogram is formed from a series of spatial covariances at a range of lag distances. Atkinson and Naser (2010) used it to measure similarity for categorical variables. We use the covariogram to measure similarity for a binary variable. The spatial covariance of a binary variable in an object $\mathrm{cov}(h)_{\mathrm{obj}}$ is computed as:

$$cov(h)_{\mathrm{obj}} = \frac{\sum_{i=1,\, i+h\in \mathrm{obj}}^{M} I(u_{i+h}=1 | u_i = 1)}{\sum_{i=1}^{M} I(u_i = 1)} \quad (1)$$

where u_i represents a pixel u at location i in an object, M is the number of pixels in the object. I is an indicator function taking a value of one where the conditional statement is true and zero where it is false, whereas $u_i = 1$ means the pixel u_i belongs to the foreground (derived by the Otsu thresholding). h is a spatial lag in a certain direction, and u_{i+h} is a pixel separated by distance h from pixel u_i. The pixel u_{i+h} is evaluated only if its location $i + h$ is not beyond the border of the current object. The condition of the numerator is stricter than the condition of the denominator. Therefore, Eq. (1) is a conditional probability ranging from 0 to 1. When h equals 0, the spatial covariance equals 1. The spatial covariance of an auto-correlated pattern is usually larger for two pixels separated by a smaller distance than a larger distance, whereas the random noise is presumed to have a constant spatial covariance (Tang et al., 2019). Since the spatial covariance is detected within an object, the maximum lag distance h cannot exceed the smaller dimension of the object. The spatial covariance along four directions (east, west, north and south) is detected. We use the average spatial covariance in the east and the west directions as one feature for the spatial covariance in the east–west direction instead of two. A similar calculation is used for the north–south direction. The spatial covariogram is then plotted with the lag distances on the x-axis and the spatial covariance on the y-axis.

To illustrate, two pairs of simulated images are shown in Fig. 2. Each image can be viewed as an object. The first pair of images includes pixels with differing spatial aggregation, and the second pair has regular but different overall spatial patterns. Since these images consist of 50 × 50 pixels, the maximum lag distance h is set at 40 pixels to plot the spatial covariogram. For the first pair (Fig. 2a), the Moran coefficient (MC) of the upper image has a high of 0.92, common for remotely sensed imagery. The lower image is a random distribution of the pixels in the upper image, so it appears as random noise and the MC equals 0.09. The second pair of images both contain contrasting 1600 low value pixels and 900 high value pixels (Fig. 2b) but their patterns differ. The upper image includes four squares and the lower five stripes. The binary results using Otsu thresholding reflect the general spatial configuration of these images and are displayed beside their original counterparts.

In the first pair of images, the histogram and cumulative histogram are identical, although the spatial distribution of the two images is very different. The spatial covariogram of the highly spatial auto-correlated image shows the typical curve common to most remotely sensed images. It has higher values for smaller than larger distances, and it becomes stabilized at a low value when a certain lag distance is reached, such as the case in the spatial covariogram for the east–west direction. The stabilized value is known as the sill and the lag distance at the sill is known as the correlation range. Correlations decrease with increasing range distances, but do not change beyond the range. However, in some cases high values reappear at longer distances because of the repetition of a pattern or spatial configuration, such as the case in the spatial covariogram for the north–south direction. The spatial covariogram of the lower image has little variation in either direction, indicating there is no spatial similarity in random noise.

In the second pair of images, the spatial covariogram in both directions effectively distinguishes between the images with stripes and squares. However, neither histograms nor cumulative histograms can distinguish these different patterns because they have similar frequency distributions of spectral values for the pixels. The spatial covariogram of the stripe image in the east–west direction has more peaks than that of the square image, indicating more repetitions of the spatial configuration are found in the stripe image. Since the spatial distance between two repetitions (i.e., two stripes) is 9 pixels, the spatial lag between two peaks in the spatial covariogram also equals 9. The distance between

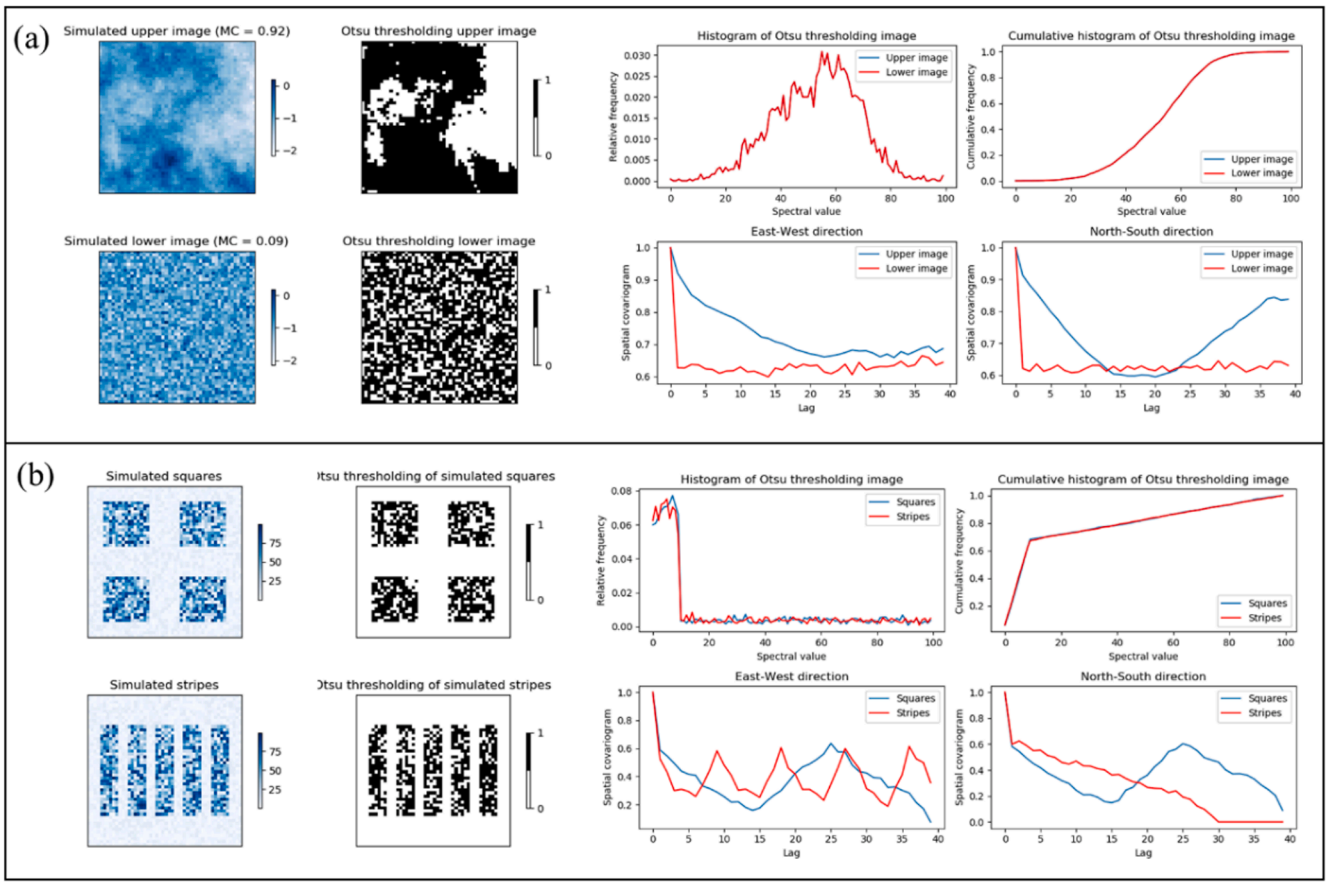

Fig. 2. Frequency distribution of spectral values and spatial covariogram for simulated images: (a) images with different spatial aggregation, (b) images with different overall patterns.

repetitions in the stripe image is smaller than that of the square image, as indicated by the two curves of their spatial covariograms for the east–west direction.

2.3. Classification

For classification, the spectral histogram and spatial covariogram are used as the spectral and spatial features respectively for an object. Since both are in the form of curves, curve matching approaches are used. This approach evaluates the divergence between two image objects by comparing their feature curves. Six curve matching approaches were used in the experiments: Kullback-Leibler (KL) divergence (Chang, 2000; Zhou and Qiu, 2015), Kolmogorov-Smirnov (KS) test (Burt and Barber, 1996; Sridharan and Qiu, 2013), curve angle mapper (CAM) (Kruse et al., 1993; Toure et al. 2013), cumulative curve angle mapper (CCAM), root sum squared differential area (RSSDA) (Hamada et al., 2007; Stow et al., 2012), and cumulative root sum squared differential area (CRSSDA) (Zhou et al., 2016). The aim is to assess if the addition of object-level spatial features (i.e., His-Cov model) can enhance classification compared to using only spectral features (i.e., His model) with these six curve matching approaches.

Given two objects, their feature curves are represented as P_1, P_2, and their cumulative feature curves are presented as F_1, F_2. The divergence D measured by the above six curve matching approaches is as follows.

$$D_{\mathrm{KL}} = \frac{1}{2}\left[\sum_{i=1}^{n} P_1(i) log\frac{P_1(i)}{P_2(i)} + \sum_{i=1}^{n} P_2(i) log\frac{P_2(i)}{P_1(i)}\right] \tag{2}$$

$$D_{\mathrm{KS}} = \sum_{i=1}^{n} max[F_1(i) - F_2(i)] \tag{3}$$

$$D_{\mathrm{CAM}} = arccos\left[\frac{\sum_{i=1}^{n} P_1(i)P_2(i)}{\sqrt{\sum_{i=1}^{n} P_1(i)^2 \sum_{i=1}^{n} P_2(i)^2}}\right] \tag{4}$$

$$D_{\mathrm{CCAM}} = arccos\left[\frac{\sum_{i=1}^{n} F_1(i)F_2(i)}{\sqrt{\sum_{i=1}^{n} F_1(i)^2 \sum_{i=1}^{n} F_2(i)^2}}\right] \tag{5}$$

$$D_{\mathrm{RSSDA}} = \sqrt{\sum_{i=1}^{n} [P_1(i) - P_2(i)]^2} \tag{6}$$

$$D_{\mathrm{CRSSDA}} = \sqrt{\sum_{i=1}^{n} [F_1(i) - F_2(i)]^2} \tag{7}$$

where i is a discrete interval of the curves, and n is the number of intervals. For the spectral feature, curves P and F are the spectral histogram and cumulative histogram, respectively, and n is the number of histogram bins. For the spatial feature, curves P and F are the spatial covariogram and cumulative spatial covariogram, respectively, and n in this case represents the number of spatial lags. For an object, the above divergence is calculated for k multispectral bands in two directions (east–west and north–south). The number of bins and number of lags can differ. The final divergence was determined by combining spectral and spatial features according to Eq. (8):

$$D_{\mathrm{hist\ -\ cov}} = w \cdot k \cdot D_{\mathrm{spectral}} + 2(1-w) \cdot D_{\mathrm{spatial}} \tag{8}$$

where D_{spectral} and D_{spatial} represent divergences derived from spectral

and spatial features, respectively. Both divergences are based on the same curve matching approach and are derived from one of the approaches in Eqs. (2)–(7). k is the number of multispectral bands. w is the weight for the spectral feature ranging from 0 to 1. This weight can be decided by cross-validation or trial-and-error.

To illustrate, two examples of KL based classification are shown in Fig. 3. The first object to be classified belongs to the parking lot class (Fig. 3a). The curve of this object and those of the samples with the five

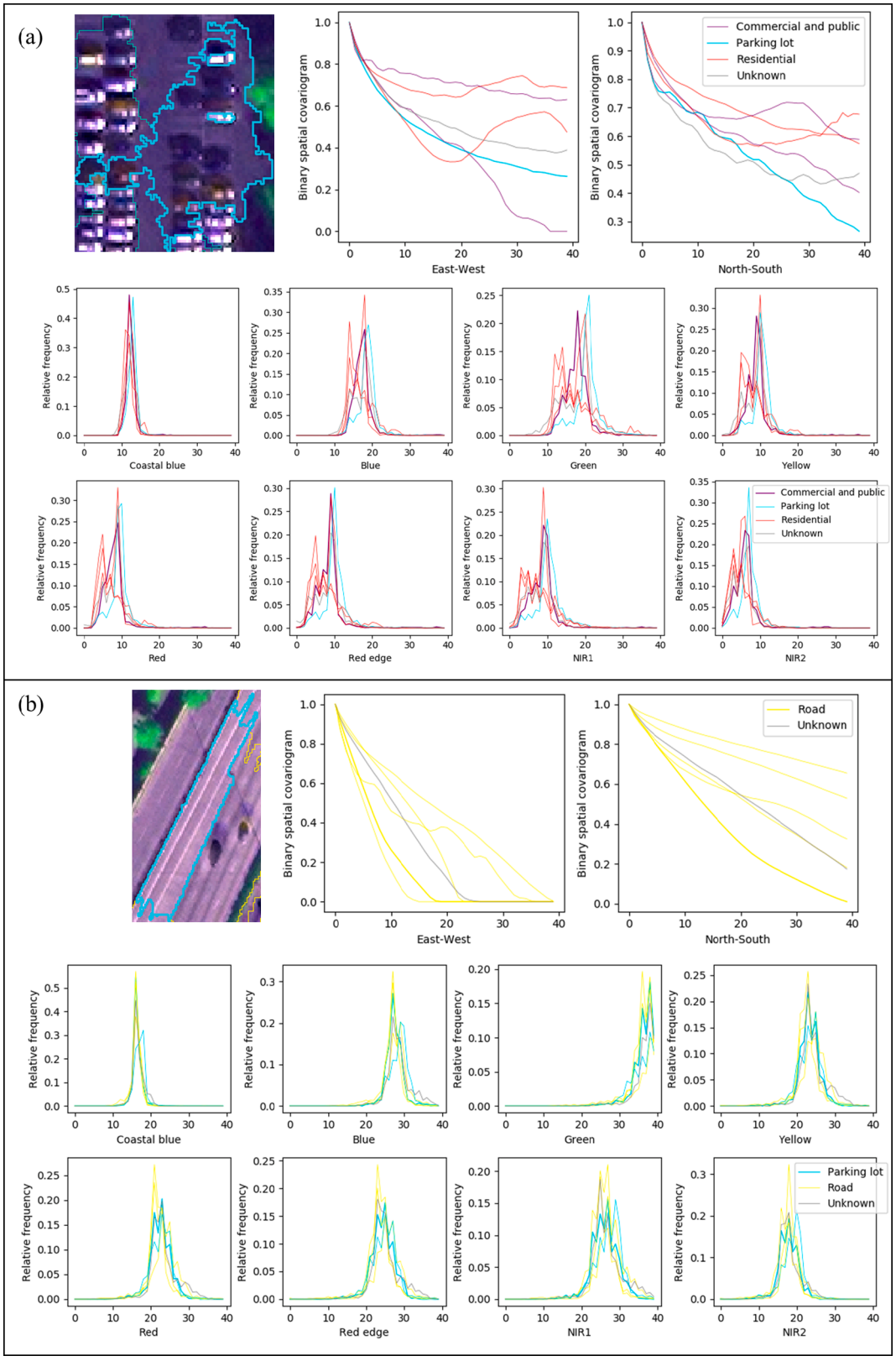

Fig. 3. Two examples of classification using curve matching approaches. The curves of spatial covariograms and curves of histograms, and the corresponding five most similar sample curves for (a) a parking lot object, (b) a road object.

smallest divergences for both the spectral and the spatial features are shown in the figure. This object is misclassified as the commercial and public class in the His model. However, the directional spatial covariograms show that the sample curve is most similar to the parking lot class, which is especially obvious in the east–west direction. Consequently, combining the spectral and spatial features in the His-Cov model correctly classifies this object to the parking lot class. The second illustrative object belongs to the road class (Fig. 3b). The curves of the spectral histogram show that the five curves with the smallest divergence from the object to be classified belong to either the road or parking lot classes. The curve with the minimum divergence belongs to a sample of the parking lot class, as shown most obviously in the red, red edge, and two near infrared (NIR) bands. Consequently, the object is misclassified as the parking lot class using the His model. However, the five sample curves of the spatial covariograms with the smallest divergence all belong to the road class, which easily corrects this object to the road class in the His-Cov model. These examples demonstrate how the spatial covariogram can help improve the classification of objects having similar spectral features.

3. Experiments

3.1. Datasets and study areas

Test were conducted on three differing HSR images with the aim of classifying complex land cover and land use classes in urbanized areas. Typical land cover classes such as water and vegetation were excluded because spectral signatures based on curve matching classifiers have been successfully demonstrated for these classes (Stow et al., 2012; Sridharan and Qiu, 2013). Although there are some trees of several different species scattered in the study area, identification of vegetation types was also excluded. Our focus is on classes having similar spectral responses and therefore easily misclassified, such as building and road, or more complex function zones such as residential, commercial and public.

The three images were acquired from different sources: IKONOS, WorldView-2 and WorldView-3. The spatial resolution of the multispectral bands for all three images was improved using the Gram–Schmidt pansharpening method. In the IKONOS image with a 1 m spatial resolution, three simple land cover classes can be visually distinguished, including building, road, and barren land. These three classes all include objects with a concreate surface, causing the classification difficult. The WorldView-2 and WorldView-3 images are with a sub-meter spatial resolution, five more detailed land use classes can be visually identified, including residential, commercial and public, road, parking lot, and barren land. The residential and the commercial and public areas in WorldView-2 are located separately in most places, therefore it is easy to visually differentiate the residential from the commercial and public classes. However, these two classes in WorldView-3 image are geographically mixed up and their spectral reflectance is more similar. Thus, the identification of objects of the residential, commercial and public classes was aided by referencing Google Earth and OpenStreetMap for the WorldView-3 image. Additionally, the parking lot and the road classes are also difficult to be classified in the WorldView-2 and WorldView-3 images, since both classes include cars and parking lines or crosswalk. The information of three study areas and the characteristics of classes are list in Table 1.

The three images and samples are shown in Fig. 4. Table 2 provides information on the samples used in the study. The samples are used for either training or testing purposes. A 33.3% training versus 66.7% testing sample split was used, selected randomly from the sample dataset to alleviate the influence of spatial autocorrelation. Since sampling has a great effect on classification accuracies (Ye et al., 2018; Li et al., 2018), all classifications were performed ten times based on different training and testing samples selected randomly and independently to ensure a robust test.

3.2. Segmentation and classification

Image segmentation based on the SLIC and FNEA were applied in this study. The quality of image segmentation is important to the outcome of any object-based classification (Liu et al., 2012). A supervised method namely adjusted rand index (ARI) (Hubert and Arabie, 1985) was used for evaluating the segmentation quality on samples. The ARI is built on counting pairs of items in which two segments agree or disagree. A large ARI value indicates a high correspondence of the segmentation result to the reference, with an upper bound of 1.

For the three study areas, the six curve matching approaches (KL, KS, CAM, CCAM, RSSDA and CRSSDA) introduced previously were used. Each approach was applied with spectral features only (i.e., the His model) as well as with both spectral and spatial features (i.e., the His-Cov model), resulting in twelve experiments on each of the images. For further comparison, four state-of-the-art classification methods based on machine learning algorithms were applied: (1) support vector machine (SVM), (2) random forest (RF) classification, (3) extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016), and (4) structural feature set (SFS) (Huang et al., 2007).

Among these four methods, SVM and RF have been used commonly in object-based classification. XGBoost, a variant of the gradient boosting machine, is an ensemble of classification and regression tree. Twenty-seven object-level statistical summaries were used in these three methods, including seven spectral features, six gray-level co-occurrence matrix (GLCM) texture features, and fourteen semivariogram features, as listed in Table 3. The spectral features are calculated for each

Table 1
The information of three study areas and characteristics of classes.

Image	IKONOS	WorldView-2	WorldView-3
Location	Beijing	Dallas	Beijing
Coordinate	39°57′57″– 39°58′30″N, 116°23′19″– 116°24′46″E	32°50′18″– 32°51′21″N, 96°45′49″– 96°47′8″W	39°59′52″–40°0′31″N, 116°23′37″–116°24′7″E
Data product	4-band bundle, S2A	8-band bundle, S2A	4-band bundle, S2A
Spatial resolution	1 m	0.5 m	0.4 m
Image size	2048 × 1024	4000 × 4000	1760 × 3035
Acquisition time	May 2000	January 2010	September 2014
Classes and characteristics	Barren land: includes bare soil and playground.	Barren land: includes soil land, playground and under-developed area.	Barren land: mainly playground.
	Road: main roads with cars inside.	Road: main roads with crosswalk and cars inside and some narrow roads.	Road: main roads with crosswalk and cars inside.
	Building: a mix of residential, commercial and public.	Commercial and public: includes shopping centers, schools, factories, etc.	Commercial and public: includes shopping centers, schools, office buildings, etc.
		Residential: buildings with smaller size than commercial and public.	Residential: buildings are of a similar size to commercial and public, but away from main roads.
		Parking lot: with parking lines and cars inside.	Parking lot: with parking lines and cars inside.

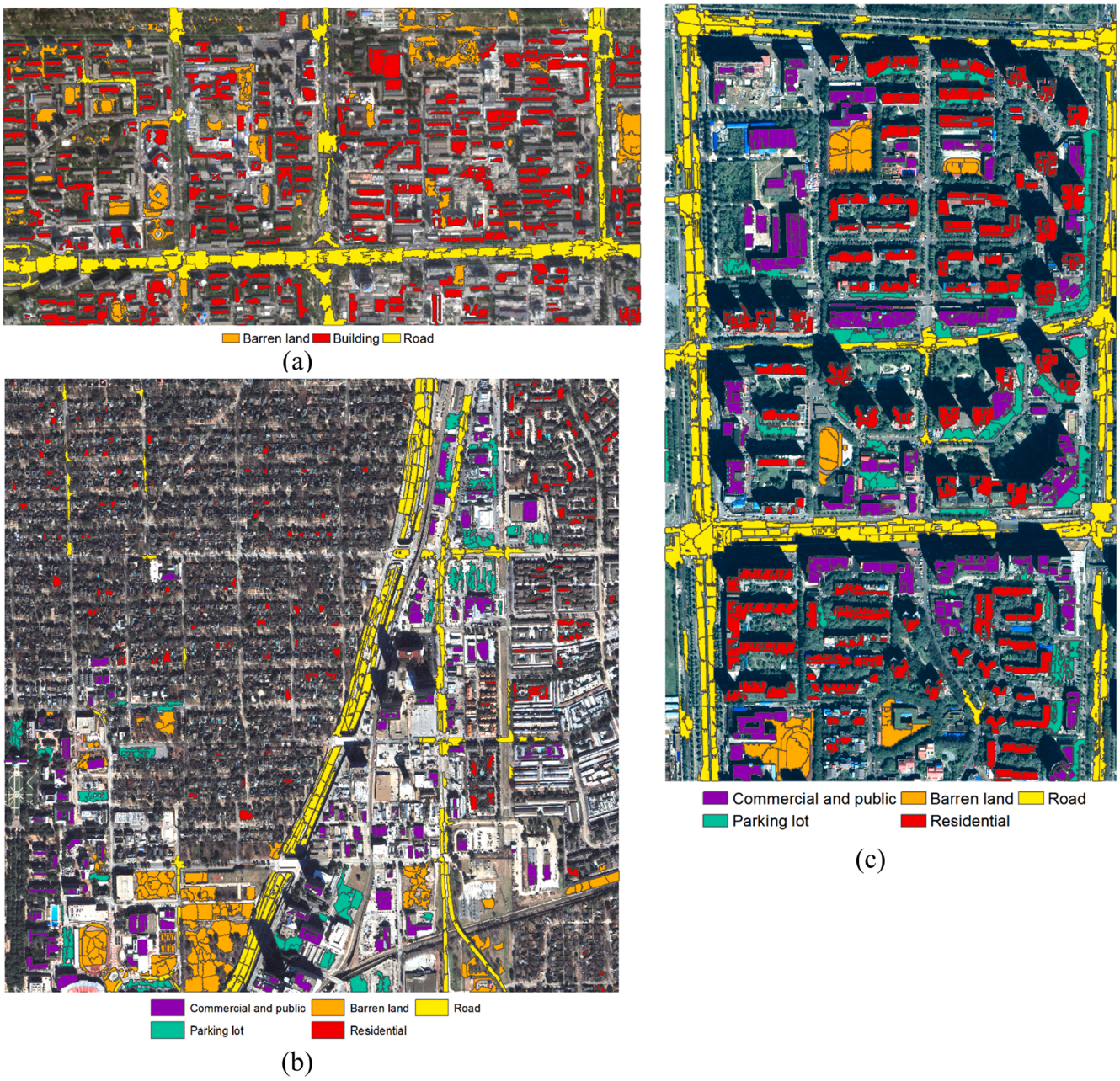

Fig. 4. Three images in true color composite used for the experiments: (a) IKONOS image and samples, (b) WorldView-2 image and samples, (c) WorldView-3 image and samples.

Table 2
The samples per class for the study areas.

IKONOS			WorldView-2			WorldView-3		
Class	# of objects	# of pixels	Class	# of objects	# of pixels	Class	# of objects	# of pixels
Barren land	102	95,791	Commercial and public	139	340,306	Commercial and public	148	270,514
Building	324	245,923	Barren land	126	433,364	Barren land	34	101,805
Road	74	142,428	Road	188	579,748	Road	199	514,983
			Parking lot	93	265,134	Parking lot	82	168,140
			Residential	151	192,990	Residential	207	405,180
Total	**500**	**484,142**	**Total**	**697**	**1,811,542**	**Total**	**670**	**1,460,622**

multispectral band of each object. One GLCM is computed for each object to describe the co-occurrences of the pixels that are separated at one-pixel lag distance interval inside the object. The average value in all four directions on all multispectral bands is derived (Haralick, 1979; Hall-Beyer, 2007). Similarly, the semivariogram features are extracted for each object, also based on the average value for all four directions and all multispectral bands. The equations for these semivariogram features can be found in Yue et al. (2013) and Wu et al. (2015). The SFS

Table 3
The object-level features used in the SVM, RF and XGBoost methods.

Spectral features	Semivariogram features
Mean	Ratio between total variance and first semivariance (RVF)
Median	Ratio between the first and the second semivariance (RSF)
Standard deviation	First derivative near the origin (FDO)
Skewness	Second derivative at third lag (SDT)
Kurtosis	First maximum lag value (FML)
First quartile	Mean of the semivariogram values up to the first maximum (MFM)
Third quartile	Variance of the semivariogram values up to the first maximum (VFM)
GLCM texture features	Difference between MFM and the first semivariance (DMF)
Contrast	Ratio between the first local maximum semivariance and MFM (RMM)
Homogeneity	Second-order difference between first lag and first maximum (SDF)
Dissimilarity	Area until the first maximum (AFM)
Energy	Distance between the first and the second local maxima (DMS)
Correlation	Distance between the first local maximum and the first local minimum (DMM)
Angular second moment (ASM)	Hole area (HA)

method uses statistical summaries of the histogram along four directions in the objects. These statistical features include length, width, pixel shape index, weighted mean, ratio, and standard deviation. Compared with the His model, the SFS method uses only six features based on summaries of the histogram of the object in four directions, rather than the complete histogram of the object itself. Also, the SFS is based on the SVM classifier and the His model is based on the curve matching approach. For the SVM-based classification, a non-linear radial basis kernel was chosen. A 10-fold cross-validation was applied to select optimal parameters for the penalization constant and the kernel parameter. For the RF classification, the number of trees was set to 10, and the supported criteria were based on the Gini impurity. For the XGBoost method, a tree-based booster was used, and the maximum depth of tree was set to 6.

The weight for the spectral feature w in Eq. (8) was set automatically from 0 to 1 with an increment of 0.01 using a k-fold cross-validation scheme at each performance, with k set at 5. The selection was repeated 50 times, and the value with the majority of votes from these 50 repetitions was selected. The number of histogram bins was set to 100. The number of spatial lags for the spatial covariaogram was set to 50 pixels, derived from the average dimension of the objects.

4. Results and analysis

4.1. Segmentation quality

To assess segmentation quality, Table 4 shows the overall ARI of segmentation results and the ARI for each class. The reference segments are delineated manually. As can be seen, the overall ARI ranges from 0.88 to 0.90 for three study areas. The relatively high accuracy of image segmentation ensures the quality of subsequent image classification. For specific classes, the segmentation has a good performance on the small objects such as the residential and the commercial and public classes. The building class in the IKONOS image, the commercial and public class and the residential class in the WorldView-2 and WorldView-3 images, dominated by the FNEA segmentation result, show high ARIs (above 0.90). However, large objects such as the barren land, the road, and the parking lot classes have low ARIs (between 0.65 and 0.87). These three classes are usually of large size and therefore are over-segmented by the SLIC segmentation. For example, a good segmentation of a road object is expected to result in a long, narrow shape. Many small and compact objects segmented by SLIC for the road class are actually more suited to a curve matching based classification. This is because roads segmented as one complete, long object have a greater chance of being misclassified by curve matching approaches, since the spatial dimensions of the roads could be very different in size and thus may greatly affect their effective ranges. These large objects are over-segmented intentionally because over-segmentation is beneficial to the His-Cov model. The final classification map will be formed when the adjacent small segments of the same classes are recombined into a large object.

Table 4
The ARI of segmentation results for three study areas.

IKONOS		WorldView-2		WorldView-3	
Class	ARI	Class	ARI	Class	ARI
Barren land	0.86	Commercial and public	0.91	Commercial and public	0.92
Building	0.91	Barren land	0.87	Barren land	0.65
Road	0.82	Road	0.84	Road	0.82
		Parking lot	0.82	Parking lot	0.81
		Residential	0.92	Residential	0.90
Overall	**0.90**	**Overall**	**0.89**	**Overall**	**0.88**

4.2. Classification results

There are sixteen classification methods (twelve curve matching approaches and four machine learning algorithms) for each study area, and each method is performed ten times using different samples. One classification result randomly selected from the ten performances is shown in Figs. 5–7. Among these, the results using the four machine learning methods are all presented. For the curve matching approaches, only the KL and CRSSDA are displayed, with His and His-Cov models, respectively, since these two approaches have the best performance, as discussed below. All the classified objects are delineated by polygons with the color of their boundary corresponding to their reference class label, and the misclassified objects are filled with the color corresponding to the misclassified class label. To provide a quantitative assessment, the classification accuracy was calculated. The overall accuracy based on the average value from the ten performances is shown in Table 5. The standard deviations of accuracies vary from 0.1 to 0.3 for all the methods, thus are not shown in the table. The overall classification accuracy from one sampling (correspond to Figs. 5–7) and F1-score of each class are shown in Tables 6 and 7.

In the classification results for the IKONOS image (Fig. 5), most small buildings with a regular shape are correctly classified. Larger buildings, however, may be misclassified as road objects, as in the SVM, XGBoost and CRSSDA results (Fig. 5a, c and g). The misclassification of the road class as buildings exists in all results, although worse in some than others. The SFS result (Fig. 5d) includes many barren land objects, most of them misclassified. This class has the smallest sample size, which suggests that the SFS method does not properly consider the unbalanced sample size. Barren land with a soil surface is easier to identify than that with a concrete surface (e.g., a playground), and the latter is sometimes confused with the smaller building such as in the RF, XGBoost, SFS and KL results (Fig. 5b-e). The His-Cov model correctly classified the road objects in the north–south direction (in the right portion of the image in Fig. 5f and h), which were misclassified by the His model based on the KL and CRSSDA approaches (Fig. 5e and g).

More complex classes are involved in the WorldView-2 image. As can be seen from classification results in Fig. 6, the residential class usually has small sized, dark colored roofs, whereas most commercial and public class objects have large sized, light colored roofs. The parking lot class and the road class are easily confused with each other since both tend to be affected by vehicles within the objects. The SVM, KL and CRSSDA approaches (Fig. 6a, e and g) misclassify roads as parking lots, whereas

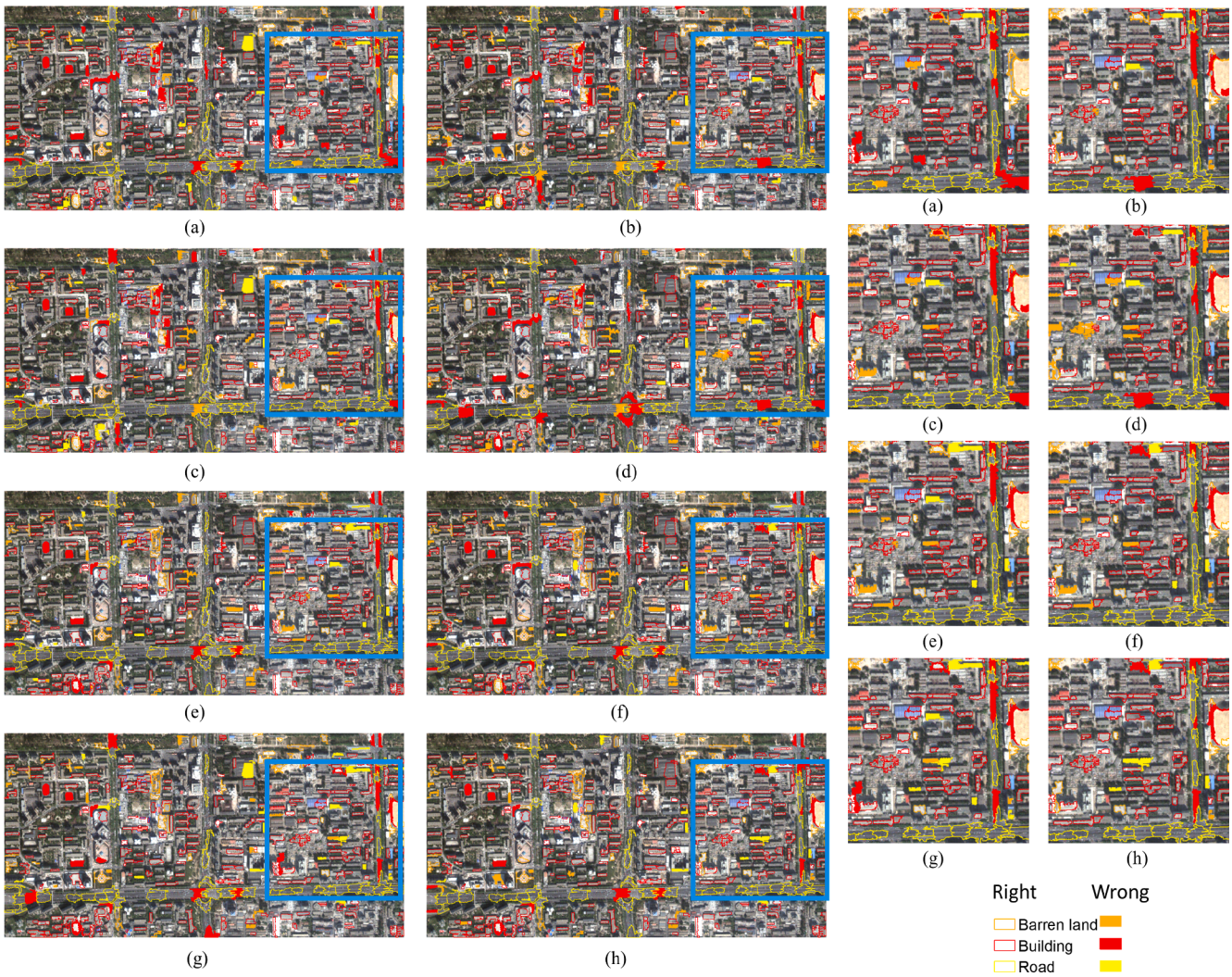

Fig. 5. Classification results using different methods for the IKONOS image (left) and their zoomed view for the subset (right): (a) SVM, (b) RF, (c) XGBoost, (d) SFS, (e) KL with His, (f) KL with His-Cov, (g) CRSSDA with His, and (h) CRSSDA with His-Cov.

the His-Cov model was able to correct some of these misclassifications with the KL and CRSSDA approaches (Fig. 6f and h). The RF and SFS methods (Fig. 6b and d) misclassified some commercial and public objects as parking lots, while the road class is mostly correctly classified by SFS (Fig. 6d).

The residential class and the commercial and public class are more easily confused with each other in the WorldView-3 image than in the WorldView-2 image because their spectral reflectance is more similar. All methods suffer from the misclassification of these two classes, although four of the curve matching approaches (Fig. 7e–h) perform slightly better than the machine learning algorithms (Fig. 7a–d) in the bottom right part of the image. Some residential class and commercial and public class objects are misclassified to the road class in the SVM and SFS results (Fig. 7a and d). Many residential objects are misclassified as parking lots in the KL and CRSSDA results (Fig. 7e and g), although the His-Cov model corrects some with its additional consideration of spatial distribution. However, the road objects in the east–west direction in the lower middle of the image are severely misclassified, and even the His-Cov model can only correct a few of them.

As revealed by the average classification accuracy in Table 5, the six statistical summaries of the histogram in the SFS method do not fully utilize the rich information within the objects, resulting in the lowest accuracy. The SVM method has the worst performance among the three methods based on the combination of spectral, GLCM and semivariogram features. The XGBoost method has the highest accuracy among the four machine learning methods. It is the only method able to achieve results comparable to the curve matching approaches based on the His-Cov model. The His-Cov model consistently produces higher accuracies than the His model for all six curve matching approaches. In fact, not only is the average accuracy for the His-Cov model higher, but also the accuracy for all the performances is highest among all the ten samplings. This result demonstrates that the spatial covariogram is a powerful object-based feature that can capture spatial distributions within objects and thus enhance curve matching approaches to classification based on spectral features only.

From the F1-score in Tables 6–8, the barren land class in the IKONOS image, the commercial and public and the barren land classes in the WorldView-2 image, and the barren land, the residential and the road classes in the WorldView-3 image have the greatest improvements. The other classes do not constantly result in the highest accuracy using the His-Cov model, but have higher accuracy than the machine learning methods and the His model for most curve matching approaches. The machine learning methods show a few exceptions: the road class in the IKONOS image using the XGBoost method, the parking lot and the road classes in the WorldView-2 image using the XGBoost, and the parking lot class in the WorldView-3 image using the SFS method achieve the accuracy that is higher than or comparable with the His-Cov model. The accuracies of the His model are generally higher than the machine

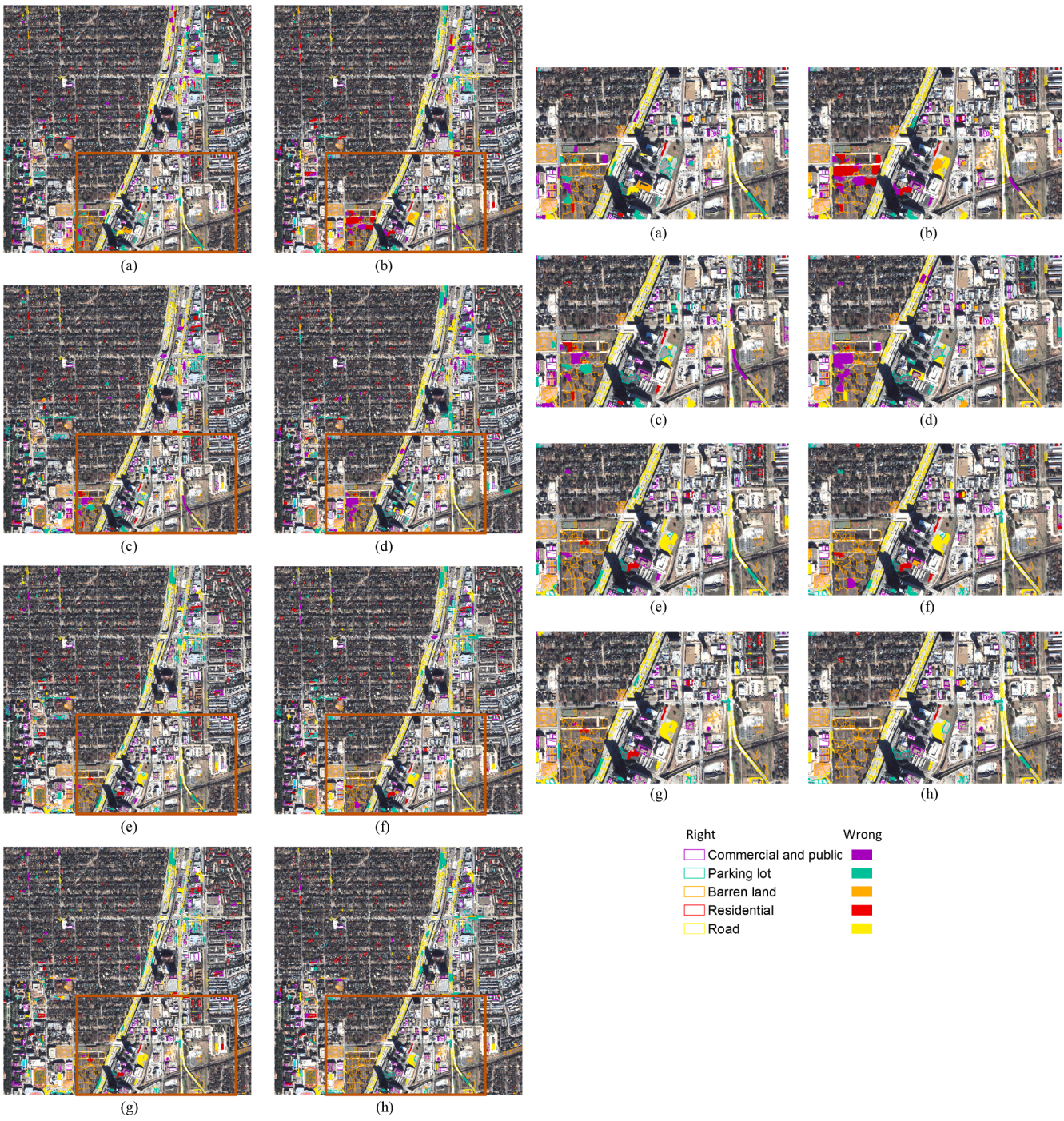

Fig. 6. Classification results using different methods for the WorldView-2 image (left) and their zoomed view for the subset (right): (a) SVM, (b) RF, (c) XGBoost, (d) SFS, (e) KL with His, (f) KL with His-Cov, (g) CRSSDA with His, and (h) CRSSDA with His-Cov.

learning methods and lower than the His-cov model, except for the parking lot class in the WorldView-2 and WorldView-3 images, where the machine learning methods perform better than the His model. The F1-score accuracy shows that the classification of spectral similar classes (e.g., barren land, road, parking lot) can be improved by spatial features, either from the semivariogram features or spatial covariogram. In general, the classes show the greatest improvements when the spatial covariogram is introduced by the His-Cov model most of the time.

4.3. His-Cov model on each class

It is of interest to further explore how the His-Cov model performs for each class compared with the His model. For the six curve matching approaches, changes in numbers and accuracies (at the object level) in each class for the His-Cov model relative to the His model are shown in Fig. 8 for one of the ten performances. This allows approaches based on the His-Cov model to be compared with those based on the His model. Positive values for the object number or accuracy indicate an increase from applying the His-Cov model, whereas negative values indicate a decrease.

For most curve matching approaches, the number of buildings increases most among all the classes in the IKONOS image, and the number of roads decreases after applying the His-Cov model (Fig. 8a). All methods have accuracy increases for building and road classes

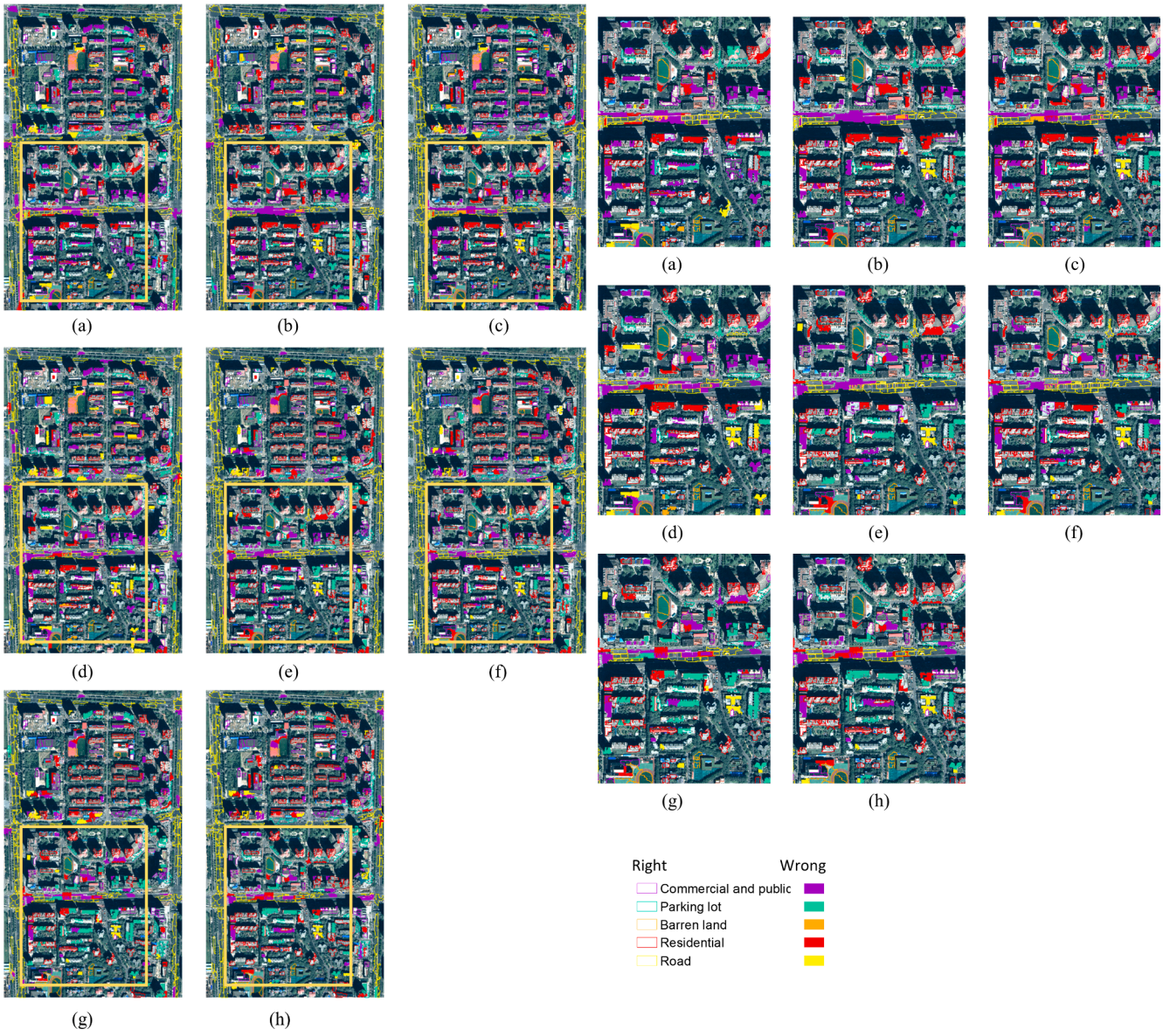

Fig. 7. Classification results using different methods for the WorldView-3 (left) and their zoomed view for the subset (right): (a) SVM, (b) RF, (c) XGBoost, (d) SFS, (e) KL with His, (f) KL with His-Cov, (g) CRSSDA with His, and (h) CRSSDA with His-Cov.

Table 5
The average of the overall accuracy (%) of the classification results over ten samplings. The highest accuracies among all methods are marked in bold.

Type	Method	IKONOS	WorldView-2	WorldView-3
Machine learning algorithms	SVM	79.37	73.61	63.36
	RF	81.14	71.85	66.47
	XGBoost	82.31	76.80	67.99
	SFS	76.88	72.39	62.86
Curve matching approaches based on His model	KL	81.83	74.64	68.57
	KS	81.14	73.31	64.41
	CAM	81.26	73.81	65.37
	CCAM	81.74	73.85	64.65
	RSSDA	80.78	73.76	64.27
	CRSSDA	81.74	74.26	64.99
Curve matching approaches based on His-Cov model	KL	82.88	**77.05**	**70.36**
	KS	82.76	75.42	67.38
	CAM	82.85	76.21	69.37
	CCAM	83.42	75.51	67.09
	RSSDA	81.98	76.51	68.03
	CRSSDA	**83.54**	75.83	67.38

Table 6
The overall accuracy (%) and the F1-score per class of classification results from one sampling for the IKONOS image.

Type	Method	Total	Barren land	Building	Road
Machine learning algorithms	SVM	82.28	0.72	0.88	0.67
	RF	80.78	0.64	0.87	0.70
	XGBoost	83.78	0.69	0.89	0.78
	SFS	73.87	0.61	0.83	0.51
Curve matching approaches based on His model	KL	82.58	0.75	0.88	0.70
	KS	81.98	0.72	0.87	0.72
	CAM	81.38	0.74	0.87	0.68
	CCAM	83.78	0.75	0.89	0.73
	RSSDA	79.88	0.70	0.85	0.70
	CRSSDA	83.78	0.75	0.89	0.73
Curve matching approaches based on His-Cov model	KL	83.78	0.72	0.89	0.76
	KS	84.08	0.77	0.89	0.71
	CAM	83.18	0.77	0.88	0.70
	CCAM	86.19	0.81	0.91	0.73
	RSSDA	81.08	0.70	0.87	0.72
	CRSSDA	86.49	0.81	0.91	0.73

Table 7
The overall accuracy (%) and the F1-score per class of classification results from one sampling for the WorldView-2 image.

Type	Method	Total	Commercial and public	Parking lot	Barren land	Residential	Road
Machine learning algorithms	SVM	72.69	0.64	0.51	0.85	0.77	0.78
	RF	69.03	0.59	0.45	0.75	0.72	0.80
	XGBoost	75.70	0.62	0.62	0.81	0.81	0.83
	SFS	72.47	0.63	0.52	0.80	0.78	0.80
Curve matching approaches based on His model	KL	74.84	0.74	0.53	0.91	0.76	0.75
	KS	74.62	0.73	0.47	0.91	0.72	0.80
	CAM	72.47	0.72	0.44	0.89	0.68	0.77
	CCAM	73.12	0.71	0.45	0.89	0.74	0.77
	RSSDA	74.41	0.76	0.48	0.89	0.72	0.78
	CRSSDA	73.55	0.71	0.47	0.89	0.75	0.76
Curve matching approaches based on His-Cov model	KL	78.28	0.75	0.62	0.89	0.78	0.81
	KS	77.63	0.76	0.53	0.90	0.79	0.81
	CAM	77.85	0.74	0.55	0.89	0.80	0.81
	CCAM	78.28	0.74	0.57	0.89	0.82	0.80
	RSSDA	78.71	0.76	0.52	0.92	0.81	0.83
	CRSSDA	78.49	0.75	0.55	0.90	0.83	0.80

Table 8
The overall accuracy (%) and the F1-score per class of classification results from one sampling for the WorldView-3 image.

Type	Method	Total	Commercial and public	Parking lot	Barren land	Residential	Road
Machine learning algorithms	SVM	60.85	0.50	0.64	0.67	0.47	0.82
	RF	66.00	0.48	0.69	0.71	0.62	0.83
	XGBoost	65.55	0.45	0.70	0.65	0.59	0.86
	SFS	60.85	0.44	0.75	0.60	0.54	0.76
Curve matching approaches based on His model	KL	67.56	0.46	0.56	0.81	0.65	0.88
	KS	65.77	0.48	0.46	0.86	0.64	0.87
	CAM	67.79	0.49	0.51	0.83	0.65	0.89
	CCAM	65.77	0.48	0.50	0.86	0.65	0.85
	RSSDA	67.11	0.51	0.47	0.83	0.64	0.89
	CRSSDA	65.77	0.49	0.51	0.86	0.65	0.84
Curve matching approaches based on His-Cov model	KL	72.26	0.49	0.72	0.81	0.69	0.90
	KS	69.57	0.49	0.59	0.86	0.69	0.87
	CAM	72.26	0.50	0.70	0.83	0.70	0.89
	CCAM	69.80	0.48	0.67	0.86	0.68	0.85
	RSSDA	70.69	0.53	0.59	0.86	0.68	0.89
	CRSSDA	70.25	0.50	0.66	0.86	0.69	0.86

(Fig. 8b), with the latter more obvious than the former. This indicates that most misclassification of road as building using the His model is corrected by applying the His-Cov model. For the WorldView-2 image in Fig. 8c, numbers in the commercial and public class and the parking lot class decrease whereas numbers in the residential class and the road class increase when introducing the His-Cov model. The accuracies in Fig. 8d shows that the parking lot, residential and road classes benefit most from the His-Cov model. A similar situation is also observed in the WorldView-3 image (Fig. 8e and f). Reductions in the parking lot class lead to the greatest improvement in accuracy. The increase in residential class numbers also generates an improvement in accuracy, but not as much as in the WorldView-2 result. The number and accuracy of the road and the barren land classes do not show an obvious change for most methods. The numbers for the commercial and public class vary in both the WorldView-2 and WorldView-3 images, but accuracy improvement is limited or even decreases slightly. This suggests that the commercial and public class is the most difficult to identify. In general, objects with a distinctive spatial distribution, such as parking lots with vehicles and parking lines, are more recognizable using the His-Cov model, whereas the objects without a distinctive spatial distribution inside, such as the barren land and the commercial and public classes, are difficult to identify even with the His-Cov model. Overall in an urban environment, the spatial covariogram has a positive impact on typical land use classes most of the time, with residential, road and parking lot classes benefitting the most from applying the His-Cov model.

4.4. Computational cost

The total computational cost of each classification method is compared in Table 9. All computations used Python on a laptop with 1.8-GHz Intel Core-i7 CPU and 16-GB memory.

The most efficient methods are curve matching approaches based on the His model since only the histogram-based spectral features are extracted. For this model, feature extraction accounts for 37% to 77% of the time used, implying classification takes more than half of the computation time for some classifiers (KL, CAM and CCAM). The SFS method is based on the SVM classifier and almost all time is spent on feature extraction. It extracts seven object-level spectral features and six spatial features from directional histograms, without calculating semivariogram features, less efficient than the curve matching approaches based on the His model.

The machine learning algorithms involving semivariogram (SVM, RF and XGBoost) are markedly more computational intense. The feature extraction for these machine learning algorithms and the curve matching approaches based on the His-Cov model accounts for more than 98% of the total computational cost. The total cost for these machine learning algorithms compared with the curve matching approaches based on the His-cov model is about 3.5, 13.7 and 8.7 times longer for the IKONOS, WorldView-2 and WorldView-3 images, respectively. The main reason is that the semivariogram is calculated based on continuous variables (i.e., pixel values) for each multi-spectral band, whereas the spatial covariogram is estimated based on binary variables (i.e., Otsu thresholding results) for only one band (i.e., the first PC). Therefore, the spatial features extracting from the binary spatial covariogram is more efficient

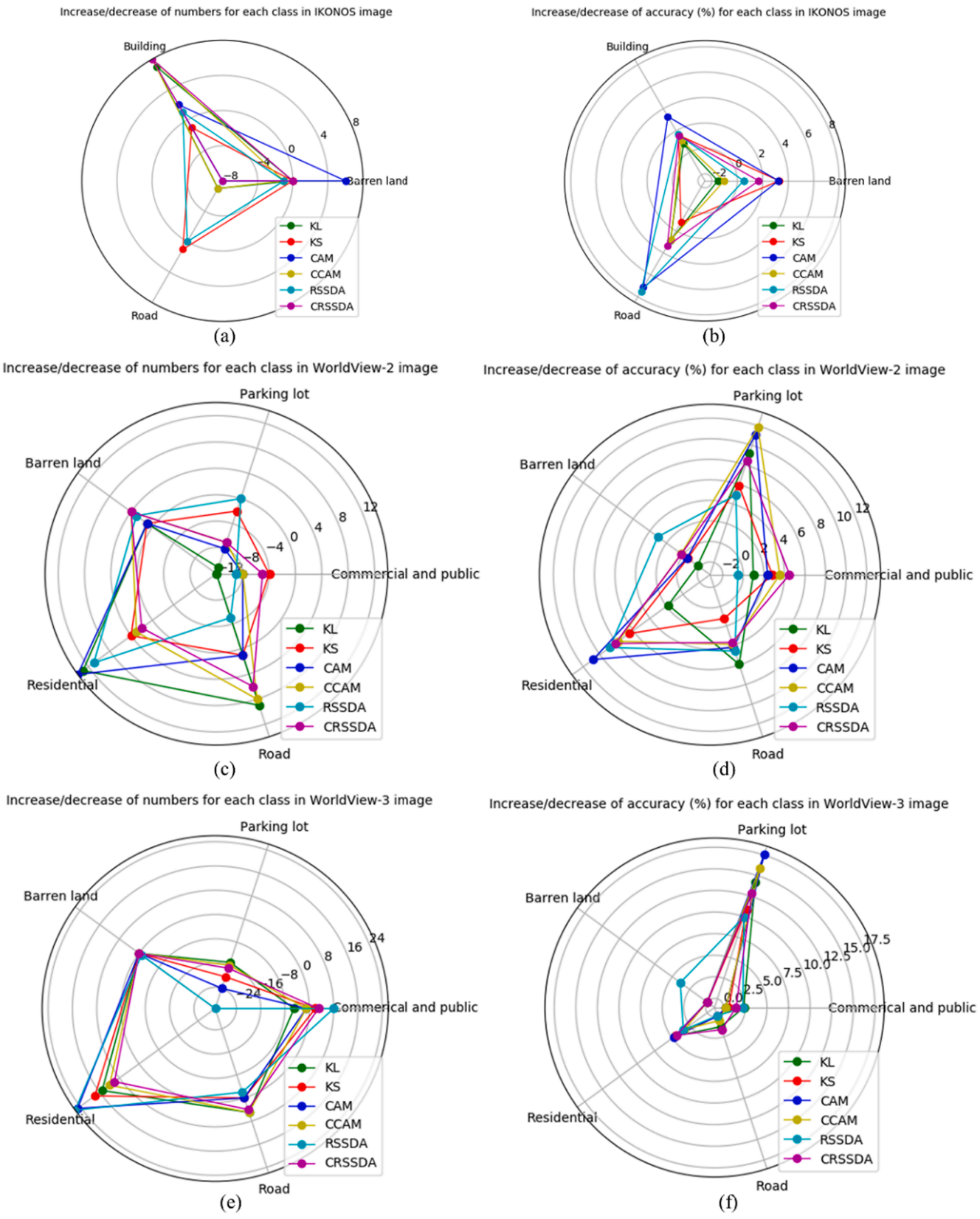

Fig. 8. Comparison of the six curving matching approaches showing increase/decrease of object numbers and accuracy for each class based on the His-Cov model relative to the His model.

Table 9
Computational cost of all classification methods (unit: second), and the time for feature extraction is reported as the percentage of the total time in parentheses.

Type	Method	IKONOS	WorldView-2	WorldView-3
Machine learning algorithms	SVM	1292.3 (100%)	19916.3 (100%)	9079.5 (100%)
	RF	1292.2 (100%)	19916.2 (100%)	9078.9 (100%)
	XGBoost	1292.4 (100%)	19917.0 (100%)	9080.1 (100%)
	SFS	223.4 (100%)	180.7 (100%)	270.0 (100%)
Curve matching approaches based on His model	KL	20.0 (47%)	95.5 (45%)	107.7 (44%)
	KS	13.0 (72%)	59.7 (72%)	62.0 (76%)
	CAM	23.9 (39%)	114.2 (38%)	112.3 (42%)
	CCAM	24.6 (38%)	115.8 (37%)	94.6 (50%)
	RSSDA	14.5 (64%)	68.0 (64%)	61.3 (77%)
	CRSSDA	14.6 (64%)	69.7 (62%)	61.5 (77%)
Curve matching approaches based on His-Cov model	KL	289.8 (99%)	1452.7 (99%)	1045.7 (99%)
	KS	287.7 (99%)	1447.3 (100%)	1039.4 (99%)
	CAM	290.7 (98%)	1454.9 (99%)	1048.3 (99%)
	CCAM	291.2 (98%)	1455.3 (99%)	1049.0 (99%)
	RSSDA	288.2 (99%)	1448.6 (100%)	1040.7 (99%)
	CRSSDA	288.5 (99%)	1449.0 (100%)	1040.9 (99%)

than that from the traditional semivariogram.

4.5. Parameter setting

The new parameter introduced in the His-Cov model is the weight w (proportion of influence for the spectral feature), which decides the balance between spectral and spatial features. This parameter was set automatically using a k-fold cross-validation, where k equals 5. The selection was repeated 50 times, and the value with the majority of votes was selected. In order to check the influence of the different weights, Fig. 9 displays a sensitivity analysis of the parameter for one performance. The weights vary from 0 to 1 with an increment of 0.01 for all six curve matching approaches. The accuracies of corresponding approaches based on the His model are shown with horizontal dashed lines. The optimal weight with the maximum accuracy is marked by a cross, and the weight selected by cross-validation is marked by a dot in the figure. Generally, an appropriate weight selection is dependent on the curve matching approach. Most approaches prefer a large weight for spectral features. For the CAM and RSSDA approaches, accuracy improves more rapidly when the spatial distribution is introduced, so median to large values of weights can all lead to a higher accuracy for the His-Cov model. The two cumulative approaches, CCAM and CRSSDA, have slower improvements in accuracy when incorporating spatial distribution, leading to a large optimal weight when maximum accuracy is achieved. The KL approach, although not based on cumulative curves, also requires a large weight (around 0.9), primarily because its accuracy for the His model is already relatively high among all six approaches. The KS approach also prefers a large weight, but the increase in accuracy is faster than the other two cumulative approaches (CCAM and CRSSDA). All the selected weights are close to or even equal their optimal weights. This sensitivity analysis demonstrates that the His-Cov model with appropriate weights improves the accuracy of all curve matching approaches relative to the His model, and provides guidance on setting optimal weights for the different curve matching approaches.

The number of lags was set to 50 according to the average dimension of objects in this study. When the dimension of an object is less than 50 pixels, the spatial covariogram reaches 0 before the lag distance reaches 50. The actual range when the spatial covariance reaches 0 is called the effective range. One concern is whether different effective ranges between samples affects the classification results. Commonly, the spatial covariogram decreases sharply as the distance lag increases. Therefore, if a sample has a large effective range and another a small one, the difference between two spatial covariogram curves at a large lag distance is usually small. In this study, we adopted two measures to prevent a possible spatial covariogram mismatch due to different effective ranges. First, ten trials with different samples ensure that effective ranges of the spatial covariogram for the training and testing samples would be balanced in general. For example, the testing object in Fig. 3b has a small effective range in the east–west direction. Similarly, some training objects of the road class also have small effective ranges in the east–west direction. These should match well with the testing object. Second, as previously mentioned, SLIC segmentation was performed to ensure that an object will not have an overly long shape and thus the effective range will not be too large.

The foreground and background of the spatial configuration from the binary map was self-defined, with the foreground assigned to pixels with values less than or equal to a threshold. Technically, the foreground and background can be switched. We chose the dark pixels as foreground because the light pixels are usually a minority in an object, such as vehicles in a road object and parking lines in a parking lot object. If the light pixels are used as foreground, some objects belonging to the same class may have very different spatial distributions.

5. Discussion

The developed His-Cov model has several advantages over other traditional OBIA classification. First, the His-Cov model is able to incorporate rich spectral and spatial features in the form of curves without conducting feature selection or optimization, which is usually required in a traditional OBIA classification. Second, the His-Cov model performs well on classifying spectral similar classes, which are difficult to be distinguished by histogram-based features in the curve matching approaches. The performance of the His-Cov model is generally superior to traditional classifiers based on the other spatial features (e.g., semivariogram) in classifying complex land cover classes such as the residential and the commercial and public classes, and the road and the parking lot classes. Finally, the spatial feature extraction from spatial covariogram is easier than from (semi)variogram. Because of the extensive computation load, a common way to calculate the (semi) variogram is through modeling from a few samples. Setting appropriate parameters for the (semi)variogram model, such as sill and range, can be subjective and problematic. The His-Cov model, on the other hand, estimates spatial covariogram from all the pixels in the object and avoids the involvement of expert knowledge in determining initial parameters during the modeling process. The His-Cov model uses binary variables to reflect the general spatial configuration of the pixels. The binary variables are less sensitive to noise, makes the spatial distribution information extracted more robust and less computationally expensive. Additionally, (hyper-)parameters are not required in the curve matching approaches, thus parameter setting for the His-Cov model based classification is easier than some traditional classifiers such as SVM.

The His-Cov model developed here integrates spectral and spatial features by treating them all as curves. Instead of using statistical summaries of the objects' features, all the pixels of an object are used to reflect as much as possible of their spectral variability and spatial distribution. By incorporating spatial distribution, the His-Cov model improves the performance of the curve matching approaches based on

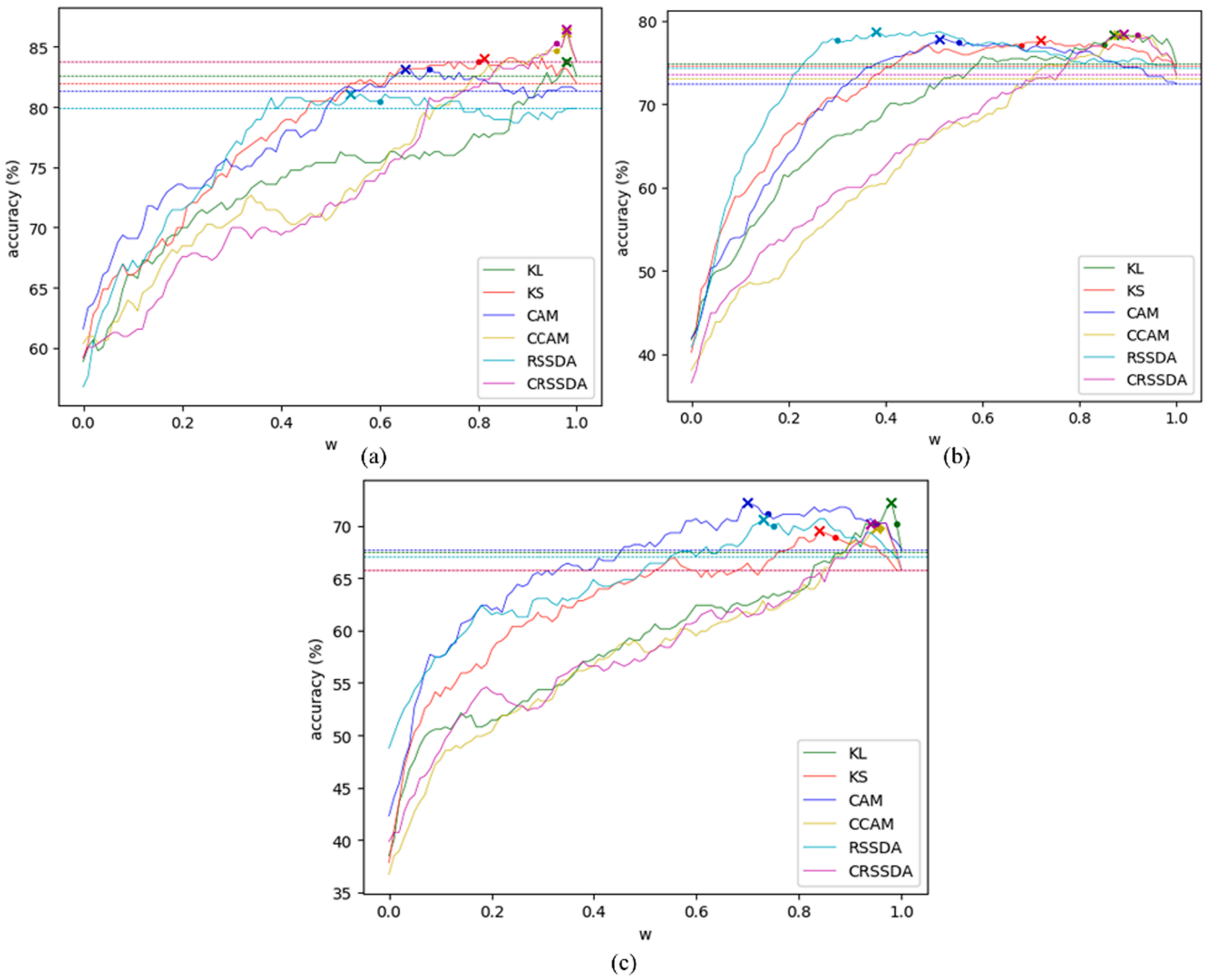

Fig. 9. Plots of overall classification accuracies against weight *w*, with the maximum accuracy indicated by a cross, and the selected weight indicated by a dot: (a) the IKONOS image, (b) the WorldView-2 image, and (c) the WorldView-3 image.

spectral features only. It also demonstrates advantages in both classification accuracy and efficiency over some state-of-the-art classifiers based on machine learning algorithms, which also consider spectral and spatial features. Objects with a distinctive spatial distribution benefit the most from the His-Cov model. In the worst scenario, if an object is very small and cannot provide useful spatial information, the His-Cov model degrades to the His model that does not consider spatial features. Although only tested in typical urban environments, the His-Cov model theoretically works in other environments. Potentially, the His-Cov could be used to combine the spatial features from HSR data with spectral features from hyperspectral data to classify, for example, tree species or crop types.

Given our experimental results, two types of misclassification require further discussion. One is the case where an object is correctly classified in the curve matching approaches with the His model but misclassified with the His-Cov model. This type of error is likely to happen to those classes without a distinctive spatial distribution such as the commercial and public class. A possible solution to the problem is to apply different weights to different classes. For example, assigning a larger spatial weight to the classes with a distinctive spatial distribution such as road and parking lot, and a smaller spatial weight to the classes without a distinctive spatial distribution such as commercial and public. Another solution is to make use of the uncertainty ratio of the classification based on the His and His-Cov models. When the two most similar samples of an object are of different classes but have very close divergence values for the spectral features (i.e., the uncertainty ratio is close to 1), the uncertainty of the classification by the His model is large. The classification of this object should be given a greater chance to change to a different class in the His-Cov model. On the other hand, if the uncertainty ratio is small in the His model, this object should be given a smaller chance to change to another class in the His-Cov model.

The second type of misclassification occurs when both the His and His-Cov models cannot correctly classify the object. Acquiring additional features may be required. For example, some road objects in the east–west direction are misclassified as the residential class or the commercial and public class in the WorldView-3 image (Fig. 7). This is because most sample objects of the road class are oriented north–south, resulting in matching problems in the east–west direction. This is not necessarily a sampling issue since topography or urban structure could result in a predominant orientation for roads. However, it is not common for buildings to be found in the middle of a road (although structures are occasionally built above highways). This knowledge of relative spatial location for object classes may be introduced into the curve matching approaches to correctly classify these objects into the road class.

6. Conclusion

This paper proposes a His-Cov model to enhance object-based classification by using curve matching approaches for both spectral variability and spatial distribution in the objects being considered. The spectral variability is derived from histograms or cumulative histograms of the pixel values in an object across all the multispectral bands, while their spatial distribution is established by the binary spatial covariogram in two directions. The method was tested on three different HSR images for land cover and land use classification in typical urban environments. The His-Cov model based on six curve matching approaches was compared with the corresponding approaches based on the His model, as well as with four advanced machine learning algorithms that also incorporate spectral and spatial features. All the results show that the proposed method can achieve higher accuracies, demonstrating the advantage of the spatial covariogram in characterizing the spatial distribution of the pixels in an object. The His-Cov model is especially suited for classifying objects containing distinctive spatial distribution information such as road and parking lot classes. Currently, however, only within-object spatial features are considered. In the future, we intend to consider the inclusion of between-object spatial relationships using the curve matching approach. Incorporating known spatial relationships between objects to be classified, such as road segments forming continuous links uninterrupted by other classes such as buildings, should further improve classification accuracy. An integration of machine learning and curve matching approaches will also be explored in an attempt to provide a fully generalized classification capability.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The paper was supported by the National Key R&D Program of China (2019YFC1520800), the National Science Foundation (NSF) (BCS-1826839), thekey project of the Aerospace Information Research Institute, CAS (Y951150Z2F), and the National Natural Science Foundation of China (41501489). We thank Dr. Ronald Briggs, Professor Emeritus of Geography and Geospatial Information Sciences at the University of Texas at Dallas, and Ms. Elizabeth Thompson for editing the language, and three anonymous reviewers for providing valuable comments and suggestions.

References

Arroyo, L.A., Healey, S.P., Cohen, W.B., Cocero, D., Manzanera, J.A., 2006. Using object-oriented classification and high-resolution imagery to map fuel types in a Mediterranean region. J. Geophys. Res. Biogeosci. 111 (4).
Atkinson, P.M., Naser, D.K., 2010. A geostatistically weighted K-NN classifier for remotely sensed imagery. Geographical Analysis 42, 204–225.
Balaguer, A., Ruiz, L.A., Hermosilla, T., Recio, J.A., 2010. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. Coumput. Geosci. 36, 231–240.
Balaguer-Beser, A., Ruiz, L.A., Hermosilla, T., Recio, J.A., 2013. Using semivariogram indices to analyse heterogeneity in spatial patterns in remotely sensed images. Comput. Geosci. 50, 115–127.
Belward, A.S., Skøien, J.O., 2015. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. ISPRS J. Photogramm. Remote Sens. 103, 115–128.
Benediktsson, J.A., Pesaresi, M., Arnason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. IEEE Trans. Geosci. Remote Sens. 41 (9), 1940–1949.
Berberoglu, S., Curran, P.J., Lloyd, C.D., Atkinson, P.M., 2007. Texture classification of Mediterranean land cover. Int. J. Appl. Earth Obs. Geoinf. 9 (3), 322–334.
Berberoglu, S., Lloyd, C.D., Atkinson, P.M., Curran, P.J., 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. Comput. Geosci. 26 (4), 385–396.
Berger, C., Voltersen, M., Hese, O., Walde, I., Schmullius, C., 2013. Robust extraction of urban land cover information from HSR multi-spectral and LiDAR data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (6), 1–16.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65 (1), 2–16.
Blaschke, T., Burnett, C., Pekkarinen, A., 2004. In: New contextual approaches using image segmentation for object-based classification. Remote Sensing Image Analysis: Including the Spatial Domain. Kluver Academic Publishers, Dordrecht, pp. 211–236.
Blaschke, T., Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GeoBIT/GIS 14 (6), 12–17.
Burt, J., Barber, G., 1996. Elementary Statistics for Geographers. The Guildford Press, New York.
Chang, C.-I., 2000. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. IEEE Trans. Inf. Theory 46 (5), 1927–1932.
Chen, Q., Gong, P., 2004. Automatic variogram parameter extraction for textural classification of the panchromatic Ikonos Imagery. IEEE Trans. Pattern Anal. Mach. Intell. 42 (5), 1106–1115.
Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13-17 August 2016; p. 785.
Chica-Olmo, M., Abarca-Hernández, F., 2000. Computing geostatistical image texture for remotely sensed data classification. Comput. Geosci. 26 (4), 373–383.
Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24 (5), 603–619.
Costa, H., Foody, G.M., Boyd, D.S., 2017. Using mixed objects in the training of object-based image classifications. Remote Sens. Environ. 190, 188–197.
Dell'Acqua, F., Gamba, P., Ferari, A., Palmason, J.A., Benediktsson, J.A., Arnason, K., 2004. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. IEEE Geosci. Remote Sens. Lett. 1 (4), 322–326.
El-naggar, A.M., 2018. Determination of optimum segmentation parameter values for extracting building from remote sensing images. Alexandria Engineering Journal 57, 3089–3097.
Gao, H., Wang, C., Wang, G., Li, Q., Zhu, J., 2020. A new crop classification method based on the time-varying feature curves of time series dual-polarization Sentinel-1 data sets. IEEE Geosci. Remote Sens. Lett. 17 (7), 1183–1187.
Geiß, C., Taubenböck, H., 2015. Object-based postclassification relearning. IEEE Geosci. Remote Sens. Lett. 12 (11), 2336–3234.
Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E., 2018. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. IEEE Geosci. Remote Sens. Lett. 15 (4), 607–611.
Hall-Beyer, M., 2007. GLCM Texture: A Tutorial v. 1.0, through2.7.
Hamada, Y., Stow, D.A., Coulter, L.L., Jafolla, J.C., Hendricks, L.W., 2007. Detecting tamarisk species (Tamarix Spp.) in Riparian habitats of Southern California using high spatial resolution hyperspectral imagery. Remote Sens. Environ. 109 (2), 237–248.
Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67 (5), 786–804.
Hay, G.J., Niemann, K.O., McLean, G., 1996. An object-specific image-texture analysis of H-resolution forest imagery. Remote Sens. Environ. 55, 108–122.
Hossain, M.D., Chen, D., 2019. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. ISPRS J. Photogramm. Remote Sens. 150, 115–134.
Huang, X., Zhang, L., Li, P., 2007. Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. IEEE Geosci. Remote Sens. Lett. 4 (2), 260–264.
Jiao, L., Liu, Y., Li, H., 2012. Characterizing land-use classes in remote sensing imagery by shape metrics. ISPRS J. Photogramm. Remote Sens. 72, 46–55.
Kruse, F.A., Lefkoff, A.B., Boardman, J.B., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The Spectral Image Processing System (SIPS) - interactive visualization and analysis of imaging spectrometer data. Remote Sens. Environ. 44, 145–163.
Laliberte, A.S., Fredrickson, E.L., Rango, A., 2007. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. Photogramm. Eng. Remote Sens. 2 (11), 197–207.
Li, Y., Chen, J., Rao, Y., 2018. A practical sampling method for assessing accuracy of detected land cover/land use change: Theoretical analysis and simulation experiments. ISPRS J. Photogramm. Remote Sens. 144, 379–389.
Liu, D., Han, L., 2017. Spectral curve shape matching using derivatives in hyperspectral images. IEEE Geosci. Remote Sens. Lett. 14 (4), 504–508.
Ma, L., Li, M., Ma, X., Cheng, L., Du, P., Liu, Y., 2017. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote Sens. 130, 277–293.
Melville, B., Lucieer, A., Aryal, J., 2018. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. Int. J. Appl. Earth Obs. Geoinf. 66, 46–55.
Newman, M.E., Mclaren, K.P., Wilson, B.S., 2011. Comparing the effects of classification techniques on landscape-level assessments: pixel-based versus object-based classification. Int. J. Remote Sens. 32 (14), 4055–4073.
Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybernet.-Syst. 9 (1), 62–66.
Pedley, M.I., Curran, P.J., 1991. Per-field classification: an example using SPOT HRV imagery. Int. J. Remote Sens. 12, 2181–2192.

Radhakrishna, A., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2274–2282.
Russ, J.C., 2002. The Image Processing Handbook, 4th ed. CRC Press, Boca Raton, FL.
Shackelford, A.K., Davis, C.H., 2003. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. IEEE Trans. Geosci. Remote Sens. 41, 2354–2364.
Shen, Y., Chen, J., Xiao, L., Pan, D., 2019. Optimizing multiscale segmentation with local spectral heterogeneity measure for high resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 157, 13–25.
Sridharan, H., Qiu, F., 2013. Developing an object-based hyperspatial image classifier with a case study using WorldView-2 data. Photogramm. Eng. Remote Sens. 79 (11), 1027–1036.
Stow, D.A., Toure, S.I., Lippitt, C.D., Lippitt, C.L., Lee, C., 2012. Frequency distribution signatures and classification of within-object pixels. Int. J. Appl. Earth Obs. Geoinf. 15 (1), 49–56.
Su, T., 2019. Scale-variable region-merging for high resolution remote sensing image segmentation. ISPRS J. Photogramm. Remote Sens. 147, 319–334.
Tang, Y., Jing, L., Li, H., Atkinson, P.M., 2016. A multiple-point spatially weighted k-NN method for object-based classification. Int. J. Appl. Earth Obs. Geoinf. 52, 263–274.
Tang, Y., Jing, L., Shi, F., Li, X., Qiu, F., 2019. A hybrid model integrating spatial pattern, spatial correlation, and edge information for image classification. Remote Sensing 11, 1599.
Toure, S.I., Stow, D.A., Weeks, J.R., Kumar, S., 2013. Histogram curve matching approaches for object-based image classification of land cover and land use. Photogramm. Eng. Remote Sens. 79 (5), 433–440.
van der Meer, F., 2000. Spectral curve shape matching with a continuum removed CCSM algorithm. Int. J. Remote Sens. 21 (16), 3179–3185.
van der Meer, F., 2006. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. Int. J. Appl. Earth Obs. Geoinf. 8, 3–17.
Vorobiova, N., Chernov, A., 2017. Curve fitting of MODIS NDVI time series in the task of early crops identification by satellite images. Procedia Eng. 201, 184–195.
Wang, G., Liu, J., He, G., 2013. A method of spatial mapping and reclassification for high-spatial-resolution remote sensing image classification. The ScientificWorld Journal, 192982.
Wu, S., Xu, B., Wang, L., 2006. Urban land-use classification using variogram-based analysis with an aerial photograph. Photogramm. Eng. Remote Sens. 72 (7), 813–822.
Wu, X., Peng, J., Shan, J., Cui, W., 2015. Evaluation of semivariogram features for object-based image classification. Geo-spatial Information Science 18 (4), 159–170.
Ye, S., Pontius Jr., R.G., Rakshit, R., 2018. A review of accuracy assessment for object-based image analysis: from per-pixel to per-polygon approaches. ISPRS J. Photogramm. Remote Sens. 141, 137–147.
Yue, A., Zhang, C., Yang, J., Su, W., Yun, W., Zhu, D., 2013. Texture extraction for object-oriented classification of high spatial resolution remotely sensed images using a semivariogram. Int. J. Remote Sens. 34 (11), 3736–3759.
Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. ISPRS J. Photogramm. Remote Sens. 140, 133–144.
Zhong, Y., Lin, A., He, L., Zhou, Z., Yuan, M., 2020. Spatiotemporal dynamics and driving forces of urban land-use expansion: a case study of the Yangtze River Economic Belt. China. Remote Sensing 12, 287.
Zhou, W., Troy, A., 2008. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. Int. J. Remote Sens. 29, 3119–3135.
Zhou, Y., Qiu, F., 2015. Fusion of high spatial resolution WorldView-2 imagery and LiDAR pseudo-waveform for object-based image analysis. ISPRS J. Photogramm. Remote Sens. 101, 221–232.
Zhou, Y., Qiu, F., Ni, F., Lou, Y., Zhang, C., Alfarhan, M., Al-Dosari, A.A., 2016. Curve matching approaches to waveform classification: a case study using ICESat data. GIScience& Remote Sens. 53 (6), 739–758.