ISPRS Journal of Photogrammetry and Remote Sensing 135 (2018) 127–141

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Angular difference feature extraction for urban scene classification using ZY-3 multi-angle high-resolution satellite imagery

Xin Huang [a,b,1,*], Huijun Chen [a,1], Jianya Gong [a,b]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, PR China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

ARTICLE INFO

Article history:
Received 1 August 2017
Received in revised form 3 October 2017
Accepted 21 November 2017

Keywords:
Multi-angle
Urban classification
High spatial resolution
Scene classification

ABSTRACT

Spaceborne multi-angle images with a high-resolution are capable of simultaneously providing spatial details and three-dimensional (3D) information to support detailed and accurate classification of complex urban scenes. In recent years, satellite-derived digital surface models (DSMs) have been increasingly utilized to provide height information to complement spectral properties for urban classification. However, in such a way, the multi-angle information is not effectively exploited, which is mainly due to the errors and difficulties of the multi-view image matching and the inaccuracy of the generated DSM over complex and dense urban scenes. Therefore, it is still a challenging task to effectively exploit the available angular information from high-resolution multi-angle images. In this paper, we investigate the potential for classifying urban scenes based on local angular properties characterized from high-resolution ZY-3 multi-view images. Specifically, three categories of angular difference features (ADFs) are proposed to describe the angular information at three levels (i.e., pixel, feature, and label levels): (1) ADF-pixel: the angular information is directly extrapolated by pixel comparison between the multi-angle images; (2) ADF-feature: the angular differences are described in the feature domains by comparing the differences between the multi-angle spatial features (e.g., morphological attribute profiles (APs)). (3) ADF-label: label-level angular features are proposed based on a group of urban primitives (e.g., buildings and shadows), in order to describe the specific angular information related to the types of primitive classes. In addition, we utilize spatial-contextual information to refine the multi-level ADF features using superpixel segmentation, for the purpose of alleviating the effects of salt-and-pepper noise and representing the main angular characteristics within a local area. The experiments on ZY-3 multi-angle images confirm that the proposed ADF features can effectively improve the accuracy of urban scene classification, with a significant increase in overall accuracy (3.8–11.7%) compared to using the spectral bands alone. Furthermore, the results indicated the superiority of the proposed ADFs in distinguishing between the spectrally similar and complex man-made classes, including roads and various types of buildings (e.g., high buildings, urban villages, and residential apartments).

1. Introduction

High-resolution satellite imagery enables a more detailed observation of the Earth at fine scales, which provides new opportunities for detailed urban land-cover mapping. However, urban classification is a challenging task due to the spectral heterogeneity and structural diversity of the complex geospatial objects (Khatami et al., 2016). For instance, it is difficult to separate man-made objects (e.g., roads and buildings) because of their spectral similarities (Pacifici et al., 2009). Although spatial and structural features such as morphological profiles (Mura et al., 2010) and textural metrics (Pacifici et al., 2009) have been used to complement spectral features in a great number of studies, the complexity of urban scenes, especially in the vertical dimension, seriously affects the interpretation accuracy.

In recent years, satellites with the ability to capture high-resolution stereo images (e.g., ZY-3) have become accessible. Such satellites provide multi-view observations and allow the retrieval of 3D structure characteristics that are difficult to obtain with a single-view mode. These multi-angle images simultaneously

* Corresponding author at: School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, PR China.
E-mail address: xhuang@whu.edu.cn (X. Huang).
[1] Xin Huang and Huijun Chen are co-first authors.

https://doi.org/10.1016/j.isprsjprs.2017.11.017
0924-2716/

provide both high-resolution multispectral bands and abundant information about 3D structures, which is particularly suitable for the interpretation of urban scenes with complex vertical structures. Therefore, there has been increasing interest in applying high-resolution multi-angle imagery to urban classification. Most of the existing high-resolution multi-angle classification studies generated a digital surface model (DSM) using image matching methods, and subsequently utilized the height information derived from the DSM to enhance the classification results. For example, a multi-angle derived DSM was stacked with spectral, textural, and morphological features using the random forest (RF) classifier by Longbotham et al. (2012). Their results showed that the inclusion of the multi-angle information bolsters the ability to classify spectrally similar classes with significant height differences, such as bridges and highways. Tian et al. (2014) proposed a method for building change detection based on stereo imagery and DSMs generated with a stereo matching methodology. The results showed that the fusion of height information and multispectral images could significantly improve the performance of change detection compared to using spectral or height information alone. Li et al. (2016) presented an approach for land-cover mapping in surface-mined and agricultural landscapes based on ZY-3 stereo satellite imagery. The mean and standard deviation filters of the spectral bands and topographic features derived from the ZY-3 stereo images were employed for classification. In spite of the progress made, a major disadvantage of these methods is that the performance of the classification is subject to the accuracy of the generated DSM (Tian et al., 2013), which can be seriously affected by inaccurate matching points, incompleteness, and blurred boundaries in the proximities of buildings (Aguilar et al., 2014). The satellite-derived DSMs also tend to underestimate the height of high buildings (Huang et al., 2017a). At the same time, the DSM may ignore the more implicit angular information contained in the multi-angle images, leading to underutilization of the discriminative features. These drawbacks emphasize the need to investigate more effective features or approaches to fully exploit the available multi-angle images and obtain a high classification accuracy in complex urban scenes.

The differences in the multi-angle images can be considered as additional features that provide information about the radiative and structural characteristics of the scenes (Diner et al., 2005), including the lateral sides of the objects (Xiao et al., 2012), materials of man-made construction (Longbotham et al., 2012), reflective properties of the land surface (Puttonen et al., 2009), and shadow-casting and mutual obscuration of three-dimensional surface elements (Lucht et al., 2000). For elevated objects with 3D structure (e.g., buildings and trees), the angular effects are particularly significant (Diner et al., 2005; Licciardi et al., 2012; Pasher and King, 2010). For example, in forest studies, changes in canopy structure, including changes in tree crown size, shape, density, and the spatial distribution of leaves, affect the directional scattering of light. Multi-angle observations of this scattering thereby reveal information about the three-dimensional structure of the vegetation (Chopping et al., 2008). For urban studies, in high-resolution multi-angle images, a lot of detailed information about the 3D structures of the elevated objects emerges, and can therefore provide cues about the properties of the buildings, such as the materials, structures, and heights. For example, the vertical structures of buildings presented in multi-angle imagery can provide strong evidence for building detection (Xiao et al., 2012). In this context, effective methods capable of synergistically integrating the cues from multi-angle imagery are needed to explore the great potential of angular information in observing 3D structures and gaining a better understanding of urban scenes. In this study, we used the ZiYuan-3 (ZY-3) multi-view images. The ZY-3 satellite, launched in January 2012, is China's first civilian high-resolution three-line array stereo satellite. The ZY-3 satellite can simultaneously collect multi-view panchromatic images, and this unique merit makes it particularly suitable for vertical feature extraction of the Earth's surface.

In this context, this paper proposes a series of novel multi-level angular difference features (ADFs), in order to make full use of the angular information contained in high-resolution multi-angle images for urban classification. The proposed method describes the angular information at three levels (pixel, feature, and label levels) to reveal the angular variation patterns of different urban scenes. Specifically, at the pixel level, the angular information is directly extrapolated by pixel comparison between the multi-angle images. At the feature level, in order to make full use of the spatial structures in the high-resolution images and describe the angular differences in the spatial domains, several angular features are extracted based on spatial features (e.g., attribute profiles (Mura et al., 2010)), which can provide a multi-level characterization of an image and can model the different kinds of structural information. At the label level, the urban scenes are represented using a group of primitives (e.g., building/shadow), including their frequency and spatial arrangement, in order to describe the angular differences related to the specific primitive types. The multi-level features can characterize the image angular information from various perspectives, and can complement each other in classifying different land covers. It should be noted that after obtaining the multi-level ADFs, the ADFs are refined based on superpixel segmentation, for the purpose of alleviating the effect of noise and representing the main angular characteristics within a local area. The performance of the proposed multi-level ADF feature set was assessed using a series of ZY-3 satellite stereo images over representative urban areas. In the experiments, we examined the use of the proposed multi-level ADFs for urban classification, and compared the ADFs with the state-of-the-art spatial features and the commonly used height feature (e.g., DSM).

The remainder of this paper is structured as follows. Section 2 introduces the proposed ADF method. The experimental analysis, which includes the description of the datasets, the classification results, as well as the feature analysis and discussion, is given in Section 3. Section 4 concludes the paper with some closing remarks and gives suggestions for future research directions.

2. Methodology

2.1. Overview

The main idea of the proposed approach is that the angular differences can be considered as additional features that allow further discrimination between land-cover classes with analogous spectral properties in urban environments. To illustrate this concept, Fig. 1 shows examples where the same urban scenes are represented under the three viewing angles of ZY-3 images. In order to thoroughly demonstrate the differences in angular variation patterns based on class type, Fig. 2 shows the normalized multi-angle panchromatic values for six urban classes, computed by averaging the normalized multi-angle panchromatic values of a ground-truth reference set (dataset 2, see Section 3.1) for each class. We have conducted a relative normalization to the multi-angle images using histogram matching method, taking the nadir images as the reference. As shown in the first row in Fig. 1, the low-lying class (e.g., road) remains almost the same across the multi-angle images, while the high buildings present apparent angular variations within a local area. This is due to the solar observational cross-section, an effect responsible for changes in the reflectance of the objects with non-flat surfaces (Matasci et al., 2015). This indicates that the degree of local angular variation contains implicit height

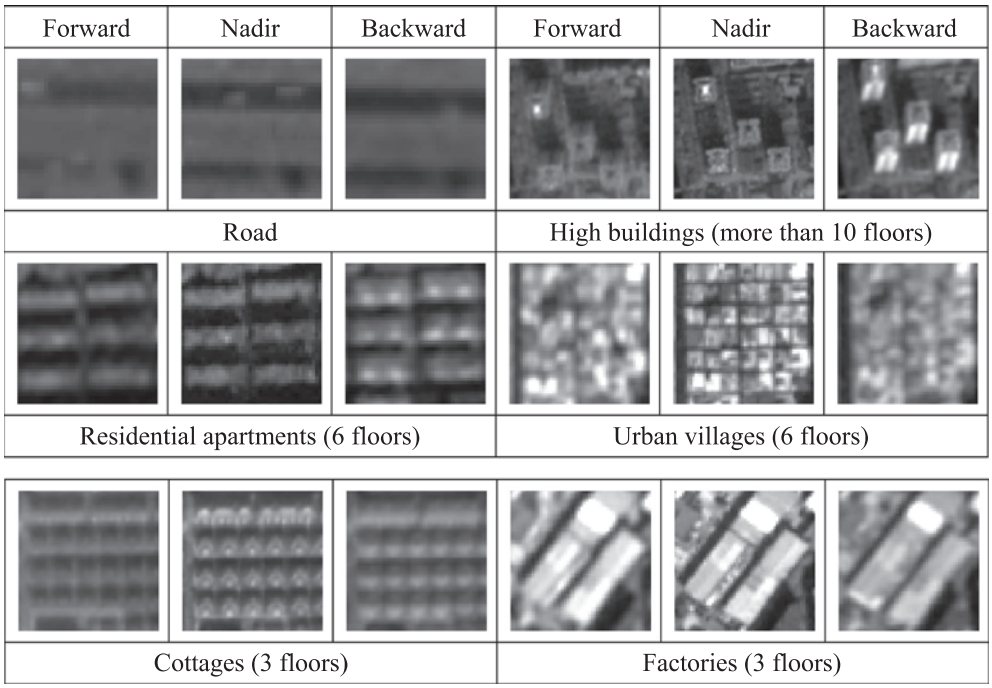

Fig. 1. ZY-3 multi-angle images for six urban classes.

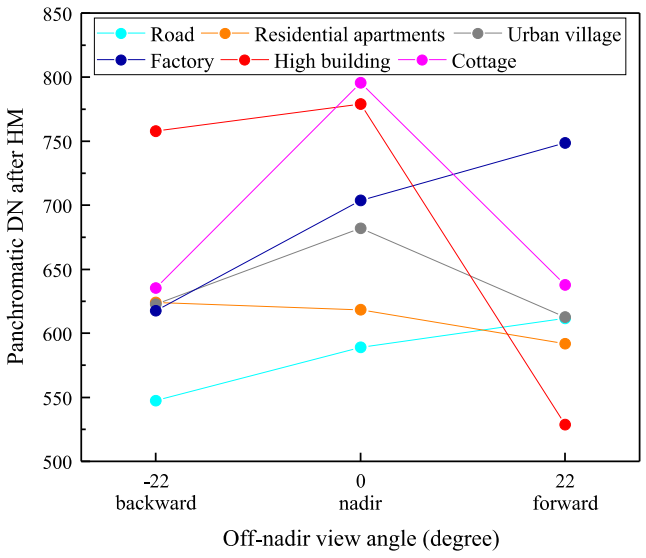

Fig. 2. Multi-angle panchromatic values for six urban classes. The multi-angle panchromatic values are normalized by histogram matching (HM).

information and can help to discriminate the classes that have similar spectral responses but different height characteristics, such as roads and different types of buildings. Note that although the widely used DSMs are able to provide height information, it has been reported that DSMs derived from ZY-3 stereo images are not accurate enough to estimate building height (RMSE = 7.78 m) for Chinese cities with complex 3D urban landscapes (Huang et al., 2017a). In particular, high buildings (more than 50 m) tend to be underestimated, due to the large disparity and occlusion, which is problematic for image matching. The local angular variations may provide a new possibility to delineate high buildings and their height information, which is particularly useful in cases where matching errors occur and the DSM is less accurate.

Furthermore, local angular variations also offer the possibility to distinguish man-made structures with similar heights. For instance, looking at the second row of Fig. 1, it can be seen that for residential apartments and urban villages (densely distributed buildings with little vegetation and public space (Huang et al., 2015)) with similar heights (both six floors), the differences in mutual obscuration and shadow-casting due to the building density are clearly presented in the multi-angle images. Specifically, the sparsely distributed residential apartments present bright lateral sides in the backward imagery, whereas the lateral sides of the urban villages are rarely shown, because of the high building density. This observation is also confirmed by Fig. 2. It is noticeable that sparse residential apartments present higher reflectance values in the nadir and backward directions due to the higher fraction of bright lateral sides shown. In contrast, the angular effects of the densely distributed urban villages are less apparent and exhibit a relatively flat curve. Similarly, as shown in Fig. 2, factories and cottages with similar heights (both three floors) exhibit dissimilar angular variation patterns that may possibly be attributed to the different reflectance properties of the materials and structures. These examples illustrate that spectrally similar man-made objects may have distinct angular properties, and that these angular properties can be exploited in urban classification.

The flowchart of the proposed approach is shown in Fig. 3. Based on the coregistered stereo images, the angular difference features (ADFs) are extracted to highlight the regions with large angular variations at three levels: (1) pixel level (ADF-pixel); (2) feature level (ADF-feature); and (3) label level (ADF-label) (see Table 1). The multi-level ADFs can provide a comprehensive characterization of the angular properties. Subsequently, the ADFs are refined by spatial smoothing based on superpixel segmentation, for the purpose of alleviating the effect of salt-and-pepper noises and representing the main angular characteristics within a local area. Prior to ADF extraction, the backward and forward imagery are resampled at the same spatial resolution as the nadir imagery. Three sets of ADFs can be generated through the combinations of different viewing angles, i.e., nadir and forward (NF), nadir and backward (NB), and forward and backward (FB), respectively.

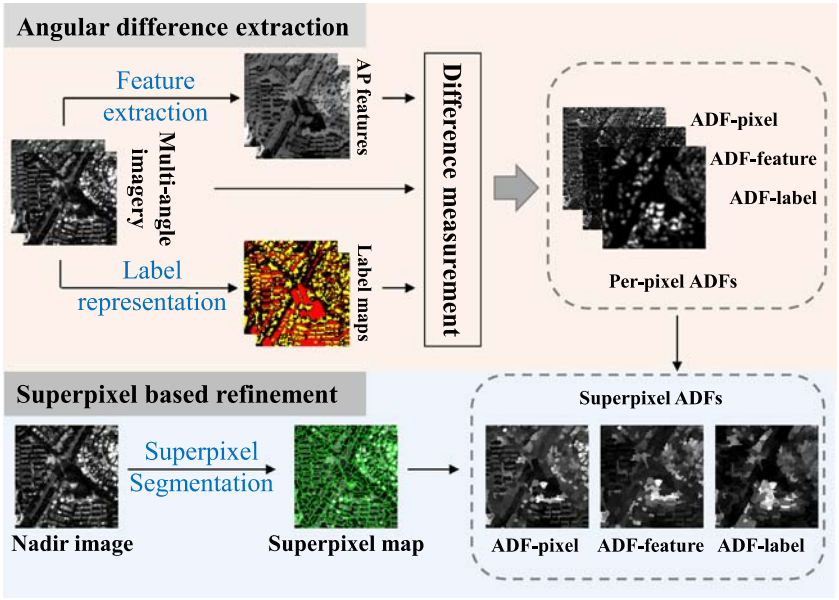

Fig. 3. Flowchart of the approach used in this study.

Table 1
Overview of the multi-level ADFs proposed in this study.

Symbol	Description
P	ADF-pixel
F(area)	ADF-feature built with the area attribute
F(diag)	ADF-feature built with the diagonal box attribute
F(iner)	ADF-feature built with the moment of inertia attribute
F(std)	ADF-feature built with the standard deviation attribute
L(shadow)	ADF-label built with the shadow primitive
L(build)	ADF-label built with the building primitive
L(back)	ADF-label built with the background primitive

2.2. Angular difference features (ADFs)

2.2.1. ADF-pixel

Under the condition that the ZY-3 multi-angle images are acquired simultaneously, it can be assumed that the apparent land-cover changes did not occur during the acquisition, and the differences between the multi-angle images are caused by the angular effects of objects. Based on this assumption, it is possible to directly extrapolate the angular information by pixel comparison between the multi-angle images. A straightforward way is image differencing, which produces an absolute residual image to represent the pixel-level angle difference. Given a pair of stereo panchromatic images X_1 and X_2 acquired over the same area from different viewing angles A_1 and A_2, respectively, the pixel-level angle difference *ADF-pixel* can be described as:

$$P = |X_1 - X_2| \tag{1}$$

ADF-pixel is expected to highlight the pixels associated with significant angular differences, and thus can be used to identify and delineate the off-ground classes (e.g., buildings) and distinguish them from the low-lying classes such as roads and soil.

2.2.2. ADF-feature

ADF-pixel describes the angular information based on the differences between stereo images on a per-pixel basis. Similarly but furthermore, the structural and geometric features extracted from multi-angle images can be utilized to represent the feature-level angle difference. On the one hand, the structural features can compensate for the inadequacy of the spectral information and make full use of the spatial details in the high-resolution images. One the other hand, the elevated objects present evident variations of spatial properties from different viewing angles (see Section 2.1), which may in turn reveal material and structural characteristics of the urban objects. This structural variation can be captured by the structural and geometric features.

To effectively describe the geometrical information of high-resolution images, morphological attribute profiles (APs), which provide a multi-level characterization of an image, are adopted in this study to model the different kinds of structural information (Mura et al., 2010). APs are a generalization of morphological profiles (MPs) (Pesaresi and Benediktsson, 2001), with the capacity to extract different kinds of spatial features by applying a series of attribute filters. We can assume that the attribute filters process an image f according to a criterion T with n morphological attribute thickening operators (ϕ^T) and attribute thinning operators (γ^T), and the AP is obtained the morphological filter by reconstruction:

$$AP(f) = \{\phi_n^T(f), \phi_{n-1}^T(f), \ldots, \phi_1^T(f), \gamma_1^T(f), \ldots, \gamma_{n-1}^T(f), \gamma_n^T(f)\} \tag{2}$$

In general, the criterion compares the value of an arbitrary attribute α measured on the component C against a given reference parameter value λ. If the criterion is fulfilled, then the regions remain unchanged; otherwise, they are set to the gray level of a darker or brighter surrounding region, according to whether the transformation performed is thickening or thinning, respectively. According to the attribute considered, different structural information can be extracted from an image (Dalla Mura et al., 2010). In this paper, four attributes are considered to construct the AP features: the area, the standard deviation, diagonal of the box and the moment of inertia.

Based on the AP and its different attributes, the feature-level angle difference, *ADF-feature*, can be defined as:

$$F(\alpha) = |AP_\alpha(X_1) - AP_\alpha(X_2)| \tag{3}$$

where $AP_\alpha(X_1)$ and $AP_\alpha(X_2)$ denote the AP features of X_1 and X_2 with attribute α, respectively. Examples of the ADF-features for four urban classes are given in Fig. 4 for a ZY-3 image from Beijing, China. It is shown that with the AP-based multi-angle feature representation, distinctive angular properties can be obtained for the different classes. The use of the ADF-features helps in discriminat-

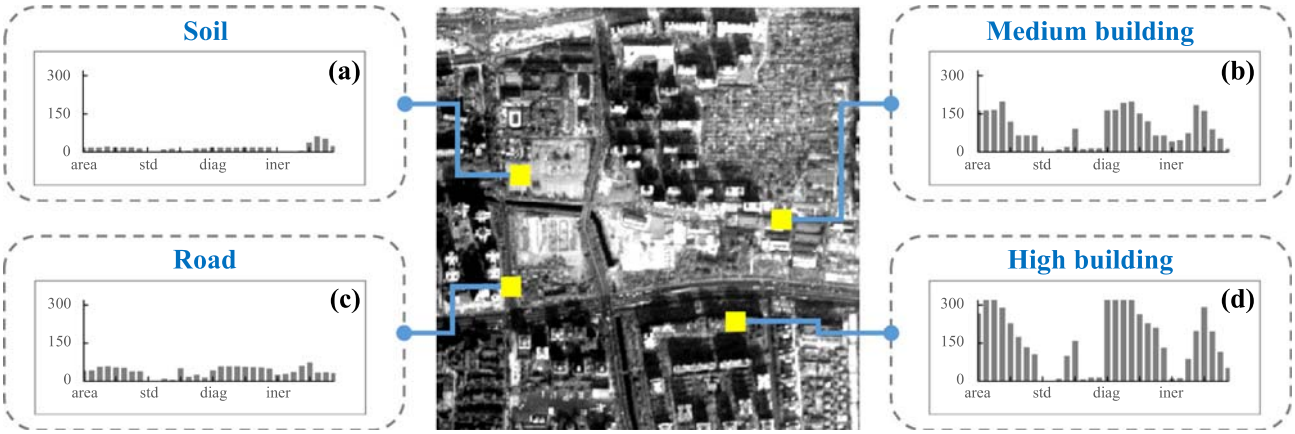

Fig. 4. ADF-feature for several typical urban classes (soil, medium building, high building, and road). The four attributes (i.e., area, moment of inertia, standard deviation, and diagonal of the box) are labeled on the horizontal axis. The parameters are set according to the suggestions in Marpu et al. (2013). The ADF-feature is calculated from the nadir and forward imagery.

ing the buildings, which exhibit strong differences under different acquisition angles, from the road and soil, which present relatively consistent structures in different viewing angles. It is interesting to see that the high building class (more than 10 floors) shows higher ADF-feature values than the spectrally similar medium building (six to nine floors) and road classes.

2.2.3. ADF-label

ADF-pixel and ADF-feature describe the intensity of the angular difference from the perspective of pixel and feature levels, respectively. At the label level, where urban primitives (e.g., buildings and shadows) are explicitly identified in each image, multi-angle information related to the specific land-cover categories can be depicted. Buildings and shadows represent urban primitives, and they exhibit significant variations in size, shape, and location under different viewing angles (Lee and Kim, 2015). Such angular variations of buildings and shadows can provide important cues for the analysis of 3D objects. As shown in Fig. 5, the proposed ADF-label is composed of two main steps: (1) label representation, in which the urban scenes are represented using a couple of primitives (i.e., buildings and shadow) that are calculated automatically; and (2) label-level angular feature extraction, which is aimed at measuring the frequency and spatial arrangement of the urban primitives and subsequently deriving the label-level angular features.

Step 1: Label representation. The urban primitives, including buildings and shadows, are extracted automatically using the morphological building index (MBI) and the morphological shadow index (MSI) (Huang et al., 2012), respectively. The MBI and MSI are chosen considering their ability to automatically generate building/shadow structures from high-resolution urban images. The MBI is an effective index to highlight building structures from high-resolution imagery by representing the spectral-spatial properties of buildings (e.g., brightness, contrast, size, and directionality) with a series of morphological operators. The calculation of MBI is based on the fact that the relatively high reflectance of roofs and the spatially adjacent shadows lead to high local contrast of buildings. The MBI is defined as:

$$\mathrm{MBI} = \frac{\sum_{s\in S}\sum_{d\in D}(\mathrm{DMP}-\mathrm{WTH}(s,d))}{N_S \times N_D} \tag{4}$$

where DMP-WTH denotes the differential morphological profiles of the white top-hat, which is able to highlight the locally bright structures; d and s represent the direction and scale of the structural element, and N_S and N_D are the total number of scales and directions, respectively. The MSI can be viewed as a twinborn index of the MBI since shadows show low reflectance but high local contrast. Consequently, the black top-hat (BTH), which is able to highlight the dark structures within the defined directions and scales, is used to construct the shadow index:

$$\mathrm{MSI} = \frac{\sum_{s\in S}\sum_{d\in D}(\mathrm{DMP}-\mathrm{BTH}(s,d))}{N_S \times N_D} \tag{5}$$

Subsequently, the label maps can be obtained by simply applying a threshold to the indices. The threshold values for the MBI and MSI were chosen according to the previous studies (Huang et al., 2017b, 2012).

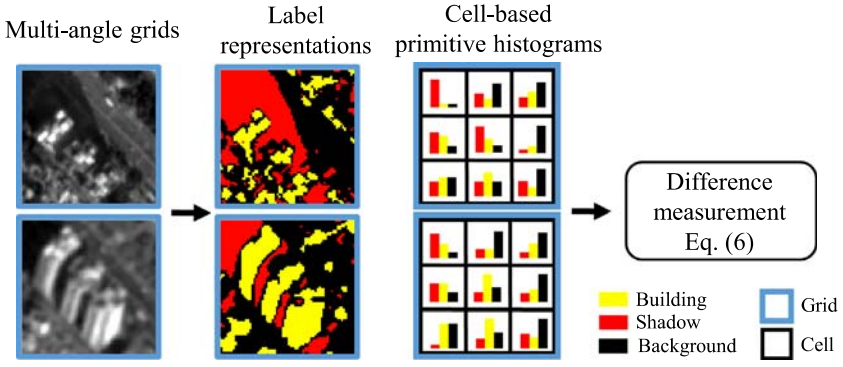

Fig. 5. Demonstrations for ADF-label.

Step 2: Difference measurement. Based on the stereo label maps derived from the first step, the composition and spatial distribution of the urban primitives are further described using a local histogram representation with a cell-grid strategy (Wen et al., 2016). ADF-label is then calculated by measuring the differences between the multi-angle primitive histograms. Specifically, an image is first divided into a series of grids with the size of $N \times N$ (pixels), which is regarded as the basic unit for calculating ADF-label. Then, as shown in Fig. 5, each grid is further divided into $n \times n$ cells, where the frequencies of the primitives in each cell are used to characterize the spatial distribution and arrangement of the primitives in the grid. In this way, a grid can be described by $n \times n$ histograms, with each histogram representing the frequencies of the primitives in each cell. The main advantage of this cell-grid strategy is the ability to simultaneously describe the frequency and spatial distribution of the urban primitives (Wen et al., 2016). Subsequently, the label-level angle difference *ADF-label* for primitive i is calculated by measuring the differences between the multi-angle primitive histograms in each grid:

$$L(i) = \sum_{x=1}^{n\times n} |H_1^x(i) - H_2^x(i)|, i \in \{building, shadow, background\} \quad (6)$$

where $H_1^x(i)$ and $H_2^x(i)$ denote the frequency of primitive i in the xth cell ($1 \le x \le n^2$), for viewing angles A_1 and A_2, respectively. A half-overlapped grid approach was used when calculating the ADF-label feature values (see Fig. 6). The main advantage of the half-overlapped grid is that sufficient contextual information is incorporated, and at the same time the loss of the spatial details (caused by the moving window) can be reduced. As shown in Fig. 6, the images are divided into a series of half-overlapped grids. The ADF-label value is computed at each grid, and the final ADF-label feature value is defined as the average of the overlapping areas.

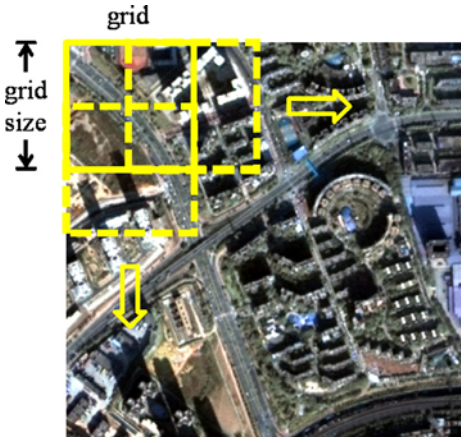

Fig. 6. The half-overlapped grid approach.

Three examples for the multi-angle label representation are shown in Fig. 7 (right), from which it can be clearly seen that the local angular variations can be effectively captured by the label maps. Note that there is a high degree of variability in the multi-angle label maps and histogram differences for the buildings, particularly for the high buildings. It appears that the angular information can provide a means of distinguishing between the spectrally similar urban classes. This can be seen in Fig. 7b, where ADF-label is displayed in false-color composite, where the high buildings (more than 10 floors) are displayed as bright blue and the medium buildings (six to nine floors) are displayed as purple-yellow. The

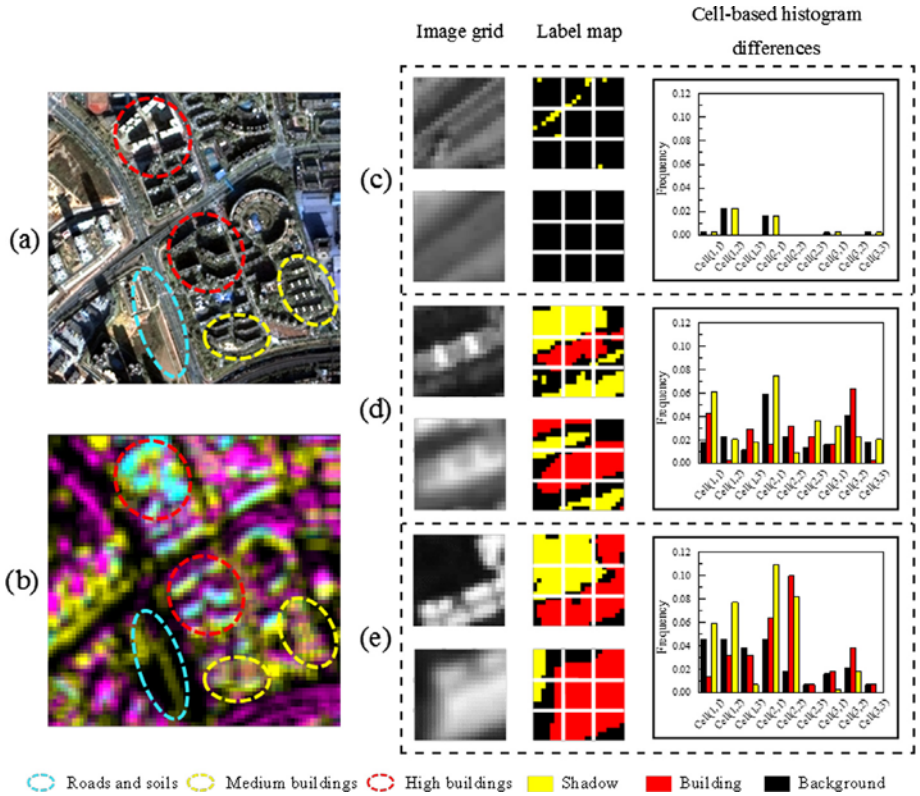

Fig. 7. Demonstration of ADF-label. Left: (a) nadir image; (b) ADF-label in false-color composite displaying background, building, and shadow primitives in red, green, and blue. The soil, medium building (six to nine floors), and high building (more than 10 floors) are marked by the blue, yellow, and red ellipses, respectively. Right: multi-angle (upper/lower nadir/backward) label representations and cell-based primitive histogram differences for the different classes, where (c), (d), and (e) represent road, medium building, and high building, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

low-lying areas (e.g., soil and roads) tend to have much lower values, and are shown in black.

2.3. Superpixel-based refinement (SBR)

It is widely acknowledged that the use of the contextual information of pixels can increase the accuracy of pixel-based land-cover classification (Bruzzone et al., 2006). Moreover, Johnson and Xie (2013) found that segment-based features, which provide relatively homogenous local information for a pixel by considering its neighborhood, are more robust and reliable than pixel-based features. In this study, the per-pixel ADFs are further refined in such a manner. After obtaining the multi-level ADFs, superpixel-based refinement (SBR) is implemented to: (1) alleviate the salt-and-pepper effect; (2) compensate for the residual misregistration errors which are unavoidable when combining multi-angle data; and (3) preserve the edges from the nadir images. Superpixels in a high-resolution scene are defined as pure perceptual uniform parcels, and a land-cover object is composed of several adjacent superpixels (Li et al., 2015). The superpixel ADFs are calculated by averaging the per-pixel ADFs within each superpixel to characterize the local angular variations. Please note that the nadir panchromatic image is utilized as the base image for the superpixel segmentation, due to its high spatial resolution.

The superpixel segmentation method utilized in this paper is based on graph partitioning and the entropy rate (ER) (Liu et al., 2011), which favors compact and homogenous non-overlapping clusters. The only free parameter T (the number of superpixels) controls the segmentation scale of the scene. Specifically, the ER first maps the base image to a graph $G = (V, E)$, with the vertices representing the pixels, and the edge weights denoting the pairwise similarities between adjacent pixels. Subsequently, the

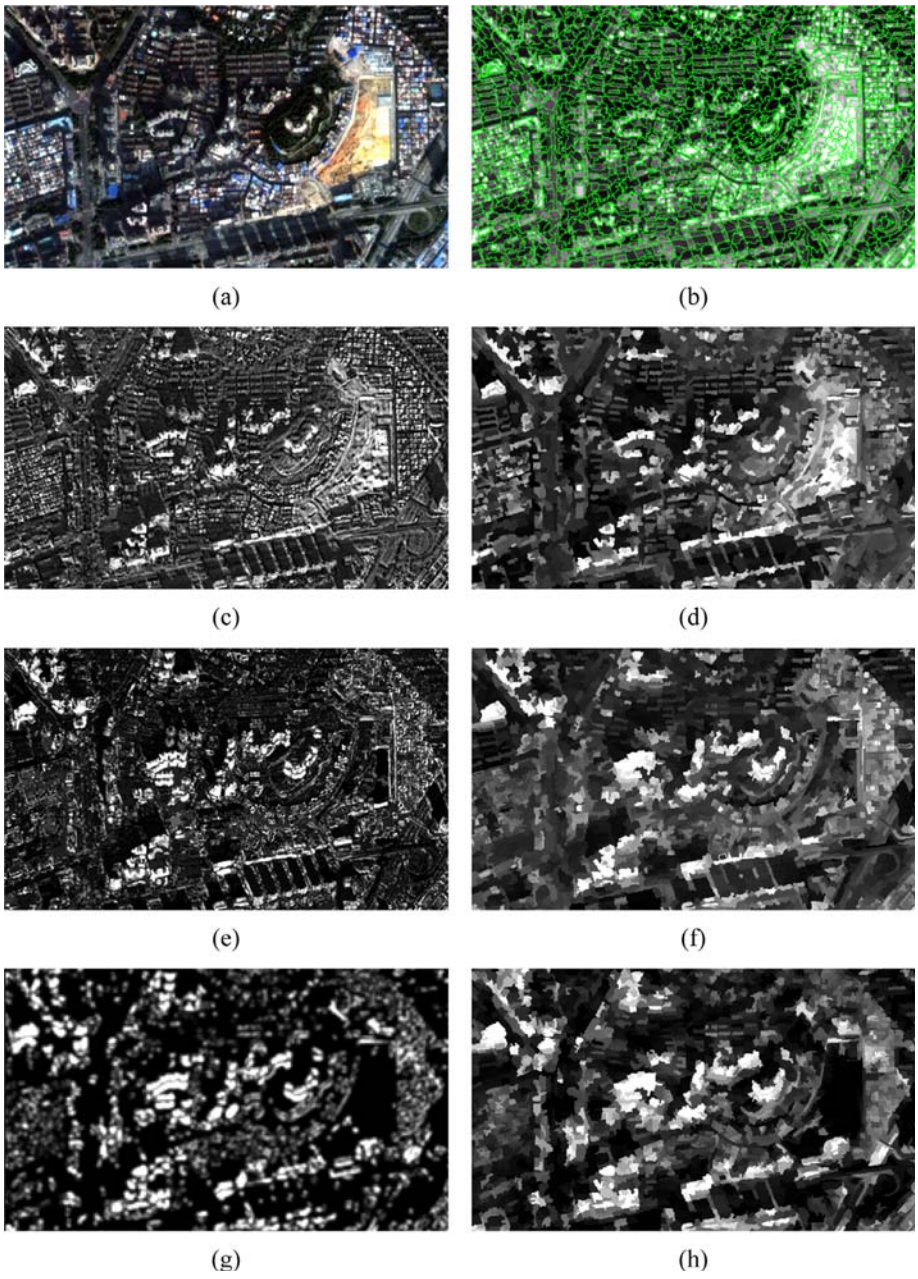

Fig. 8. The multi-level ADFs with and without superpixel based refinement (SBR). (a) True-color nadir image. (b) Superpixel segmentation. (c) ADF-pixel (i.e., P) without SBR. (d) P with SBR. (e) ADF-feature built on the area attribute (i.e., F(area)) without SBR. (f) F(area) with SBR. (g) The ADF-label built on shadow primitives (i.e., L(shadow)) without SBR. (h) L(shadow) with SBR. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

segmentation is accomplished through partitioning the graph G into T connected subgraphs. This is achieved by selecting a subset of edges $A \subseteq E$ and maximizing an objective function with respect to the edge set. In order to obtain compact, homogeneous, and balanced superpixels, both the ER $H(\bullet)$ and a balancing term $B(\bullet)$ are combined into the objective function:

$$\max_{A}\{H(A) + \lambda B(A)\}, \text{ subject to } A \subseteq E \quad (7)$$

where λ is the weight of the balancing term.

Fig. 8 compares the multi-level ADFs with and without SBR. Although the per-pixel ADF is able to highlight the pixels associated with significant angular differences, it suffers from salt-and-pepper noise (the first column) to some extent. After applying the SBR (the second column), most of the noise can be removed, and at the same time the general patterns related to angular characteristics are retained. In addition, by incorporating the boundaries from the nadir images, the building edges are better preserved.

3. Experiments and discussion

3.1. Datasets

The ZY-3 satellite is China's first civilian high-resolution satellite specifically designed for along-track stereo imagery collection, with a spatial resolution of 2.1 m for the nadir angle and 3.6 m for the forward and backward angles. The incidence angles of the forward and backward cameras are ±22° with respect to the nadir camera. The ZY-3 three-line array images are acquired nearly simultaneously. Such images are particularly suitable for extracting angular properties because, when the multi-angle images are acquired simultaneously, it can be assumed that the apparent land-cover changes did not occur during the acquisition, and the differences between the multi-angle images are mainly caused by the angular effects of objects. Therefore, this unique merit makes it potential to extrapolate angular information by measuring the differences between ZY-3 multi-angle images. Although ZY-3 images have been previously used in urban studies, including classification (Huang et al., 2014) and change monitoring (Huang et al., 2017a), the man-made structures were roughly divided into roads and buildings in these studies, and the potential of the multi-angle ZY-3 images in distinguishing different categories of buildings has rarely been exploited. More importantly, the multi-angle characteristic of the ZY-3 imagery has never been explored in the current literature, mainly due to the errors and difficulties of multi-view image matching and the inaccuracy of the generated DSM over complex and dense urban scenes (Huang et al., 2017a). Therefore, it is interesting to investigate whether the ZY-3 multi-angle images are able to differentiate different types of man-made constructions in complex urban environments (especially in the vertical dimension) by incorporating the angular difference information.

Three ZY-3 multi-angle datasets were used to evaluate the performance of the proposed method. The first two datasets used in the experiments were acquired over the city of Shenzhen, China, in December 2013, and the third dataset is an image of Beijing, China, acquired in October 2012. The three test images are made up of 824 × 830, 1098 × 1097, and 1197 × 1194 pixels, respectively (Fig. 9a). Pan-sharpening was carried out by applying the Gram-Schmidt procedure to the nadir panchromatic image and the four-band multispectral image. The backward and forward imagery were resampled at the same spatial resolution as the nadir imagery, and were registered to the nadir imagery using the nearest-neighbor method, with a root-mean-square error of less than one pixel. A relative normalization to the multi-angle images was conducted using the histogram matching method, taking the nadir images as the reference. The test areas represent a series of typical urban areas with complex scenes, e.g., various man-made structures with heterogeneous sizes and shapes, including roads and different types of buildings. The key challenge in classifying these images lies in the confusion among these man-made class types, which usually present analogous spectral features in high-resolution images.

3.2. Experimental setup

We compared the proposed multi-level ADFs with the APs extracted from nadir imagery and the nDSM (normalized DSM) derived from ZY-3 stereo imagery in classification. The DSM was generated from the ZY-3 stereo imagery using the semi-global matching (SGM) algorithm (Hirschmüller, 2008; Qin, 2016, 2014). Subsequently, the nDSM was derived from the DSM by morphological top-hat by reconstruction (Qin and Fang, 2014). The parameters of the APs were set according to the suggestions of Marpu et al. (2013). The proposed ADFs are superpixel features since the segment-based features are more robust and reliable than per-pixel ones by considering local contexts (Johnson and Xie, 2013). For a fair comparison, the compared features (i.e., spectral bands, APs and nDSM) were also refined by the superpixel processing.

The nadir images are used as the base image for the superpixel segmentation. According to the image resolution and the object sizes in the test areas, the number of superpixels was set to 3000 for dataset 1, and 5000 for datasets 2 and 3. The cell size of ADF-label was set to 3 pixels, and each grid was divided into 3 × 3 cells (n = 3) to capture the subtle angular differences. The ZY-3 data provide three panchromatic stereo angles, i.e., 22° backward, nadir, and 22° forward, resulting in three possible stereo pairs. Therefore, three sets of ADFs were generated through the combinations of different viewing angles, i.e., nadir and forward (NF), nadir and backward (NB), and forward and backward (FB). These three sets of ADFs were concatenated into one stacked feature vector and used as input for the classification. Seven groups of feature sets were compared (Table 2). As shown in Table 2, at first, each level of ADF was considered separately (cases 4–6), and then all the ADFs were taken into account simultaneously (case 7).

Nine land-cover classes were defined in the experiment: cottages, residential apartments (six to nine floors), high buildings (more than 10 floors), roads, factories, urban villages (densely distributed buildings with little public space), vegetation, bare soil, and shadows. The classes were defined to investigate the ability of the ADFs to discriminate between different man-made class types. Fig. 9d shows the reference data manually delineated from the corresponding nadir images through field investigation and visual interpretation of the ZY-3 images and Google Earth.

For each test site, 100 samples per class were randomly chosen from the reference dataset to train the random forests (RF) classifier (Breiman, 2001). The numbers of training and test samples for each dataset are shown in Table 3. The number of decision trees of RF was set to 100. All the experiments were repeated 10 times with different and randomly generated training samples, and the mean and standard deviation of the classification results are reported for the assessment of the classification performance. Three measures were calculated to evaluate the classification performance: (1) the overall accuracy (OA); (2) the kappa coefficient (KC); and (3) the kappa coefficient for man-made constructions (KCM), including cottages, residential apartments, high buildings, roads, factories, and urban villages, which is used to evaluate the ability of the classifier to distinguish between different types of man-made structures.

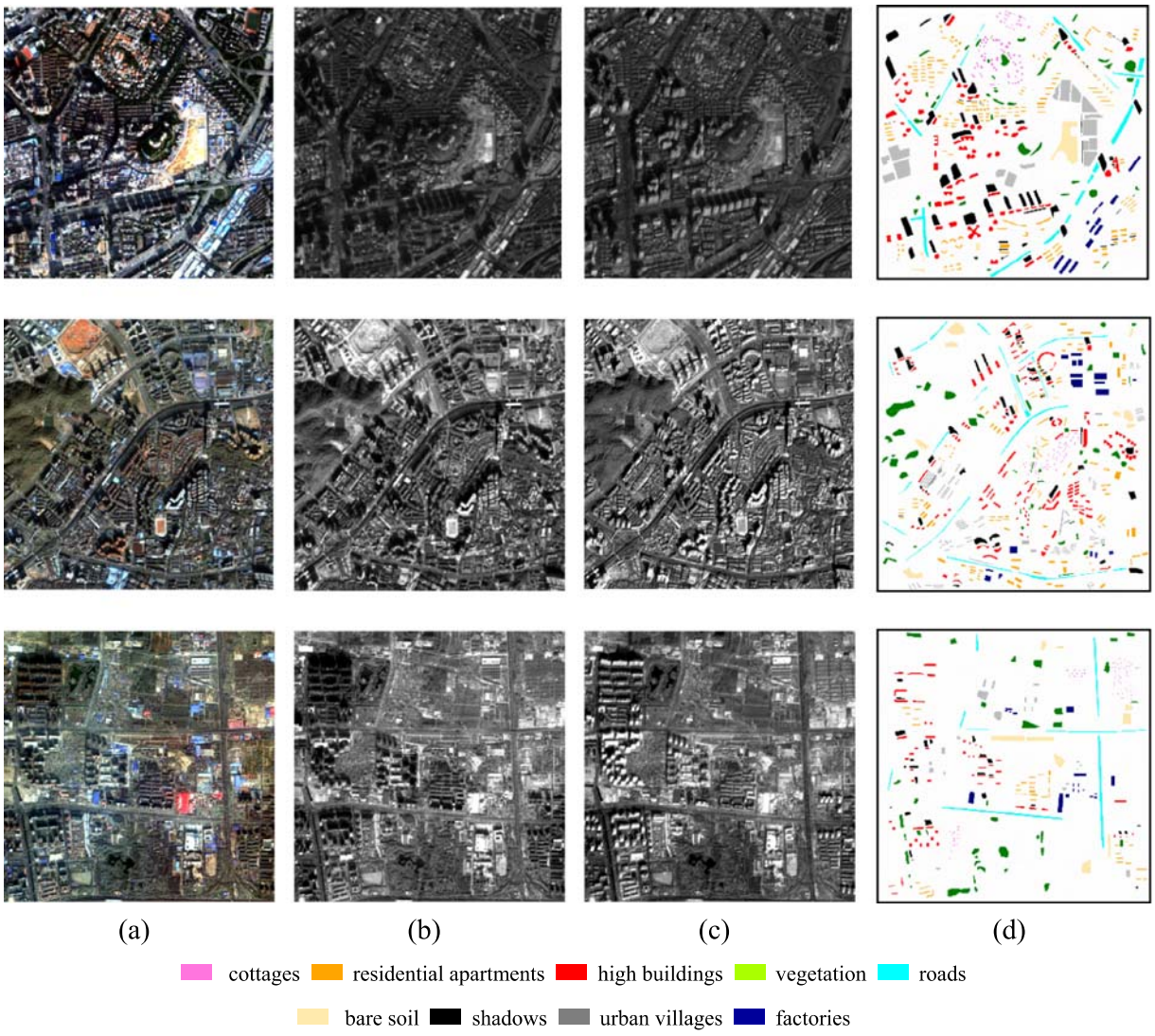

Fig. 9. The three datasets: (a) true-color nadir image; (b) forward image; (c) backward image; (d) ground-truth reference selected from the nadir image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
The number of input features for each combination of features tested in each study area.

Case	Name	Input features	Dimension of features
1	S	Nadir spectral bands	4
2	S + AP	Nadir spectral bands and nadir AP	36
3	S + nDSM	Nadir spectral bands and nDSM	5
4	S + P	Nadir spectral bands and ADF-pixel	7
5	S + F	Nadir spectral bands and ADF-feature	100
6	S + L	Nadir spectral bands and ADF-label	13
7	S + PFL	Nadir spectral bands and all ADF	112

3.3. Classification results

The classification results of the three ZY-3 test sites are provided in Tables 4–6, respectively. The result with the highest accuracy is underlined for each class. The classification maps are presented in Figs. 10–12. In general, from Tables 4–6, it can be seen that the combination of the multi-level ADFs and spectral bands (S + PFL) yields the highest overall accuracy and kappa coefficient among all the datasets. In dataset 1 and dataset 2 with the heterogeneous urban scenes, the S + PFL feature set presents a significant OA improvement (8.8–11.7%) compared to using spectral bands alone. In dataset 3, which is a sub-urban area, S + PFL achieves a lower OA improvement (3.8%). This suggests the superiority of considering angular information when interpreting complex urban scenes. The promising performance when using angular difference information can also be confirmed by the results of the feature sets when each level of ADF is considered separately (i.e., S + P, S + F, and S + L). In fact, in most cases, the results achieved with the single-level ADF are better than those achieved by the compared features (i.e., AP and nDSM).

We recall that KCM is the kappa coefficient between the man-made classes, i.e., cottages, residential apartments, high buildings, roads, factories, and urban villages. The relatively inferior KCM compared with kappa coefficient indicates the difficulty of distinguishing between the man-made class types (Tables 4–6). According to KCM, in all cases, the optimal feature set is the combination of multi-level ADFs and spectral bands (S + PFL), with a KCM improvement of 0.12–0.23, compared to using only spectral bands. Furthermore, it can be seen that by including the ADFs, the accuracy of most of the classes is significantly improved. Most of the man-made classes classified in these experiments show a

Table 3
The number of training and test samples for each class.

Class	Dataset 1		Dataset 2		Dataset 3	
	Training	Test	Training	Test	Training	Test
Cottages	100	1890	100	996	100	1210
Residential apartments	100	10,404	100	15,446	100	4150
High buildings	100	13,043	100	22,814	100	9658
Vegetation	100	7956	100	18,944	100	19,818
Roads	100	9183	100	10,982	100	13,225
Bare soil	100	6153	100	9816	100	13,092
Shadows	100	20,376	100	14,979	100	6023
Urban villages	100	18,441	100	10,555	100	6457
Factories	100	3435	100	8978	100	6015
Total	900	90,881	900	113,510	900	79,648

Table 4
The classification accuracies for ZY-3 dataset 1.

Class	Raw	Compared features		The proposed ADF features			
	S	S + AP	S + nDSM	S + P	S + F	S + L	S + PFL
Cottages	0.894 ± 0.018	0.920 ± 0.027	0.895 ± 0.023	0.881 ± 0.020	0.933 ± 0.029	0.907 ± 0.021	0.942 ± 0.026
Res. Apa.	0.716 ± 0.038	0.767 ± 0.031	0.752 ± 0.029	0.787 ± 0.029	0.838 ± 0.021	0.837 ± 0.030	0.857 ± 0.032
High Bui.	0.650 ± 0.041	0.721 ± 0.032	0.755 ± 0.021	0.801 ± 0.021	0.902 ± 0.018	0.845 ± 0.029	0.907 ± 0.017
Vegetation	0.946 ± 0.018	0.946 ± 0.024	0.941 ± 0.021	0.945 ± 0.020	0.949 ± 0.019	0.945 ± 0.022	0.948 ± 0.014
Roads	0.814 ± 0.050	0.933 ± 0.027	0.932 ± 0.025	0.944 ± 0.021	0.961 ± 0.016	0.908 ± 0.026	0.960 ± 0.015
Bare soil	0.970 ± 0.019	0.984 ± 0.014	0.996 ± 0.003	0.977 ± 0.013	0.973 ± 0.017	0.981 ± 0.013	0.975 ± 0.014
Shadows	0.934 ± 0.020	0.943 ± 0.013	0.936 ± 0.016	0.949 ± 0.016	0.953 ± 0.017	0.947 ± 0.013	0.953 ± 0.015
Urban Vil.	0.720 ± 0.030	0.884 ± 0.036	0.886 ± 0.025	0.865 ± 0.030	0.934 ± 0.019	0.852 ± 0.028	0.937 ± 0.016
Factories	0.970 ± 0.018	0.970 ± 0.023	0.989 ± 0.008	0.961 ± 0.018	0.978 ± 0.013	0.952 ± 0.029	0.971 ± 0.023
OA	0.817 ± 0.011	0.882 ± 0.006	0.884 ± 0.004	0.892 ± 0.005	0.931 ± 0.006	0.898 ± 0.006	0.934 ± 0.005
KC	0.786 ± 0.012	0.861 ± 0.006	0.864 ± 0.005	0.874 ± 0.006	0.919 ± 0.006	0.881 ± 0.007	0.923 ± 0.006
KCM	0.700 ± 0.018	0.830 ± 0.014	0.831 ± 0.010	0.848 ± 0.011	0.926 ± 0.014	0.859 ± 0.009	0.929 ± 0.013

Table 5
The classification accuracies for ZY-3 dataset 2.

Class	Raw	Compared features		The proposed ADF features			
	S	S + AP	S + nDSM	S + P	S + F	S + L	S + PFL
Cottages	0.917 ± 0.028	0.963 ± 0.029	0.964 ± 0.014	0.952 ± 0.024	0.959 ± 0.023	0.949 ± 0.025	0.962 ± 0.014
Res. Apa.	0.658 ± 0.030	0.726 ± 0.016	0.719 ± 0.041	0.719 ± 0.041	0.745 ± 0.039	0.761 ± 0.032	0.776 ± 0.033
High Bui.	0.566 ± 0.035	0.716 ± 0.033	0.795 ± 0.024	0.791 ± 0.020	0.862 ± 0.021	0.842 ± 0.020	0.870 ± 0.024
Vegetation	0.952 ± 0.010	0.931 ± 0.017	0.938 ± 0.015	0.938 ± 0.012	0.888 ± 0.016	0.927 ± 0.018	0.879 ± 0.021
Roads	0.796 ± 0.034	0.861 ± 0.015	0.848 ± 0.028	0.840 ± 0.036	0.873 ± 0.023	0.864 ± 0.025	0.881 ± 0.016
Bare soil	0.919 ± 0.025	0.957 ± 0.021	0.953 ± 0.017	0.944 ± 0.023	0.954 ± 0.022	0.936 ± 0.021	0.952 ± 0.022
Shadows	0.960 ± 0.012	0.969 ± 0.014	0.963 ± 0.011	0.963 ± 0.010	0.964 ± 0.009	0.959 ± 0.016	0.959 ± 0.014
Urban Vil.	0.664 ± 0.022	0.829 ± 0.036	0.726 ± 0.015	0.797 ± 0.035	0.801 ± 0.027	0.790 ± 0.027	0.796 ± 0.033
Factories	0.959 ± 0.023	0.960 ± 0.021	0.968 ± 0.018	0.971 ± 0.019	0.956 ± 0.022	0.951 ± 0.034	0.954 ± 0.026
OA	0.791 ± 0.011	0.853 ± 0.008	0.858 ± 0.009	0.863 ± 0.008	0.876 ± 0.005	0.876 ± 0.008	0.879 ± 0.007
KC	0.761 ± 0.013	0.831 ± 0.009	0.837 ± 0.010	0.842 ± 0.009	0.857 ± 0.006	0.857 ± 0.009	0.861 ± 0.008
KCM	0.657 ± 0.021	0.769 ± 0.014	0.780 ± 0.020	0.786 ± 0.013	0.840 ± 0.010	0.817 ± 0.015	0.848 ± 0.016

Table 6
The classification accuracies for ZY-3 dataset 3.

Class	Raw	Compared features		The proposed ADF features			
	S	S + AP	S + nDSM	S + P	S + F	S + L	S + PFL
Cottages	0.904 ± 0.033	0.927 ± 0.039	0.916 ± 0.029	0.966 ± 0.016	0.972 ± 0.014	0.952 ± 0.024	0.982 ± 0.009
Res. Apa.	0.856 ± 0.022	0.899 ± 0.029	0.927 ± 0.018	0.876 ± 0.034	0.908 ± 0.023	0.925 ± 0.025	0.921 ± 0.029
High Bui.	0.806 ± 0.019	0.816 ± 0.036	0.847 ± 0.029	0.905 ± 0.010	0.906 ± 0.010	0.884 ± 0.013	0.910 ± 0.010
Vegetation	0.928 ± 0.026	0.934 ± 0.027	0.923 ± 0.037	0.924 ± 0.031	0.879 ± 0.030	0.926 ± 0.033	0.884 ± 0.043
Roads	0.783 ± 0.041	0.920 ± 0.027	0.856 ± 0.024	0.873 ± 0.027	0.931 ± 0.028	0.870 ± 0.030	0.931 ± 0.020
Bare soil	0.946 ± 0.023	0.950 ± 0.035	0.965 ± 0.019	0.966 ± 0.014	0.963 ± 0.027	0.950 ± 0.019	0.959 ± 0.021
Shadows	0.952 ± 0.020	0.955 ± 0.021	0.950 ± 0.027	0.947 ± 0.020	0.944 ± 0.017	0.951 ± 0.022	0.937 ± 0.013
Urban Vil.	0.882 ± 0.033	0.920 ± 0.028	0.913 ± 0.031	0.946 ± 0.027	0.946 ± 0.015	0.915 ± 0.027	0.948 ± 0.017
Factories	0.970 ± 0.024	0.975 ± 0.016	0.980 ± 0.013	0.980 ± 0.018	0.979 ± 0.012	0.973 ± 0.015	0.981 ± 0.014
OA	0.889 ± 0.009	0.922 ± 0.011	0.915 ± 0.010	0.926 ± 0.009	0.926 ± 0.008	0.920 ± 0.009	0.927 ± 0.010
KC	0.870 ± 0.011	0.908 ± 0.013	0.900 ± 0.011	0.913 ± 0.010	0.913 ± 0.009	0.907 ± 0.011	0.915 ± 0.011
KCM	0.839 ± 0.013	0.921 ± 0.013	0.901 ± 0.013	0.920 ± 0.015	0.957 ± 0.009	0.922 ± 0.014	0.960 ± 0.009

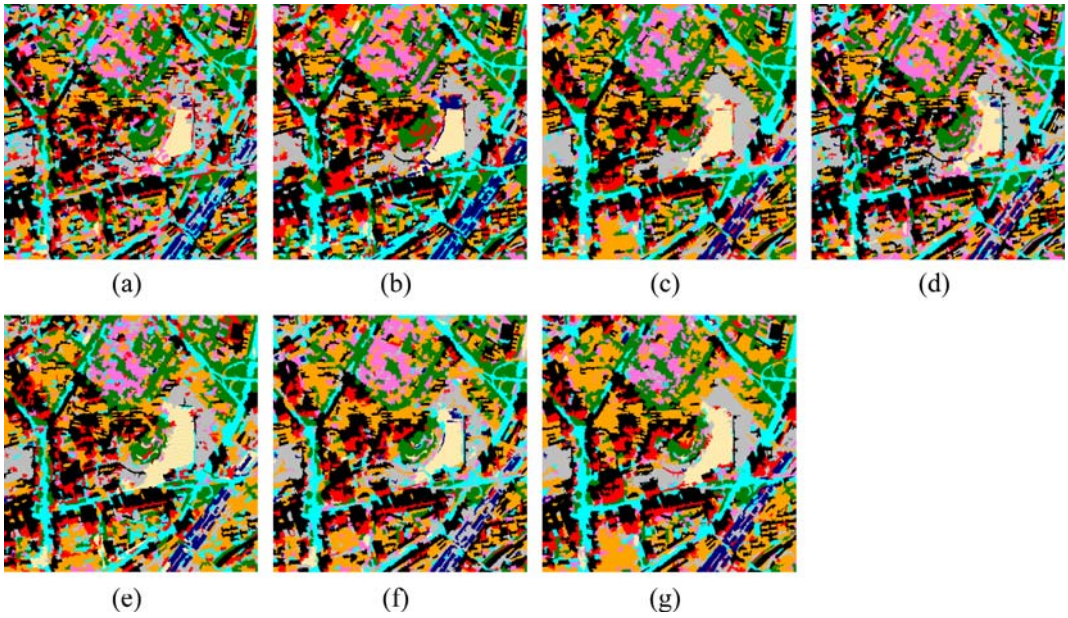

Fig. 10. Classification results with dataset 1: (a) S; (b) S + nDSM; (c) S + AP; (d) S + P; (r) S + F; (f) S + L; and (g) S + PFL (see the color legend in Fig. 9).

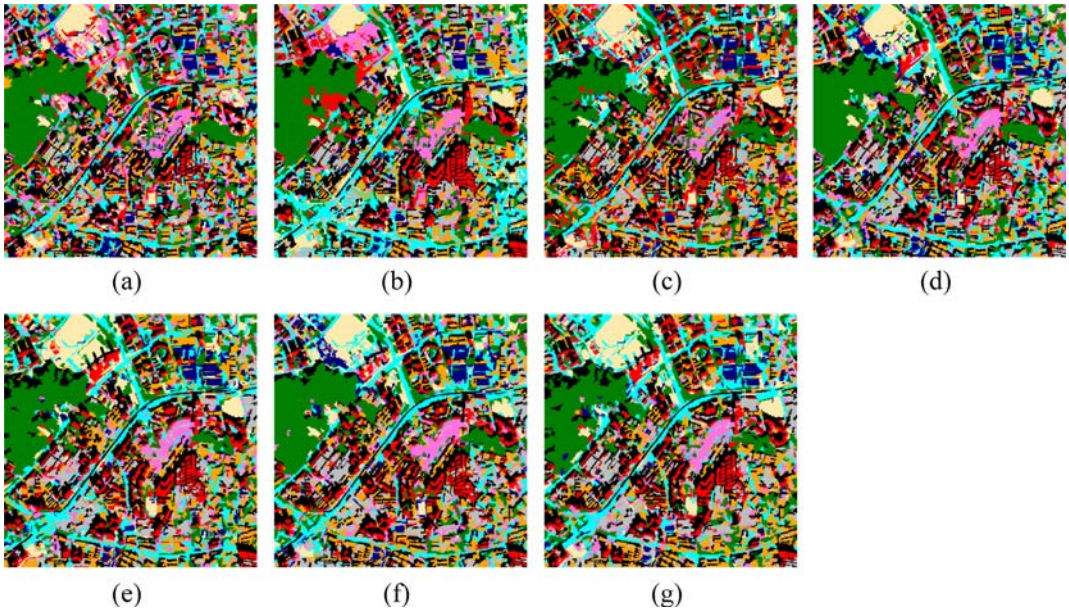

Fig. 11. Classification results with dataset 2: (a) S; (b) S + nDSM; (c) S + AP; (d) S + P; (r) S + F; (f) S + L; and (g) S + PFL (see the color legend in Fig. 9).

significant improvement that can be attributed to the angular difference information. For example, in dataset 2, the most heterogeneous scene, the spectral bands show a relative inability to classify the residential apartments, high buildings, and urban villages (see the first column of Tables 4–6), and these classes have an accuracy of less than 70%. After adding the multi-level ADFs, the accuracy increase for each of these classes is as much as 11.8%, 30.4%, and 13.2%, respectively. The cottages and factories show a relatively low accuracy improvement, because the spectral bands (S) already achieve an accuracy of over 89% for these classes, which is possibly due to the prominence of the distinctive colors of these classes.

By focusing on the accuracy of the ADFs at each level, the largest OA improvement is obtained by ADF-feature (3.7–11.4%), followed by ADF-label (3.1–8.5%), while ADF-pixel, as the raw and low-level form of angular difference information, provides the modest but still significant improvement (3.7–7.5%). On the one hand, this suggests that extracting information from each image, whether at the feature level or label level, can make better use of the multi-angle information and provide an additional improvement in the classification performance. On the other hand, the satisfactory performance of ADF-pixel, in spite of it being the rawest form of angular difference information with the lowest feature dimension (only one feature for an angle combination), is important to note, and its performance can further indicate the effectiveness of the angular difference information.

Comparing the two types of features derived from the stereo images (i.e., nDSM and ADFs) with the AP, it can be seen that ADFs significantly outperform APs, and nDSM yields very close accuracy to APs. The improved classification accuracy of ADFs, as opposed to APs, is mainly due to the improved classification of man-made

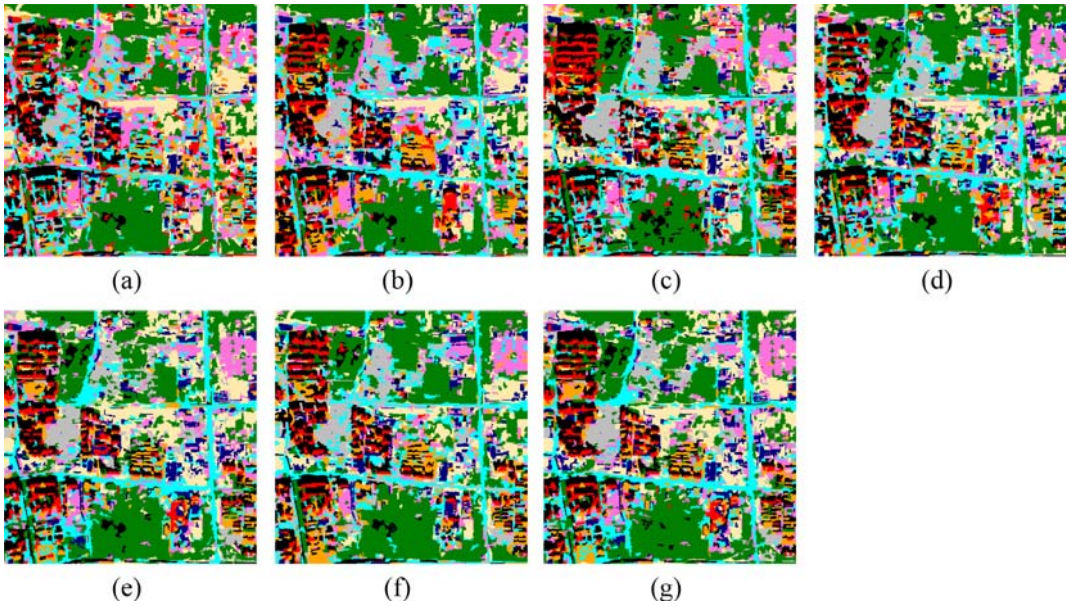

Fig. 12. Classification results with dataset 3: (a) S; (b) S + nDSM; (c) S + AP; (d) S + P; (r) S + F; (f) S + L; and (g) S + PFL (see the color legend in Fig. 9).

structures, as ADFs largely improve KCM. This indicates that multi-angle information greatly increase the inter-class separability between the most commonly misclassified man-made categories in urban scenes.

When comparing ADFs and nDSM, the former presents a higher accuracy. As mentioned previously, this is due to the fact that the classification accuracy of nDSM can be affected by the matching errors, which can be attributed to the severe occlusions and shadows in complex urban scenes. For instance, as shown in Fig. 13, the high buildings (marked with yellow boxes) are not recognized in the nDSM but, instead, they are successfully identified by the ADFs, since ADFs can capture the angular variation characteristics, instead of extracting the elevation parallaxes which depend strongly on the matching efficacy, as in the case of the nDSM. This leads to an accuracy increment for high buildings from 6.3% to 15.2%, compared to nDSM. Another explanation for the superior performance of ADFs is that the multi-level ADFs are able to further explore the implicit angular information that is ignored by the nDSM. Thus, they can be more useful for the delineation of man-made class types with similar heights. In particular, in dataset 1, the average nDSM values for residential apartments and urban villages are 16.5 m and 20.7 m, respectively, making it difficult to distinguish between residential apartments and urban villages using the S + nDSM feature set, due to their similar height and color. However, the S + PFL feature set has a better ability to differentiate between these two classes. For instance, the average values of F (area) for residential apartments and urban villages are 34.8 and 62.7, respectively, showing an increase in the accuracy of residential apartments in dataset 1 from 75.2% (S + nDSM) to 85.7% (S + PFL), and an increase in the accuracy of urban villages from 88.6% (S + nDSM) to 93.7% (S + PFL). A similar phenomenon can also be observed for dataset 2, where the average nDSM value for residential apartments and urban villages is 18.5 m and 21.2 m, respectively. The multi-level ADFs (S + PFL) show gains in classification accuracy for both classes compared to nDSM, with an improvement of 5.7% for residential apartments and 7.0% for urban villages. As the results show, ADFs offer obvious advantages over nDSM for the delineation of man-made class types, and can further improve the accuracy of urban classification.

3.4. Feature analysis of the multi-level ADFs

As reported in Section 3.3, the multi-level ADFs (S + PFL) can significantly improve the classification results. Furthermore, it would be interesting to quantitatively analyze the contribution of the multi-level ADFs to the urban classification. Therefore, a feature contribution analysis was performed according to the feature importance quantified by permutation importance (Breiman, 2001). This method randomly permutes the values of a feature in all samples, and then classifies the out-of-bag samples (OOB, the unchosen samples for training a decision tree) by the RF classifier. The average decrease in accuracy over all the trees, caused by the feature permutation, is regarded as the feature importance.

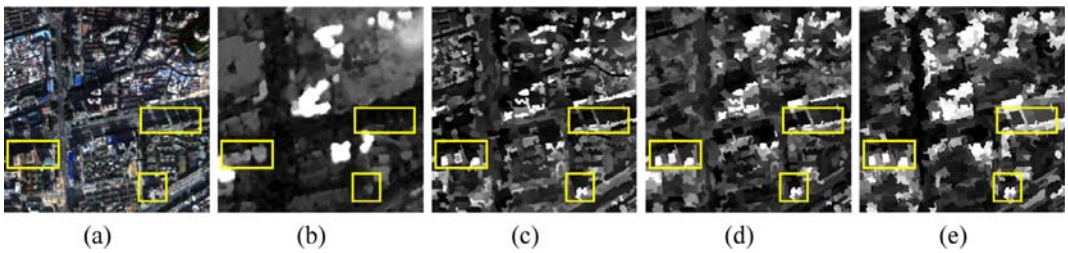

Fig. 13. Exemplary comparison between ADFs and nDSM with dataset 1: (a) true-color nadir image; (b) nDSM; (c) ADF-pixel (i.e., P); (d) ADF-feature built with the area attribute (i.e., F(area)); and (e) ADF-label built with the shadow primitive (i.e., L(shadow)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The average and total feature contributions of the multi-level ADFs are listed in Table 7. It is interesting to note that, according to the average feature contributions, ADF-pixel appears to be the most informative feature, followed by ADF-label. The contribution of an individual attribute in the ADF-feature is relatively small. However, it should be noted that the overall contribution of ADF-feature is significant, considering that it is composed of a series of APs that depict the urban scene characteristics from various perspectives. With respect to ADF-label, L(shadow) and L(build) both have a large influence on the classification, which shows the important role of shadows and building primitives for urban scene representation and class separation.

We further analyzed which ADF features are the most relevant for the various classes. However, ADF-feature has a significantly higher feature dimension, which complicates the comparison of the contribution of the different ADFs. To analyze the importance of ADF-feature regardless of the choice of scale, the feature contribution for each attribute was averaged over the different scales. Without loss of generality, dataset 1 was utilized for the class-specific feature contribution analysis (Fig. 14). The per-class contribution of the features varies according to the class categories. The spectral bands contribute most significantly to vegetation and shadow, which show distinctive spectral properties. However, as the categories of most interest in urban classification, i.e., buildings and roads, with similar spectral properties, they can only be effectively identified by the proposed ADFs. Specifically, ADF-pixel shows the largest contribution to discrimination of road and cottage. This is possibly because ADF-pixel is a raw form of angular difference with the least information loss. ADF-label plays a key role in identifying soil. In general, the analysis of the class-specific contributions indicates that it is necessary to include different ADF features which can complement each other in classifying different land covers, by characterizing the angular properties from different perspectives.

Moreover, it should be noted that ADFs can be built on different combinations of viewing angles, i.e. nadir and forward (NF), nadir and backward (NB), forward and backward (FB). Therefore, the influence and performance of different combinations of angles for classification were analyzed. Specifically, in this research, we attempt two approaches to combine the ADFs calculated from different pairs: stacking (i.e., three sets of ADFs were concatenated into one stacked feature vector), and the weighted average (i.e., three sets of ADFs were summed). The results show that the accuracy achieved by considering each pair separately is comparable to the accuracy obtained by the stacked ADF (Table 8). Moreover, the classification results of ADFs constructed on different angle combinations exhibit similar accuracies, signifying that all the multi-angle combinations contain rich and relevant angular information

Table 7
The contribution of the multi-level ADFs. The values in brackets correspond to the dimension of the features. See Table 1 for a description of the feature set symbols. AC = average contribution, TC = total contribution.

	ADF-pixel (3)	ADF-feature (96)				ADF-label (9)		
	P	F(area)	F(std)	F(diag)	F(iner)	L(back)	L(build)	L(shadow)
Dataset 1								
AC	0.020	0.012	0.012	0.013	0.011	0.021	0.027	0.018
TC	0.060	0.294	0.300	0.304	0.260	0.062	0.082	0.055
Dataset 2								
AC	0.021	0.011	0.014	0.017	0.009	0.018	0.023	0.023
TC	0.064	0.273	0.327	0.411	0.206	0.053	0.068	0.070
Dataset 3								
AC	0.027	0.015	0.012	0.017	0.016	0.025	0.017	0.023
TC	0.082	0.359	0.283	0.418	0.381	0.075	0.051	0.068

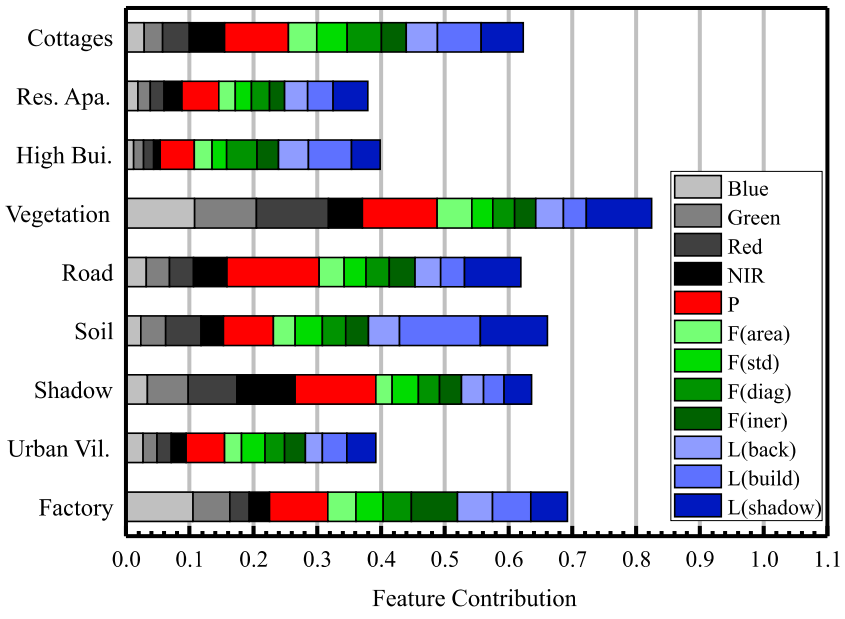

Fig. 14. The class-specific feature contributions of the multi-level ADFs. The symbols in the legend correspond to the specific features in Table 1.

Table 8
The classification results for multi-level ADFs built on different angle combinations. NB, NF, and BF denote the angle combinations of nadir and backward, nadir and forward, and backward and forward, respectively. 'Stacked' means the three pairs are concatenated in a feature vector. 'Weighted' denotes the weighted average of the three pairs. 'Equal-Weight' means the weights for NB, NF, and BF are equal, and in the case of 'Different-Weight', the weight for the NB, NF, and BF is 1, 1, and 2, respectively.

Dataset	Angle combinations			Stacked	Weighted	
	NB	NF	BF		Equal-Weight	Different-Weight
1	0.902 ± 0.010	0.905 ± 0.008	0.910 ± 0.006	0.923 ± 0.006	0.911 ± 0.007	0.914 ± 0.006
2	0.857 ± 0.016	0.848 ± 0.011	0.840 ± 0.005	0.861 ± 0.008	0.858 ± 0.012	0.858 ± 0.010
3	0.922 ± 0.016	0.919 ± 0.010	0.924 ± 0.010	0.915 ± 0.011	0.915 ± 0.010	0.913 ± 0.010

Table 9
Kappa values of classification without superpixel based refinement. KC means kappa coefficient. KCM denotes kappa coefficient between man-made constructions.

Feature set	Dataset 1		Dataset 2		Dataset 3	
	KC	KCM	KC	KCM	KC	KCM
S	0.574 ± 0.008	0.320 ± 0.008	0.558 ± 0.006	0.323 ± 0.010	0.664 ± 0.009	0.468 ± 0.013
S + AP	0.721 ± 0.007	0.569 ± 0.014	0.697 ± 0.006	0.540 ± 0.012	0.838 ± 0.011	0.787 ± 0.019
S + nDSM	0.774 ± 0.008	0.654 ± 0.011	0.725 ± 0.007	0.584 ± 0.010	0.791 ± 0.008	0.712 ± 0.014
S + P	0.663 ± 0.008	0.477 ± 0.012	0.639 ± 0.006	0.438 ± 0.007	0.781 ± 0.007	0.693 ± 0.011
S + F	0.816 ± 0.005	0.745 ± 0.005	0.703 ± 0.007	0.589 ± 0.013	0.797 ± 0.014	0.844 ± 0.011
S + L	0.740 ± 0.004	0.603 ± 0.007	0.704 ± 0.010	0.563 ± 0.011	0.797 ± 0.008	0.744 ± 0.009
S + PFL	0.842 ± 0.005	0.780 ± 0.007	0.725 ± 0.010	0.627 ± 0.012	0.802 ± 0.016	0.862 ± 0.009

for interpreting urban scenes. In addition, it can be seen that the results achieved by the stacked ADFs are slightly better than those by the weighted ADFs. Moreover, the weights have little influence on the classification performances.

3.5. Analysis of superpixel-based refinement

In order to investigate the effectiveness of superpixel based refinement, the classification results for the pixel-wise ADFs, as well as other features (e.g., AP and nDSM), are presented (Table 9). It can be observed that the ADFs can produce the highest overall accuracy in most cases, except for the KC in dataset 2 and 3, where the ADFs still give close accuracies to the highest value. Moreover, in terms of the KCM (i.e., kappa coefficient between the man-made classes), the best feature set is always the combination of spectral bands and the multi-level ADFs (S + PFL), which once again confirms the effectiveness of the proposed ADFs over other features when classifying man-made objects. By reference to the classification results of the superpixel feature sets (Section 3.3), it can be observed that by incorporating superpixel refinement, the ADFs can achieve greater accuracy increments compared to other features (e.g., nDSM, APs), which indicates that the spatial refinement can further enhance the performance of ADFs, probably by reducing the salt-and-pepper effect.

3.6. Effects of the superpixel segmentation scale on the classification results

Since the superpixel processing is an effective refinement for ADFs, the relationship between the classification results and the superpixel segmentation scale was further analyzed. Fig. 15 presents the effect of the segmentation size on the performance of the S + PFL feature set (i.e., the combination of spectral bands and the multi-level ADFs). In this figure, the horizontal axis indicates the number of superpixels in the image, and the vertical axis shows the corresponding average kappa of the 10 independent tests. It can be seen that with the increase of the superpixel segment number, the classification accuracy quickly increases before reaching a maximum, and then becomes relatively stable. This trend can be expected, since the segmentation scale should be large enough so that sufficient local angular information is captured but, at the same time, the scale should produce a reasonable over-segmentation, so as to depict the subtle angular differences and better describe the angular properties of small or narrow objects (e.g., cottages and roads) in the heterogeneous urban environment.

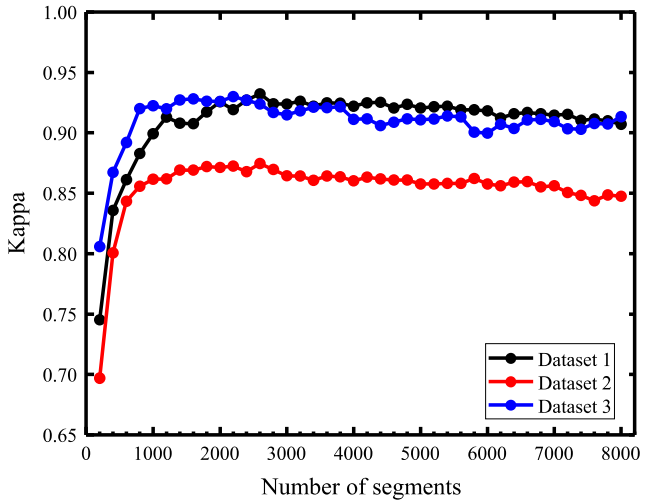

Fig. 15. Relationships between the kappa values of the classification results and the segmentation scales.

4. Conclusion

Multi-angle high-resolution images can provide abundant angular information in both the spectral and spatial domains. In this study, we have proposed novel multi-level angular difference features (ADFs) for urban scene classification. The basis of ADFs is that multi-angle observations provide a source of signal variability on the urban structures, which in turn reveals the material and structural characteristics of the urban objects. In greater detail, the proposed method is composed of two main steps: (1) multi-level angular feature extraction, which reveals the angular properties at three levels (i.e., pixel, feature, and label levels); and (2) superpixel-based refinement, in which the ADF features are refined based on superpixel segmentation, for the purpose of alleviating

the effects of noise and representing the main angular characteristics within a local area. The experimental analysis was carried out on three ZY-3 multi-angle datasets acquired over complex urban scenes. Our results indicate that ADFs can characterize urban classes and give promising results for detailed urban classification. The classification results are far better than using the spectral bands alone. More importantly and interestingly, the proposed ADF features can outperform the state-of-the-art nDSM, as the latter can be subject to stereo image matching errors, especially for high buildings in dense and heterogeneous urban scenes. In addition, ADFs also provided much better classification performances than the state-of-the-art spatial and structural features, i.e., morphological attribute profiles, which have been proven to be very effective in high-resolution remote sensing urban classification tasks (Bhangale et al., 2017; Licciardi et al., 2012; Mura et al., 2010). Furthermore, the results indicate the superiority of the proposed ADFs in discriminating spectrally similar man-made classes, including roads and various types of buildings such as high buildings, urban villages, and residential apartments. On the one hand, ADFs contain implicit height information and can help to discriminate the spectrally similar classes with different height characteristics. On the other hand, ADFs can also offer the possibility to distinguish between man-made classes with similar heights but with different spatial structures, such as residential apartments and urban villages.

The proposed ADFs could still be further improved in certain aspects. For instance, the current framework only considers the vector stacking approach for the integration of the multi-level ADFs, which does not necessarily result in the optimal performance for multi-feature classification (Huang and Zhang, 2013). In addition, the ADF features have the potential to estimate building height, which will be investigated in our future research.

Acknowledgements

The research was supported by the National Natural Science Foundation of China under Grants 41522110 and 41771360, the Hubei Provincial Natural Science Foundation of China under Grant 2017CFA029, and the National Key Research and Development Program of China under Grant 2016YFB0501403.

References

Aguilar, M.Á., Saldaña, M., Aguilar, F.J., 2014. Generation and quality assessment of stereo-extracted DSM from GeoEye-1 and WorldView-2 imagery. IEEE Trans. Geosci. Remote Sens. 52, 1259–1271.

Bhangale, U., Durbha, S.S., King, R.L., Younan, N.H., Vatsavai, R., 2017. High performance GPU computing based approaches for oil spill detection from multi-temporal remote sensing data. Remote Sens. Environ.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bruzzone, L., Member, S., Carlin, L., Member, S., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans. Geosci. Remote Sens. 44, 2587–2600.

Chopping, M., Moisen, G.G., Su, L., Laliberte, A., Rango, A., Martonchik, J.V., Peters, D. P.C., 2008. Large area mapping of southwestern forest crown cover, canopy height, and biomass using the NASA Multiangle Imaging Spectro-Radiometer. Remote Sens. Environ. 112, 2051–2063.

Dalla Mura, M., Atli Benediktsson, J., Waske, B., Bruzzone, L., 2010. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. Int. J. Remote Sens. 31, 5975–5991.

Diner, D.J., Braswell, B.H., Davies, R., Gobron, N., Hu, J., Jin, Y., Kahn, R.A., Knyazikhin, Y., Loeb, N., Muller, J.P., Nolin, A.W., Pinty, B., Schaaf, C.B., Seiz, G., Stroeve, J., 2005. The value of multiangle measurements for retrieving structurally and radiatively consistent properties of clouds, aerosols, and surfaces. Remote Sens. Environ. 97, 495–518.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30, 328–341.

Huang, X., Liu, H., Zhang, L., 2015. Spatiotemporal detection and analysis of urban villages in mega city regions of china using high-resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 53, 3639–3657.

Huang, X., Lu, Q., Zhang, L., 2014. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. ISPRS J. Photogramm. Remote Sens. 90, 36–48.

Huang, X., Wen, D., Li, J., Qin, R., 2017a. Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery. Remote Sens. Environ. 196, 56–75.

Huang, X., Yuan, W., Li, J., Zhang, L., 2017b. A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 654–668.

Huang, X., Zhang, L., 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 51, 257–272.

Huang, X., Zhang, L., Member, S., 2012. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 161–172.

Johnson, B., Xie, Z., 2013. Classifying a high resolution image of an urban area using super-object information. ISPRS J. Photogramm. Remote Sens. 83, 40–49.

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. 177, 89–100.

Lee, T., Kim, T., 2015. Automatic building height extraction by volumetric shadow analysis of monoscopic imagery. Int. J. Remote Sens. 1161.

Li, J., Zhang, H., Zhang, L., 2015. Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. Geosci. Remote Sensing, IEEE Trans. 53, 5338–5351.

Li, X., Chen, W., Cheng, X., Wang, L., 2016. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. Remote Sens. 8, 514.

Licciardi, G.A., Villa, A., Mura, M.D., Bruzzone, L., Chanussot, J., Benediktsson, J.A., 2012. Retrieval of the height of buildings from worldview-2 multi-angular imagery using attribute filters and geometric invariant moments. Sel. Top. Appl. Earth Obs. Remote Sensing, IEEE J. 5, 71–79.

Liu, M.Y., Tuzel, O., Ramalingam, S., Chellappa, R., 2011. Entropy rate superpixel segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2097–2104.

Longbotham, N., Chaapel, C., Bleiler, L., Padwick, C., Emery, W.J., Pacifici, F., 2012. Very high resolution multiangle urban classification analysis. IEEE Trans. Geosci. Remote Sens. 50, 1155–1170.

Lucht, W., Schaaf, C.B., Strahler, A.H., 2000. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. IEEE Trans. Geosci. Remote Sens. 38, 977–998.

Marpu, P.R., Pedergnana, M., Mura, M.D., Benediktsson, J.A., Bruzzone, L., 2013. Automatic generation of standard deviation attribute profiles for spectral – spatial classification of remote sensing data. IEEE Geosci. Remote Sens. Lett. 10, 293–297.

Matasci, G., Longbotham, N., Pacifici, F., Kanevski, M., Tuia, D., 2015. Understanding angular effects in VHR imagery and their significance for urban land-cover model portability: a study of two multi-angle in-track image sequences. ISPRS J. Photogramm. Remote Sens. 107, 99–111.

Mura, M.D., Member, S., Benediktsson, J.A., 2010. Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans. Geosci. Remote Sens. 48, 3747–3762.

Pacifici, F., Chini, M., Emery, W.J., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sens. Environ. 113, 1276–1292.

Pasher, J., King, D.J., 2010. Multivariate forest structure modelling and mapping using high resolution airborne imagery and topographic information. Remote Sens. Environ. 114, 1718–1732.

Pesaresi, M., Benediktsson, J.A., 2001. A new approach for the morphological segmentation of high-resolution satellite imagery. IEEE Trans. Geosci. Remote Sens. 39, 309–320.

Puttonen, E., Suomalainen, J., Hakala, T., Peltoniemi, J., 2009. Measurement of reflectance properties of asphalt surfaces and their usability as reference targets for aerial photos. IEEE Trans. Geosci. Remote Sens. 47, 2330–2339.

Qin, R., 2016. Rpc Stereo Processor (Rsp) – a Software Package for Digital Surface Model and Orthophoto Generation From Satellite Stereo Imagery. ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. III-1, 77–82.

Qin, R., 2014. Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery. ISPRS J. Photogramm. Remote Sens. 96, 179–192.

Qin, R., Fang, W., 2014. A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization. Photogramm. Eng. Remote Sensing 80, 37–47.

Tian, J., Cui, S., Reinartz, P., 2014. Building change detection based on satellite stereo imagery and digital surface models. IEEE Trans. Geosci. Remote Sens. 52, 406–417.

Tian, J., Reinartz, P., Angelo, P., Ehlers, M., 2013. Region-based automatic building and forest change detection on Cartosat-1 stereo imagery. ISPRS J. Photogramm. Remote Sens. 79, 226–239.

Wen, D., Huang, X., Zhang, L., Benediktsson, J.A., 2016. A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation. IEEE Trans. Geosci. Remote Sens. 54, 609–625.

Xiao, J., Gerke, M., Vosselman, G., 2012. Building extraction from oblique airborne imagery based on robust façade detection. ISPRS J. Photogramm. Remote Sens. 68, 56–68.