SCIENCE CHINA
Earth Sciences

•REVIEW•

April 2020 Vol.63 No.4: 463–475
https://doi.org/10.1007/s11430-019-9547-x

Advances in urban information extraction from high-resolution remote sensing imagery

Jianya GONG[1], Chun LIU[2*] & Xin HUANG[1,2†]

[1] *School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;*
[2] *State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China*

Received April 26, 2019; revised September 6, 2019; accepted November 8, 2019; published online December 25, 2019

Abstract The study of urban area is one of the hottest research topics in the field of remote sensing. With the accumulation of high-resolution (HR) remote sensing data and emerging of new satellite sensors, HR observation of urban areas has become increasingly possible, which provides us with more elaborate urban information. However, the strong heterogeneity in the spectral and spatial domain of HR imagery brings great challenges to urban remote sensing. In recent years, numerous approaches were proposed to deal with HR image interpretation over complex urban scenes, including a series of features from low level to high level, as well as state-of-the-art methods depicting not only the urban extent, but also the intra-urban variations. In this paper, we aim to summarize the major advances in HR urban remote sensing from the aspects of feature representation and information extraction. Moreover, the future trends are discussed from the perspectives of methodology, urban structure and pattern characterization, big data challenge, and global mapping.

Keywords High-resolution, Urban remote sensing, Feature extraction, Land use/land cover classification, Change detection

Citation: Gong J, Liu C, Huang X. 2020. Advances in urban information extraction from high-resolution remote sensing imagery. Science China Earth Sciences, 63: 463–475, https://doi.org/10.1007/s11430-019-9547-x

1. Introduction

Urban area is the core of human habitation, and is also the most active region for social and economic activities. According to the United Nations, the proportion of global urban residents increased from 30% to 55% between 1950 and 2018, and it is predicted to reach 68% by 2050. About 90% of this growth will take place in Africa and Asia, where the level of urbanization is relatively low at present (United Nations, 2018). Although cities account for only a small share (< 3%) of the Earth's land surface, they significantly impact both natural and human systems from regional to global scales (Gamba and Herold, 2009). The rapid urban expansion is accompanied by the disappearance of the surrounding cultivated land, forest, and water areas, which brings a series of resource, environmental, and ecological problems (Huang J K et al., 2007; Gong et al., 2012; Pesaresi et al., 2015).

Urban remote sensing, as one of the most important branches in the field of remote sensing, mainly refers to the use of remote sensing technologies to obtain urban information for monitoring, understanding, and predicting the various urban phenomena, and to support decision making in urban planning, disaster response, and sustainable development. In recent decades, with the considerable developments of aerospace technology, a number of high-resolution (HR) satellites have been launched (Table 1). The spatial resolution of civilian/commercial remote sensing imagery has achieved meter and even sub-meter levels. Through different observation modes, such as satellite networking and along-track or cross-track imaging, many of the HR satellites (e.g.,

* Corresponding author (email: liuchun919@whu.edu.cn)
† Corresponding author (email: xhuang@whu.edu.cn)

Table 1 Typical high-resolution satellites[a)]

Satellite	Launch time	Country	Bands and spatial resolution	Swath width	Revisit cycle
IKONOS-2	1999-09	America	a) PAN: nadir 0.82 m; off-nadir 1 m b) 4 MS (blue, green, red, near-infrared): nadir 3.20 m; off-nadir 4 m	11.3 km	3 days
QuickBird-2	2001-10	America	a) PAN: 0.61 m b) 4 MS (blue, green, red, near-infrared): 2.40 m	16.5 km	1–3.5 days
Cartosat-1	2005-05	India	PAN: 2.50 m	26 km	5 days
Beijing-1	2005-10	China	a) PAN: 4 m b) 3 MS (green, red, near-infrared): 32 m	a) PAN: 24 km b) MS: 600 km	3–7 days
ALOS-PRISM	2006-01	Japan	PAN: 2.50 m	35 km	5 days
EROS-B	2006-04	Israel	PAN: 0.70 m	7 km	2–10 days
Cartosat-2	2A: 2007-01 2B: 2008-04	India	PAN: 0.80 m	9.6 km	4 days
TianHui-1	01: 2010-08 02: 2012-05 03: 2015-10	China	a) PAN: 5 m b) HR PAN: 2 m c) 4 MS (blue, green, red, near-infrared): 10 m	60 km	1 day
ZiYuan-3	01: 2012-01 02: 2016-05	China	a) PAN: nadir 2.10 m; off-nadir 3.50 m (01), 2.50 m (02) b) 4 MS (blue, green, red, near-infrared): 5.80 m	50 km	5 days
SPOT-6/7	SPOT-6:2012-09 SPOT-7:2014-06	France	a) PAN: 1.50 m b) 4 MS (blue, green, red, near-infrared): 6 m	60 km	1–5 days
GaoFen-1	2013-04	China	a) PAN: 2 m b) 4 MS (blue, green, red, near-infrared): 8 m and 16 m	68 km with two HR cameras and 830 km with four wide-field imager	≤ 4 days
WorldView-3	2014-08	America	a) PAN: nadir 0.31 m; off-nadir 0.34 m b) 8 MS (red, red edge, coastal, blue, green, yellow, near-infrared1, near-infrared2): nadir 1.24 m; off-nadir 1.38 m c) 8 short-wave infrared: nadir 3.70 m; off-nadir 4.10 m d) 12 CAVIS: nadir 30 m	13.1 km	1–4.5 days
GaoFen-2	2014-08	China	a) PAN: 0.81 m b) 4 MS (blue, green, red, near-infrared): 3.24 m	45 km	≤ 4 days
SuperView-1	01/02: 2016-12 03/04: 2018-01	China	a) PAN: 0.5 m b) 4 MS (blue, green, red, near-infrared): 2 m	12 km	1 day with 4 satellites

a) PAN, panchromatic band; MS, multispectral bands

QuickBird, Cartosat-1/2, ZiYuan-3 (ZY-3), and SPOT-6/7) have the ability of stereo mapping. HR imagery can substantially reduce the phenomenon of mixed-pixels with enhanced spatial details of ground objects. Meanwhile, multi-angle observation is able to provide three-dimensional (3D) information, which increase the dimension of urban information extraction (Huang X et al., 2018; Huang et al., 2017a; Peng et al., 2017). Nevertheless, the problems such as spectral heterogeneity, shadow, occlusion, and disparity are distinct in HR images, especially in urban settings, which presents new challenges for urban information extraction. Hence the traditional methods that rely purely upon spectral characteristics may be insufficient to tackle these problems (Huang et al., 2007b; Peng et al., 2015).

Compared to natural surfaces, urban areas have more distinct textural and structural variations in HR images, due to the inclusion of artificial surfaces such as buildings and roads. Given these characteristics, many studies have incorporated both spectral and spatial features to improve the interpretation accuracy of HR images over urban areas. In this paper, we mainly focus on introducing state-of-the-art urban features, as well as urban information extraction based on HR remote sensing imagery from the following aspects: (1) detection of urban targets, e.g., buildings, roads, impervious surfaces, urban vegetation, and water bodies; (2) classification of urban land use/land cover, such as urban scene recognition and functional zone mapping; (3) change detection, i.e., dynamic monitoring of the urban landscape; and (4) urban ecology and climate, e.g., urban heat island and ecosystem service assessment. A framework of HR urban remote sensing is illustrated in Figure 1. Finally, the future research trends in HR urban remote sensing are prospected.

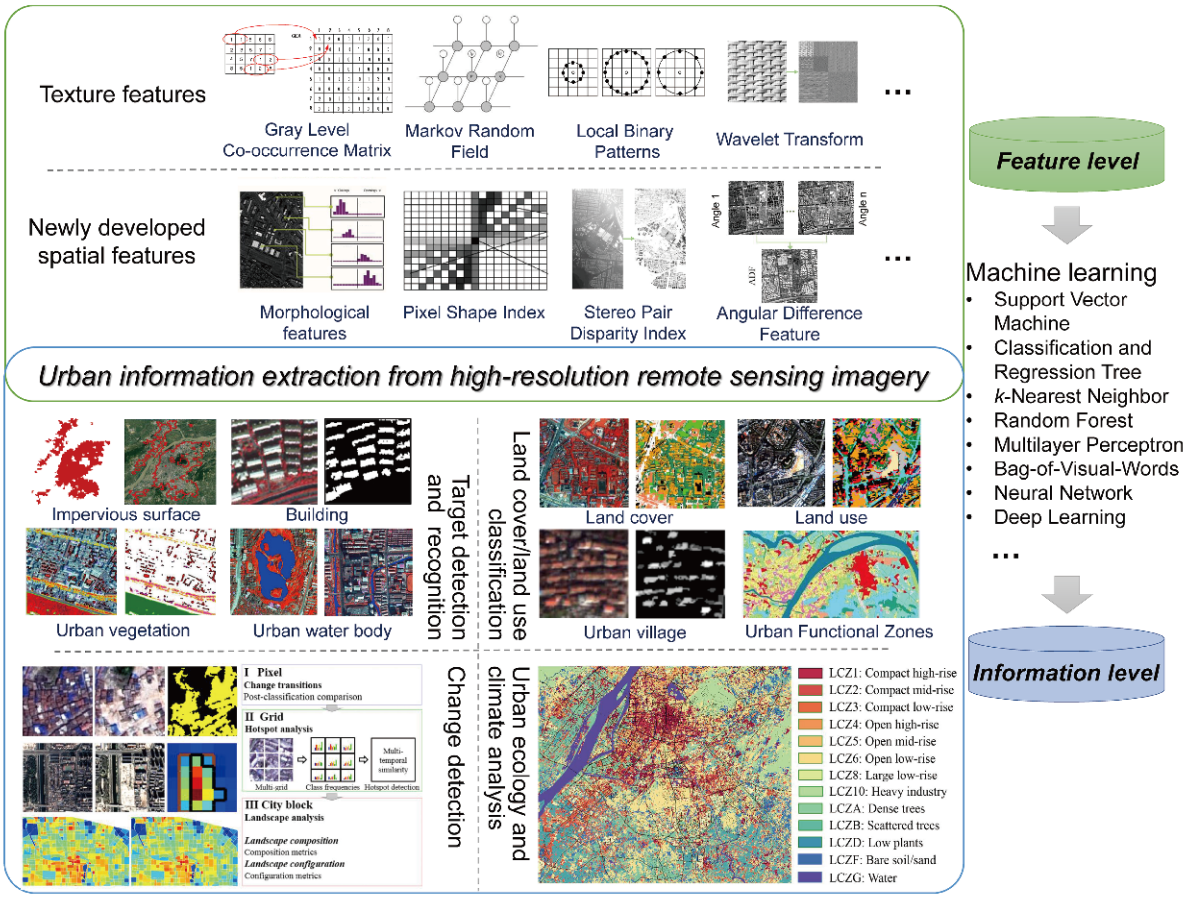

Figure 1 The framework of urban information extraction from high-resolution remote sensing imagery.

2. Urban feature extraction from high-resolution remote sensing imagery

The widely used texture features are calculated based mainly on statistical and spatial-frequency domain analysis, e.g., the gray-level co-occurrence matrix (Haralick et al., 1973), wavelet transform (Mallat, 1989), and local binary patterns (Ojala et al., 1996). Moreover, a series of planar and stereo features, e.g., the pixel shape index (Zhang et al., 2006), PanTex (Pesaresi et al., 2008), morphological features (Pesaresi and Benediktsson, 2001), and angular features (Huang X et al., 2018), were specially designed for HR images to characterize urban environments. In addition, the recent advances in the domain of machine learning (e.g., deep learning) show strong abilities for high-level feature representation, which show promising avenues to address complex HR urban remote sensing problems (Zhu et al., 2017).

2.1 Textural features

2.1.1 Gray-level co-occurrence matrix

The gray-level co-occurrence matrix (GLCM) is a classical statistical texture extraction method. A series of statistical measures, e.g., homogeneity (HOM), contrast (CON), angular second moment (ASM), and entropy (ENT), were defined to characterize the co-occurrence matrix to reflect the grayscale changes and textural features of the image (Haralick et al., 1973). For HR urban remote sensing, Puissant et al. (2005) and Su et al. (2008) combined the spectral information and the GLCM textures of HR images for urban classification, demonstrating the effectiveness of the GLCM features to complement the spectral information. To overcome the influence of the "window effect" of the spatial features, Huang et al. (2007a) presented an adaptive multi-scale feature fusion method, which automatically selects the best window size according to the spectral and boundary information, and integrates the multi-scale features to extract ground objects of different sizes. Standard GLCM textures are calculated based on one band, therefore the first principal component or one of the bands is widely employed when dealing with multi/hyperspectral imagery (Pacifici et al., 2009). In order to exploit all the useful information, Huang et al. (2014b) proposed a multichannel GLCM calculation method via image coding techniques to extract the synthesized texture features from multi/hyperspectral bands.

2.1.2 Wavelet transform

Wavelet transform (WT), which was developed in the 1980s (Mallat, 1989), has been widely applied in texture analysis. The standard WT is based on orthogonal wavelet basis, aiming to obtain the multi-scale information, and extract the high- and low-frequency features of each layer. Myint et al. (2004) compared the WT, GLCM, spatial autocorrelation, and fractal approaches for extracting urban textures from HR images, suggesting that WT was more effective than the other methods. Ouma et al. (2006) constructed multi-scale textural features through WT, and combined them with the spectral features to extract urban trees from QuickBird imagery. The 3D-WT processes the multispectral imagery as a cube and extracts spectral and spatial information simultaneously, hence it provides a more adequate feature representation for multi/hyperspectral images (Guo et al., 2014; Huang and Zhang, 2012b; Li Q et al., 2017; Qian et al., 2013; Yoo et al., 2009). For instance, Yoo et al. (2009) constructed the urban complexity index (UCI) based on 3D-WT, in order to discriminate complex urban areas and natural surfaces. Accordingly, Huang and Zhang (2012b) proposed the multi-scale UCI (M-UCI) to further enhance the performance over urban and suburban areas.

2.1.3 Local binary patterns

The local binary patterns (LBP) descriptor was proposed by Ojala et al. (1996) to characterize the local textural features of an image, and has been widely used in the fields of image registration, target tracking, etc. The basic idea is to compare the gray value of the center pixel with its neighboring pixels, so that a set of binary codes of the center pixel can be obtained. Furthermore, improved versions of LBP were developed with illumination and rotation invariant properties (Ojala et al., 2002). LBP has been widely applied to characterize the complex urban textures in HR images. For example, Song et al. (2010) combined the LBP and spectral features to classify HR imagery, and achieved higher accuracy with the addition of LBP. Musci et al. (2013) extracted the LBP texture features of urban areas from QuickBird and IKONOS images for land cover classification, and obtained better results compared to the GLCM features. Li W et al. (2015) extracted local features of HR hyperspectral imagery based on LBP and obtained good classification results.

2.2 Spatial features for high-resolution imagery

2.2.1 Pixel shape index

The pixel shape index (PSI) is a spatial index describing local shape features (Zhang et al., 2006). The idea of the algorithm is to define a set of anisotropic direction lines radiating from the central pixel to its surrounding pixels. Under the constraints of the spectrum and the space, the number of neighboring pixels with spectral similarity to the central pixel along each direction line is counted as the length of this direction line. The length values of all the directional lines then constitute a histogram, and the mean value of the histogram is finally defined as the PSI value of the center pixel. PSI can detect more than 20 directions, which makes up for the insufficient scanning directions of the GLCM and sufficiently explores the spatial context features in HR images. The authors combined the spectral information and PSI to conduct urban classification, and demonstrated the superiority of PSI by comparing it with texture features such as the GLCM and WT. Furthermore, the structural feature set (SFS) was defined as an extension of PSI based on the histograms of the direction lines, including six operators such as length-width ratio, weighted PSI, and variance (Huang et al., 2007b). PSI is more efficient for urban structures than natural surfaces (Huang and Zhang, 2012b), so it is often used as a local spatial feature together with other features to classify complex urban areas (Li Q et al., 2017; Zhang et al., 2013).

2.2.2 PanTex

PanTex (Pesaresi et al., 2008) is a rotation-invariant built-up presence index computed based on the GLCM. Specifically, PanTex extracts the GLCM from panchromatic imagery by using the offset vectors in 10 directions and the CON measure. Only when the texture values in all the directions are high can the pixel be considered as a built-up pixel. Therefore, the minimum value of all the directions is taken as the PanTex value of the pixel. Since PanTex has false alarms from scattered trees, high-brightness bare soil, rocks, etc., it was further improved by utilizing the normalized difference vegetation index (NDVI) and morphological filtering, which increased the accuracy by 2.44% and 20.76%, respectively (Pesaresi and Gerhardinger, 2011). This index has also been employed by the European Commission's Joint Research Centre in the Global Human Settlement Layer (GHSL) project to extract large-scale built-up areas in Europe from HR images (Florczyk et al., 2016; Pesaresi et al., 2011, 2013).

2.2.3 Morphological features

Morphological feature refers to the spatial structure features of an image obtained through basic morphological operations (e.g., erosion and dilation, opening and closing) with the structural element (SE). Pesaresi and Benediktsson (2001) proposed the multi-scale morphological profiles (MPs) and applied it to HR image classification successfully. Since the strength of the morphological feature response is determined by the SE radius and the local structure size, MPs extracts multi-scale bright and dark structures of the image with a set of SEs of different sizes (Pesaresi and Benediktsson, 2001). To detect the morphological features of different scales more effectively, the derivative morpholo-

gical profiles (DMPs) were further defined as the sequential differences of the MPs between two adjacent scales. MPs and DMPs have been utilized for HR image processing in many studies, achieving satisfactory results (Benediktsson et al., 2003; Chanussot et al., 2006; Chini et al., 2009; Tuia et al., 2009). In order to apply MPs to hyperspectral imagery, Benediktsson et al. (2005) proposed extended MPs (EMPs) by using the principal components as the base images to calculate the MPs. Since the spectral information of hyperspectral imagery was not sufficiently explored by EMPs, Fauvel et al. (2008) further fused the spectral information and MPs for urban classification. Dalla Mura et al. (2010b) proposed morphological attribute profiles (APs) to obtain the morphological attributes, such as area and standard deviation. Similarly, extended APs (EAPs) have also been proposed and applied to hyperspectral imagery (Dalla Mura et al., 2010a). Huang et al. (2014a) investigated the influence of the different base image strategies for MPs, and constructed the multiple morphological profiles (MMPs) for hyperspectral image classification. Ghamisi et al. (2015) reviewed and summarized the different morphological features. Since DMPs only consider the difference between adjacent scales and ignore the cross-scale information, Huang et al. (2016) developed generalized DMPs (GDMPs) to obtain the difference between arbitrary scales, which can better describe the multi-scale property of complex urban scenes.

In recent years, the morphological building index (MBI) was proposed for unsupervised building extraction (Huang and Zhang, 2011). MBI describes the spectral and spatial features of buildings, such as brightness, structure, and anisotropy, based on morphological operators. Since buildings and their shadows have similar structures and are spatially adjacent, Huang and Zhang (2012a) also constructed the morphological shadow index (MSI) for the automatic detection of building shadows. Experiments were conducted on GeoEye-1, IKONOS, and WorldView-2 images of Wuhan, Hangzhou, and Washington DC, which confirmed the superiority of this algorithm. The MBI can effectively detect buildings from HR remote sensing images, but it may also induce false alarms from bright bare ground and roads. In order to further strengthen the efficacy of the MBI in suburban, mountainous, and agricultural areas, Huang et al. (2017b) proposed a post-processing framework and obtained more accurate building extraction results by applying spectral, shadow, and shape constraints successively on the initial MBI results, to filter out commission errors such as bright vegetation, soil, playgrounds, and roads. In addition, a number of studies combined the MBI, MSI, and spectral features for urban classification, built-up area extraction, change detection, etc. (Huang et al., 2017a; Li Q et al., 2017; Wen et al., 2016; Zhang and Huang, 2018; Liu et al., 2019), confirming the effectiveness of these indices in urban feature extraction.

2.2.4 Stereo features

HR stereo observation satellites can acquire images from multiple viewing angles, which can be employed to produce a digital surface model (DSM), and applied to height estimation and 3D reconstruction. In recent years, many researchers have found that the application of stereo features, e.g., a DSM, to urban classification can increase the separability of land cover types such as building structures and roads (Longbotham et al., 2012; Peng et al., 2015; Qin, 2014; Qin and Fang, 2014; Tian et al., 2014). In terms of the fact that the elevation of roads, grassland, and bare land are relatively consistent while buildings have abrupt changes to the surrounding ground surface, Peng et al. (2017) proposed the stereo pair disparity index (SPDI) to extract built-up areas by describing the intensity of the elevation change. This method firstly generates disparity maps from multi-view images by using stereo matching algorithm (e.g., semiglobal matching), and then calculates the disparity gradients with multi-directional offset vectors. Finally, the built-up areas can be extracted from the gradient features. However, in dense urban areas, the quality of a DSM or disparity map is susceptible to many factors such as the base-height ratio, the image matching algorithm, and occlusions, which further influence the accuracy of urban information extraction. In order to fully exploit the angular information of HR stereo imagery, Huang X et al. (2018) proposed the angular difference feature (ADF) to describe the dissimilarities between different viewing images from the pixel, feature, and label levels. The ZY-3 multi-angle images were utilized in the experiments, and the results indicated that the joint use of ADF and spectral features can significantly improve urban classification accuracy, and the ADF can help to distinguish complex artificial structures with spectral similarity (e.g., roads, high-rise houses, urban villages, and residential buildings).

Demonstrations of some typical spatial features are displayed in Figure 2. The eight textural, structural and stereo features were extracted from ZY-3 imagery in central Shanghai. Note that the stereo features (DSM and ADF) were generated from multi-view images, while the remaining features were calculated from the nadir panchromatic image.

2.3 Deep learning-based feature representation

The previous features were designed based mainly on domain-specific knowledge. Nevertheless, their discriminative ability may still be limited to tackle complex or large-scale urban analysis. In recent years, deep learning-based methods have been increasingly investigated in the field of remote sensing, such as land use classification (Huang B et al., 2018), scene recognition (Li Y C et al., 2017), and urban expansion monitoring (He et al., 2019). Unlike handcrafted features, deep neural networks (DNNs) can directly extract

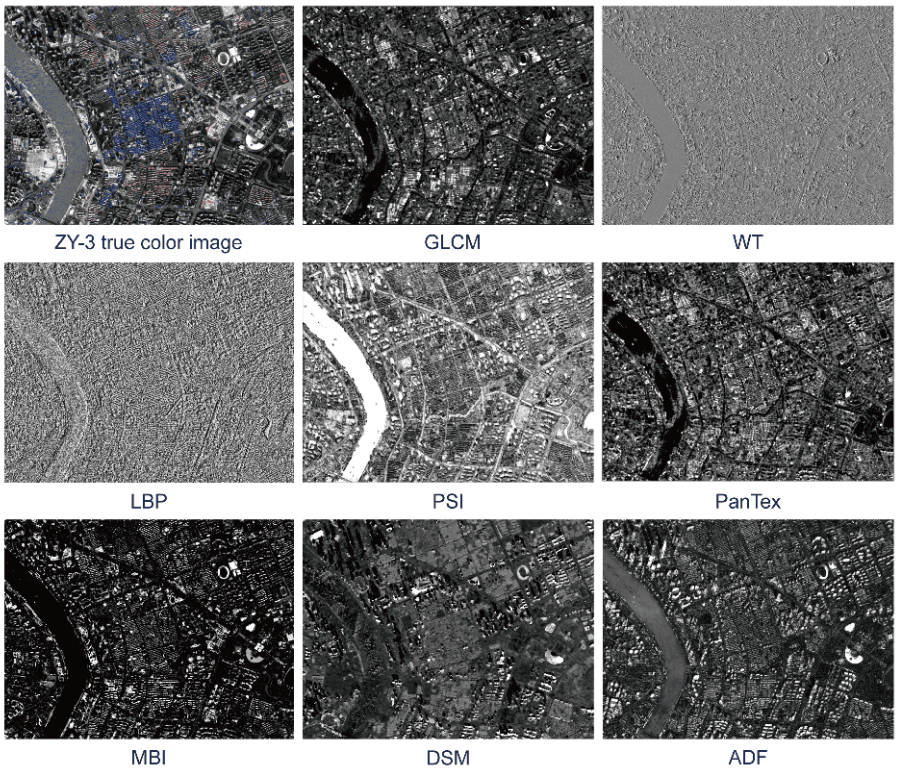

Figure 2 Illustrations of the typical spatial features extracted from ZY-3 stereo imagery in central Shanghai.

high-level features from data based on neural networks with deep architecture in an automated fashion (Reichstein et al., 2019; Zhu et al., 2017). Sufficient learning of a deep neural network is very difficult and costly since it relies on massive samples and computational resources. The pre-training and fine-tuning strategies are widely adopted (Li et al., 2019). For instance, Marmanis et al. (2016) directly extracted the features by transferring the convolutional neural networks (CNNs) pre-trained on natural images (e.g., ImageNet) for HR remote sensing scene classification. Hu et al. (2015) proposed two scenarios to generate global feature representations for HR image scenes by using the CNNs, which achieved remarkable classification accuracies on two public HR datasets. Nogueira et al. (2017) adapted the existing CNN architectures to HR image classification by fine-tuning with a small sample set of HR images, demonstrating the effectiveness of fine-tuned networks for performance improvements compared with using only the pre-trained networks.

3. Urban information extraction from high-resolution remote sensing imagery

3.1 Target recognition

Target recognition is one of the major tasks of information extraction. Urban targets, mainly including buildings, built-up, impervious surface, roads, vegetation, and water bodies, are of great interest to researchers and city planners since they are vital indicators of human distribution, economic development and city's livability (Weng, 2012). With the availability of HR data, these detailed targets are now possible to be recognized. Numerous approaches have been proposed to extract the thematic information of the urban target of interest. In this section, some representative methods are briefly introduced.

3.1.1 Impervious surfaces, buildings, and roads

Impervious surface, or other similar semantic abstractions such as artificial surfaces, urban footprint, human settlement, and built-up areas, are the major components of urban land. A large amount of researches have been focused on impervious surface extraction from coarse and medium resolution images over large areas (Chen et al., 2016; Li X et al., 2015; Schneider et al., 2010; Weng, 2012). HR imagery can significantly reduce the problems of mixed-pixels and blurred boundaries but also brings new challenges, such as the confusion between different ground objects (e.g., bright impervious layers and bare soil, dark impervious layers and water) and the problems brought by shadows. To address these issues, Hu and Weng (2011) proposed an object-based method to extract impervious surfaces for residential and central business district (CBD) areas from IKONOS imagery, which obtained high accuracies and precise feature

boundaries. The attributes considered in the rule set included spectral, spatial, and textural features, which were used to comprehensively describe the properties of impervious surfaces. To tackle the underestimation of impervious area caused by shadows, Zhang and Huang (2018) presented a two-stage object-based framework by integrating multiple features. Specifically, the spatial relationships of different land covers (e.g., the distance between shadow and vegetation) were further considered in the second stage to extract more detailed impervious surface information in shaded areas. Since synthetic aperture radar (SAR) is sensitive to the structural or geometric features of built environment, integration of optical and SAR images at pixel, feature, and decision levels are explored to improve the estimation of impervious layers (Shao et al., 2016; Zhang et al., 2014). More recently, Liu et al. (2019) proposed a framework for built-up extraction by characterizing building properties from structural, textural and vertical aspects, demonstrating the complementation of multi-feature fusion. It was also suggested that the employ of stereo features can effectively reduce the omission errors of dark built-up areas.

For building extraction, knowledge-based automatic approaches, which construct the building extraction rules by considering the spectral, shape, texture, and spatial characteristics of buildings, are commonly adopted. Generally, the sizes of most urban buildings are within a certain range; buildings have vertical structures and high reflectivity; buildings and their shadows are spatially adjacent (Huang et al., 2017b; Pesaresi et al., 2008). These characteristics have been extensively exploited to infer the existence of buildings in HR images, by characterizing the brightness, local contrast, shape, height, and the spatial relationships between buildings and their shadows with the information of solar incident angle. For instance, in Ok et al. (2013), the shadow areas in the image were firstly extracted, then the candidate buildings were detected according to the spatial directional relationship of shadows and buildings. Since this method highly depend on initial shadow extraction accuracy, post-processing and optimization of the shadow were conducted in Ok (2013) in order to further improve building detection accuracy. The MBI and MSI discussed in Section 2.2.3 were also employed in many studies for automatic building detection (Huang and Zhang, 2012a). When dealing with complex urban scenes, supervised approaches are widely used for building extraction. For example, the widely used object-oriented multi-feature fusion methods, for which the selection of the segmentation scale and feature extraction are the two important steps. The optimization of the segmentation scale ensures the independence and integrity of the extracted buildings (Tian and Chen, 2007), while spectral, texture, shape and stereo features are commonly applied to depict the within-object information (Fauvel et al., 2008; Liu C et al., 2017; Zhang et al., 2017).

Road detection from HR images is a challenging task because of the spectral and spatial complexity of road networks. Similar to building extraction, the spectrum, shape, and topology properties of road are extensively considered. For example, roads generally present a curvilinear shape; the width of a road does not change drastically; and roads have apparent edge lines and crossings, and often form networks. Based on this knowledge, Huang and Zhang (2009) proposed an object-based method to extract road centerlines from HR imagery, by integrating multi-scale spectral-structural information based on support vector machine (SVM). Poullis (2014) presented an automatic road extraction method that combined tensor encoding, Gabor filter, and Graph-Cuts for the inference of road features. Shanmugam and Kaliaperumal (2015) proposed the active deformable model for semi-automatic road extraction. This method first selected the road seed points manually, from which the propagation started and was restrained by the width and color of the road. Then, the interconnected road networks can be extracted. Sghaier and Lepage (2016) applied the beamlet transform in road border detection to find the most suitable scale for each road in HR image. The SFS structural feature, mathematical morphology, and Canny detector were also employed in preprocessing steps for edge candidate selection.

3.1.2 Urban vegetation and water areas

Vegetation and water have distinct spectral characteristics, hence some classic spectral indices, such as the NDVI, the normalized difference water index (NDWI), and the modified NDWI (MNDWI), were designed for their detection based on simple band operations. However, most of the existing indices are not very appropriate to be directly applied to HR images in urban environment due to the spectral similarity of some urban structures (e.g., temporary building, shadow) which also show high response of these indices. Moreover, many HR imagery have only four spectral bands (i.e., blue, green, red, near-infrared), hence lack the prerequisite channel to calculate some indices (e.g., MNDWI). To tackle these problems, Kumar et al. (2012) created a new vegetation index by taking advantages of the NIR-2 and red edge bands of WorldView-2 imagery to extract vegetation, which obtained improved accuracy compared to the conventional NDVI. Xie et al. (2016) proposed a HR water index based on combinations of WorldView-2 eight-band data. This new water index was further combined with the MSI, in order to alleviate the disturbance from shadows. According to their tests, the water index calculated by coastal-NIR2 or green-NIR2 bands achieved the best performance to highlight urban water areas in multiple big cities.

In addition to the binary extraction of vegetation and water, some studies further concerned the identification of their subclasses. For instance, Wen et al. (2017) conducted semantic classification of the urban tree functions from HR

imagery. Unlike general vegetation extraction and tree species classification, this study considered the location and function of trees, by dividing trees into park trees, roadside trees and residential-institutional trees based on a multi-level framework (pixel-object-patch). Specifically, the vegetation index was firstly calculated for vegetation extraction at the pixel level. Object-oriented segmentation was then performed over the vegetated area, and the spectral and texture features were extracted at the object level to distinguish between trees and ground vegetation. Finally, the semantic functions of the trees were obtained based on the area, shape, structure, and spatial relationships at the patch level. Huang et al. (2015b) identified different water-body types including lakes, rivers, canals and ponds in Wuhan and Shenzhen. A two-layer machine learning framework was presented, which first detected preliminary water areas by using water, shadow, and vegetation indices. The geometrical and textural features were then extracted at the object level, and the different urban water types were finally identified.

In general, courtesy of the rich spatial details in HR imagery, massive efforts have been devoted to propose automatic or semi-automatic approaches for urban object extraction, by characterizing the physical properties (i.e., texture, shape, height) of the target of interest. Although impressive results were obtained, their applicability may be limited since the spatial contexts of complex and varying urban scenes are often difficult to be described as a set of "rules". Current studies that integrate domain knowledge with machine learning techniques (e.g., deep learning) show a promising direction for urban target detection from HR imagery (Zhou et al., 2016; Zhu et al., 2017).

3.2 Land use/land cover mapping

Urban land use/land cover (LULC) information are crucial data to understand the complex interactions between human and the environment (Kuang et al., 2016; Yu et al., 2016). Land cover focuses on the physical property of the land surface, e.g., impervious surfaces, vegetation and water, while land use places emphasis on the social functional attributes, such as residential, industrial or commercial. Numerous methods have been proposed for LULC mapping, which can generally be categorized into three types according to their basic processing units (i.e. pixels, objects, and moving windows). Pixel- and object-based approaches are widely used for land cover mapping (Myint et al., 2011). While classifying land use is more difficult since it relates to human activities and one land use type is often mixed by multiple land covers. The texture, geometry, contexture, land cover proportion, or other auxiliary data (e.g., Google Street View) are often incorporated to recognize land use patterns and configurations (Li X et al., 2017; Zhang et al., 2019). Machine learning techniques for supervised classification have been extensively exploited for urban LULC mapping from HR imagery, such as the classification and regression tree (CART), *k*-nearest neighbor (KNN), random forest (RF), SVM, and multilayer perceptron (MLP) (Qian et al., 2015; Zhang and Huang, 2018). Since traditional low-level features may be insufficient to characterize urban land use, mid-level features are constructed by means of dictionary learning and sparse coding such as the popular bag-of-visual-words (BOVW) and latent Dirichlet allocation (LDA) models. Based on these scene models, Huang et al. (2015a) extracted the urban villages from QuickBird and WorldView-2 images covering Wuhan and Shenzhen; Zhang and Du (2015) employed city blocks as the processing unit to map the urban functional zones in Zhuhai and Beijing. Moreover, a variety of landscape metrics were calculated in order to better describe complex urban scenes. For example, Voltersen et al. (2014) combined HR images and a normalized DSM (nDSM) to classify basic urban features. Several landscape metrics of buildings and vegetation, e.g., volume, height, and vegetation fraction, were then extracted at the block level to describe the urban structure types, dividing the urban land use into residential, commercial, and industrial areas, parks, woodland, etc. Liu H et al. (2017) correlated the physical structure properties of urban villages (e.g., high building density and scarce vegetation) and the landscape metrics for urban village mapping. Recently, more endeavors have been devoted to employ the new deep learning methods for LULC mapping, which achieved state-of-the-art performances by learning the most discriminative features hierarchically (Huang B et al., 2018; Zhang et al., 2019; Zhu et al., 2017).

3.3 Change detection

In developing regions, many cities are undergoing rapid urban expansion as well as internal formation and demolition. Timely and efficient monitoring of urban changes helps us to understand human activity and provide a decision-making basis for urban planning (Marin et al., 2015). HR remote sensing imagery enables detection of subtle urban changes, but also poses great difficulties to the traditional methods. The major challenge of change detection from HR imagery is the confusion of radiometric and real semantic changes. False alarms are mainly induced by the distinct spectral heterogeneity of the multi-temporal HR images (i.e., the spectral behavior of the same object varies at different dates), due to different imaging conditions, mis-registration, disparity of vertical structures, etc. (Bruzzone and Bovolo, 2012). The widely used change detection methods include machine learning approaches (Volpi et al., 2013) and automatic methods such as the multilevel change vector analysis (Bovolo, 2009), the pulse-coupled neural networks (Pacifici and Del Frate, 2010), and the multi-temporal morphological attribute profiles (Falco et al., 2013). In order to tackle the

problems of spectral complexity and mis-registration of the multi-temporal HR images, Wen et al. (2016) employed several basic urban primitives (building, vegetation, and water) to represent complex urban scenes and utilized blocks as the basic unit to calculate their composition and arrangement. The information of the primitive features in the corresponding blocks between multi-temporal images were compared, and then the changed area and type can be identified. In addition, Huang et al. (2017a) proposed a multi-level (pixel, grid, and city block) framework for urban change analysis from HR images. The ZY-3 stereo images were employed to produce the multi-temporal orthographic images covering Wuhan and Beijing, and the multi-features and rules were integrated for land cover classification. The authors also compared the results of ZY-3 and Landsat images, suggesting that HR imagery was indispensable for subtle urban change detection.

3.4 Ecology and climate analysis

Urban LULC types have profound impacts on the urban ecological environment and climate (Kuang et al., 2017; Shi et al., 2016). On the basis of LULC data, more in-depth information of a city can be explored, such as the quantitative assessment of ecosystem service capacities or simulation of local climate. Burkhard et al. (2012) proposed a clear and applicable conceptual framework for ecosystem service mapping. The authors linked the different land cover types to ecosystem service supply and demand, and synthesized the expert knowledge from many studies to give the capacity scores of different land cover types. Following the ecosystem service concept and assessment scheme, Haas and Ban (2017) assessed the ecological changes in the urban core districts of Shanghai in China. The authors first generated the basic urban LULC types from multi-temporal IKONOS and GeoEye-1 images. The LULC changes were then analyzed in terms of ecosystem service supply and demand, and the ecosystem balance of central Shanghai from 2000 to 2009 was modeled.

On the other hand, the distribution and variation of the urban climate have received broad attention, as the urban climate is closely related to environmental and human health issues. The dense population and heterogeneous landscapes in urban areas make the intra-urban climate distinctly different. Stewart and Oke (2012) introduced a universal classification scheme for the land surface, called local climate zones (LCZ), for urban climate studies. According to the building properties (e.g., height, density, material), the LCZ system classifies 10 built types, such as compact high-rise, open low-rise, and heavy industry. Meanwhile, seven land cover categories, including dense trees, low plants, water, etc., are also defined to represent natural landscapes. The spatial scale of LCZ studies is generally between 100 and 1000 m. Bechtel et al. (2015) investigated the feasibility of using remote sensing imagery for LCZ mapping, and developed the World Urban Database and Access Portal Tools (WUDAPT) for global HR LCZ mapping. Some follow-up studies were carried out, for instance, Bechtel et al. (2016) integrated multispectral data and SAR imagery for LCZ mapping in arid areas. Wang et al. (2018) performed LCZ classification in arid desert cities using open-source image data and software. The LCZ properties were evaluated and compared to the original value ranges in Stewart and Oke (2012), and their relationships with the surface urban heat island effect were also analyzed.

4. Conclusions and perspectives

As an emerging research field, the development of high-resolution (HR) urban remote sensing is inseparable from sensor technology, photogrammetry, image processing technique, etc. High-resolution, multi-temporal, multi-angle, and multi-platform urban observations allow more elaborate urban application, which also call for more effective data interpretation approaches. In this paper, we examined the major advances in HR urban remote sensing from the feature level to the scene level. A series of advanced textural, structural and stereo features, as well as the new techniques for urban information extraction were summarized. Although HR urban remote sensing has achieved substantial improvements, many of the existing researches only involve small-scale applications and their robustness and generalization ability are unproved. There are still great challenges to meet the requirements of practical production or commercialization. The trends for future development are discussed from the following perspectives.

(1) Progress in methodology. Recent developments in machine learning, especially deep learning, have achieved notable success in plenty of scientific fields. For instance, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are the two important branches for spatial learning and sequence learning, respectively, which are highly effective in urban remote sensing tasks. Nevertheless, applying deep learning to remotely sensed image interpretation is still in its infancy mainly due to the strong data heterogeneity, inadequate sample annotation, and high complexity of the model. Specialized training datasets for HR remote sensing imagery are still few in number. There is an urgent need for creating large-volume benchmark data with abundant, diverse, and reliable representation of various urban landscapes. Unsupervised, semi-supervised, or weakly supervised learning approaches are also promising to reduce the work of manual annotation.

(2) Characterizing urban structure and pattern. Identifying the internal pattern, configuration or functional attributes (e.

g., commercial/residential/industrial areas, urban villages, ecological/leisure land) of urban land is significant for effective urban management and planning. By virtue of HR stereo imagery (e.g., ZY-3, WorldView), 3D urban form can be conveniently derived, which can support the future researches on urban dynamics in both horizontal and vertical dimensions, as well as providing new avenues for multidisciplinary applications such as socioeconomic research, disaster response, environmental assessment, and ecological modelling.

(3) Big data challenge. The astonishing geospatial data acquisition ability and the trend of elaborating urban information extraction over time and space calls for increased computational power to address the big data challenges (Ma et al., 2015). The abundant high-performance computing resources and cloud technology offer promising solutions (Li et al., 2016; Sun et al., 2019). One such example is GEE (Gorelick et al., 2017), which has archived petabytes of EO data and allows interactive data process and algorithm development over its online system, facilitating lots of researchers to carry out global or continental urban mapping and studies (Gong et al., 2019a, 2019b; Liu et al., 2018).

(4) Global mapping. The exponentially growing HR data facilitates urban remote sensing towards larger scales hence increase our knowledge about the fast urbanizing world. Historically, urban mapping at the global scale are mostly relied on coarse resolution imagery (e.g., MODIS). With the availability of the global Landsat archives, some fine resolution global products were generated in recent years, such as the FROM-GLC30 (Gong et al., 2013), GlobeLand30 (Chen et al., 2015), GHSL (Pesaresi et al., 2015), and MGIS (Liu et al., 2018), which provide us valuable information about the location and extent of urban areas worldwide. More recently, the global land cover map at 10 m resolution, i.e., FROM-GLC10 (Gong et al., 2019), was derived from Sentinel-2 data at so far the highest spatial resolution. In addition, new satellites, such as GaoFen series, ZY-3 constellation, PlanetScope, and Luojia-1, can acquire optical, SAR and nighttime light data at high spatial/temporal resolution, which have great potentials for global urban information extraction in the future.

Acknowledgements *This work was supported by the National Natural Science Foundation of China (Grant Nos. 41771360 & 41842035), the National Program for Support of Top-notch Young Professionals, the Hubei Provincial Natural Science Foundation of China (Grant No. 2017CFA029), the National Key Research and Development Program of China (Grant No. 2016YFB0501403), and the Shenzhen Science and Technology Program (Grant No. JCYJ20180306170645080).*

References

Bechtel B, Alexander P, Böhner J, Ching J, Conrad O, Feddema J, Mills G, See L, Stewart I. 2015. Mapping local climate zones for a worldwide database of the form and function of cities. ISPRS Int Geo-Inf, 4: 199–219

Bechtel B, See L, Mills G, Foley M. 2016. Classification of local climate zones using SAR and multispectral data in an arid environment. IEEE J Sel Top Appl Earth Observ Remote Sens, 9: 3097–3105

Benediktsson J A, Palmason J A, Sveinsson J R. 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans Geosci Remote Sens, 43: 480–491

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: 1940–1949

Bovolo F. 2009. A multilevel parcel-based approach to change detection in very high resolution multitemporal images. IEEE Geosci Remote Sens Lett, 6: 33–37

Bruzzone L, Bovolo F. 2012. A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. Proc IEEE, 101: 609–630

Burkhard B, Kroll F, Nedkov S, Müller F. 2012. Mapping ecosystem service supply, demand and budgets. Ecol Indic, 21: 17–29

Chanussot J, Benediktsson J A, Fauvel M. 2006. Classification of remote sensing images from urban areas using a fuzzy possibilistic model. IEEE Geosci Remote Sens Lett, 3: 40–44

Chen J, Chen J, Liao A P, Cao X, Chen L J, Chen X H, He C Y, Han G, Peng S, Lu M, Zhang W W, Tong X H, Mills J. 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J Photogramm Remote Sens, 103: 7–27

Chen X H, Cao X, Liao A P, Chen L J, Peng S, Lu M, Chen J, Zhang W W, Zhang H W, Han G, Wu H, Li R. 2016. Global mapping of artificial surfaces at 30-m resolution. Sci China Earth Sci, 59: 2295–2306

Chini M, Pierdicca N, Emery W J. 2009. Exploiting SAR and VHR optical images to quantify damage caused by the 2003 Bam earthquake. IEEE Trans Geosci Remote Sens, 47: 145–152

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

Dalla Mura M, Benediktsson J A, Waske B, Bruzzone L. 2010b. Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans Geosci Remote Sens, 48: 3747–3762

Falco N, Mura M D, Bovolo F, Benediktsson J A, Bruzzone L. 2013. Change detection in VHR images based on morphological attribute profiles. IEEE Geosci Remote Sens Lett, 10: 636–640

Fauvel M, Benediktsson J A, Chanussot J, Sveinsson J R. 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans Geosci Remote Sens, 46: 3804–3814

Florczyk A J, Ferri S, Syrris V, Kemper T, Halkia M, Soille P, Pesaresi M. 2016. A new European settlement map from optical remotely sensed data. IEEE J Sel Top Appl Earth Observ Remote Sens, 9: 1978–1992

Gamba P, Herold M. 2009. Global Mapping of Human Settlement: Experiences, Datasets, and Prospects. Boca Raton (FL): CRC Press. 374

Ghamisi P, Dalla Mura M, Benediktsson J A. 2015. A survey on spectral–spatial classification techniques based on attribute profiles. IEEE Trans Geosci Remote Sens, 53: 2335–2353

Gong P, Li X, Zhang W. 2019a. 40-year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing. Chin Sci Bull, 64: 756–763

Gong P, Liang S, Carlton E J, Jiang Q, Wu J, Wang L, Remais J V. 2012. Urbanisation and health in China. Lancet, 379: 843–852

Gong P, Liu H, Zhang M, Li C, Wang J, Huang H, Clinton N, Ji L, Li W, Bai Y, Chen B, Xu B, Zhu Z, Yuan C, Ping Suen H, Guo J, Xu N, Li W, Zhao Y, Yang J, Yu C, Wang X, Fu H, Yu L, Dronova I, Hui F, Cheng X, Shi X, Xiao F, Liu Q, Song L. 2019b. Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. Chin Sci Bull, 64: 370–373

Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, Niu Z, Huang X, Fu H,

Liu S, Li C, Li X, Fu W, Liu C, Xu Y, Wang X, Cheng Q, Hu L, Yao W, Zhang H, Zhu P, Zhao Z, Zhang H, Zheng Y, Ji L, Zhang Y, Chen H, Yan A, Guo J, Yu L, Wang L, Liu X, Shi T, Zhu M, Chen Y, Yang G, Tang P, Xu B, Giri C, Clinton N, Zhu Z, Chen J, Chen J. 2013. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. Int J Remote Sens, 34: 2607–2654

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ, 202: 18–27

Guo X, Huang X, Zhang L. 2014. Three-dimensional wavelet texture feature extraction and classification for multi/hyperspectral imagery. IEEE Geosci Remote Sens Lett, 11: 2183–2187

Haas J, Ban Y. 2017. Mapping and monitoring urban ecosystem services using multitemporal high-resolution satellite data. IEEE J Sel Top Appl Earth Observ Remote Sens, 10: 669–680

Haralick R M, Shanmugam K, Dinstein I H. 1973. Textural features for image classification. IEEE Trans Syst Man Cybern, SMC-3: 610–621

He C, Liu Z, Gou S, Zhang Q, Zhang J, Xu L. 2019. Detecting global urban expansion over the last three decades using a fully convolutional network. Environ Res Lett, 14: 034008

Hu F, Xia G S, Hu J, Zhang L. 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens, 7: 14680–14707

Hu X, Weng Q. 2011. Impervious surface area extraction from IKONOS imagery using an object-based fuzzy method. Geocarto Int, 26: 3–20

Huang B, Zhao B, Song Y. 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sens Environ, 214: 73–86

Huang J K, Zhu L F, Deng X Z. 2007. Regional differences and determinants of built-up area expansion in China. Sci China Ser D-Earth Sci, 50: 1835–1843

Huang X, Chen H, Gong J. 2018. Angular difference feature extraction for urban scene classification using ZY-3 multi-angle high-resolution satellite imagery. ISPRS J Photogramm Remote Sens, 135: 127–141

Huang X, Guan X, Benediktsson J A, Zhang L, Li J, Plaza A, Dalla Mura M. 2014a. Multiple morphological profiles from multicomponent-base images for hyperspectral image classification. IEEE J Sel Top Appl Earth Observ Remote Sens, 7: 4653–4669

Huang X, Han X, Zhang L, Gong J, Liao W, Benediktsson J A. 2016. Generalized differential morphological profiles for remote sensing image classification. IEEE J Sel Top Appl Earth Observ Remote Sens, 9: 1736–1751

Huang X, Liu H, Zhang L. 2015a. 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, Liu X, Zhang L. 2014b. A multichannel gray level co-occurrence matrix for multi/hyperspectral image texture representation. Remote Sens, 6: 8424–8445

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, Xie C, Fang X, Zhang L. 2015b. Combining pixel- and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery. IEEE J Sel Top Appl Earth Observ Remote Sens, 8: 2097–2110

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 Observ Remote Sens, 10: 654–668

Huang X, Zhang L. 2009. Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines. Int J Remote Sens, 30: 1977–1987

Huang X, Zhang L. 2011. A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery. Photogramm Eng Remote Sens, 77: 721–732

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

Huang X, Zhang L. 2012b. A multiscale urban complexity index based on 3D wavelet transform for spectral-spatial feature extraction and classification: An evaluation on the 8-channel WorldView-2 imagery. Int J Remote Sens, 33: 2641–2656

Huang X, Zhang L, Li P. 2007a. An adaptive multiscale information fusion approach for feature extraction and classification of IKONOS multispectral imagery over urban areas. IEEE Geosci Remote Sens Lett, 4: 654–658

Huang X, Zhang L, Li P. 2007b. Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. IEEE Geosci Remote Sens Lett, 4: 260–264

Kuang W H, Chen L J, Liu J Y, Xiang W N, Chi W F, Lu D S, Yang T R, Pan T, Liu A L. 2016. Remote sensing-based artificial surface cover classification in Asia and spatial pattern analysis. Sci China Earth Sci, 59: 1720–1737

Kuang W H, Yang T R, Liu A L, Zhang C, Lu D S, Chi W F. 2017. An EcoCity model for regulating urban land cover structure and thermal environment: Taking Beijing as an example. Sci China Earth Sci, 60: 1098–1109

Kumar A, Pandey A C, Jeyaseelan A T. 2012. Built-up and vegetation extraction and density mapping using WorldView-II. Geocarto Int, 27: 557–568

Li J, Huang X, Gong J. 2019. Deep neural network for remote-sensing image interpretation: Status and perspectives. Natl Sci Rev, doi: 10.1093/nsr/nwz058

Li Q, Huang X, Wen D, Liu H. 2017. Integrating multiple textural features for remote sensing image change detection. Photogramm Eng Remote Sens, 83: 109–121

Li S, Dragicevic S, Castro F A, Sester M, Winter S, Coltekin A, Pettit C, Jiang B, Haworth J, Stein A, Cheng T. 2016. Geospatial big data handling theory and methods: A review and research challenges. ISPRS J Photogramm Remote Sens, 115: 119–133

Li W, Chen C, Su H, Du Q. 2015. Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans Geosci Remote Sens, 53: 3681–3693

Li X, Gong P, Liang L. 2015. A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data. Remote Sens Environ, 166: 78–90

Li X, Zhang C, Li W. 2017. Building block level urban land-use information retrieval based on Google Street View images. GISci Remote Sens, 54: 819–835

Li Y S, Huang X, Liu H. 2017. Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. Photogramm Eng Remote Sens, 83: 567–579

Liu C, Huang X, Wen D, Chen H, Gong J. 2017. Assessing the quality of building height extraction from ZiYuan-3 multi-view imagery. Remote Sens Lett, 8: 907–916

Liu C, Huang X, Zhu Z, Chen H, Tang X, Gong J. 2019. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: Analysis of 45 global cities. Remote Sens Environ, 226: 51–73

Liu H, Huang X, Wen D, Li J. 2017. The use of landscape metrics and transfer learning to explore urban villages in China. Remote Sens, 9: 365

Liu X, Hu G, Chen Y, Li X, Xu X, Li S, Pei F, Wang S. 2018. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. Remote Sens Environ, 209: 227–239

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

Ma Y, Wu H, Wang L, Huang B, Ranjan R, Zomaya A, Jie W. 2015. Remote sensing Big Data computing: Challenges and opportunities. Futur Gener Comp Syst, 51: 47–60

Mallat S G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans Pattern Anal Mach Intell, 11: 674–

693
Marin C, Bovolo F, Bruzzone L. 2015. Building change detection in multitemporal very high resolution SAR images. IEEE Trans Geosci Remote Sens, 53: 2664–2682
Marmanis D, Datcu M, Esch T, Stilla U. 2016. Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci Remote Sens Lett, 13: 105–109
Musci M, Feitosa R Q, Costa G A O P, Velloso M L F. 2013. Assessment of binary coding techniques for texture characterization in remote sensing imagery. IEEE Geosci Remote Sens Lett, 10: 1607–1611
Myint S W, Gober P, Brazel A, Grossman-Clarke S, Weng Q. 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens Environ, 115: 1145–1161
Myint S W, Lam N S N, Tyler J M. 2004. Wavelets for urban spatial feature discrimination. Photogramm Eng Remote Sens, 70: 803–812
Nogueira K, Penatti O A B, dos Santos J A. 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognit, 61: 539–556
Ojala T, Pietikäinen M, Harwood D. 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit, 29: 51–59
Ojala T, Pietikainen M, Maenpaa T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell, 24: 971–987
Ok A O. 2013. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. ISPRS J Photogramm Remote Sens, 86: 21–40
Ok A O, Senaras C, Yuksel B. 2013. Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. IEEE Trans Geosci Remote Sens, 51: 1701–1717
Ouma Y O, Ngigi T G, Tateishi R. 2006. On the optimization and selection of wavelet texture for feature extraction from high-resolution satellite imagery with application towards urban-tree delineation. Int J Remote Sens, 27: 73–104
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
Pacifici F, Del Frate F. 2010. Automatic change detection in very high resolution images with pulse-coupled neural networks. IEEE Geosci Remote Sens Lett, 7: 58–62
Peng F, Gong J, Wang L, Wu H, Liu P. 2017. A new stereo pair disparity index (SPDI) for detecting built-up areas from high-resolution stereo imagery. Remote Sens, 9: 633
Peng F, Wang L, Gong J, Wu H. 2015. Development of a framework for stereo image retrieval with both height and planar features. IEEE J Sel Top Appl Earth Observ Remote Sens, 8: 800–815
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
Pesaresi M, Ehrilch D, Florczyk A, Freire S, Julea A, Kemper T, Soille P, Syrris V. 2015. GHS Built-up Grid, Derived from Landsat, Multitemporal (1975, 1990, 2000, 2014). European Commission, Joint Research Centre (JRC)
Pesaresi M, Ehrlich D, Caravaggi I, Kauffmann M, Louvrier C. 2011. Toward global automatic built-up area recognition using optical VHR imagery. IEEE J Sel Top Appl Earth Observ Remote Sens, 4: 923–934
Pesaresi M, Gerhardinger A. 2011. Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation. IEEE J Sel Top Appl Earth Observ Remote Sens, 4: 16–26
Pesaresi M, Gerhardinger A, Kayitakire F Ç. 2008. A robust built-up area presence index by anisotropic rotation-invariant textural measure. IEEE J Sel Top Appl Earth Observ Remote Sens, 1: 180–192
Pesaresi M, Huadong G, Blaes X, Ehrlich D, Ferri S, Gueguen L, Halkia M, Kauffmann M, Kemper T, Lu L, Marin-Herrera M A, Ouzounis G K, Scavazzon M, Soille P, Syrris V, Zanchetta L. 2013. A global human settlement layer from optical HR/VHR RS data: Concept and first results. IEEE J Sel Top Appl Earth Observ Remote Sens, 6: 2102–2131
Poullis C. 2014. Tensor-Cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images. ISPRS J Photogramm Remote Sens, 95: 93–108
Puissant A, Hirsch J, Weber C. 2005. The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. Int J Remote Sens, 26: 733–745
Qian Y, Ye M, Zhou J. 2013. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. IEEE Trans Geosci Remote Sens, 51: 2276–2291
Qian Y, Zhou W, Yan J, Li W, Han L. 2015. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. Remote Sens, 7: 153–168
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 Sens, 80: 873–883
Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat N. 2019. Deep learning and process understanding for data-driven Earth system science. Nature, 566: 195–204
Schneider A, Friedl M A, Potere D. 2010. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. Remote Sens Environ, 114: 1733–1746
Sghaier M O, Lepage R. 2016. Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. IEEE J Sel Top Appl Earth Observ Remote Sens, 9: 1946–1958
Shanmugam L, Kaliaperumal V. 2015. Water flow based geometric active deformable model for road network. ISPRS J Photogramm Remote Sens, 102: 140–147
Shao Z, Fu H, Fu P, Yin L. 2016. Mapping urban impervious surface by fusing optical and SAR data at the decision level. Remote Sens, 8: 945–966
Shi X L, Nie S P, Ju W M, Yu L. 2016. Climate effects of the GlobeLand30 land cover dataset on the Beijing Climate Center climate model simulations. Sci China Earth Sci, 59: 1754–1764
Song C, Yang F, Li P. 2010. Rotation invariant texture measured by local binary pattern for remote sensing image classification. Wuhan: 2010 Second International Workshop on Education Technology and Computer Science. 3: 3–6
Stewart I D, Oke T R. 2012. Local climate zones for urban temperature studies. Bull Amer Meteorol Soc, 93: 1879–1900
Su W, Li J, Chen Y, Liu Z, Zhang J, Low T M, Suppiah I, Hashim S A M. 2008. Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. Int J Remote Sens, 29: 3105–3117
Sun J, Zhang Y, Wu Z, Zhu Y, Yin X, Ding Z, Wei Z, Plaza J, Plaza A. 2019. An efficient and scalable framework for processing remotely sensed big data in cloud computing environments. IEEE Trans Geosci Remote Sens, 57: 4294–4308
Tian J, Chen D M. 2007. Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. Int J Remote Sens, 28: 4625–4644
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
Tuia D, Pacifici F, Kanevski M, Emery W J. 2009. Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. IEEE Trans Geosci Remote Sens, 47: 3866–3879
United Nations. 2018. 2018 Revision of World Urbanization Prospects. Population Division, Department of Economic and Social Affairs: United Nations Publications
Volpi M, Tuia D, Bovolo F, Kanevski M, Bruzzone L. 2013. Supervised

change detection in VHR images using contextual information and support vector machines. Int J Appl Earth Observ Geoinf, 20: 77–85
Voltersen M, Berger C, Hese S, Schmullius C. 2014. Object-based land cover mapping and comprehensive feature calculation for an automated derivation of urban structure types at block level. Remote Sens Environ, 154: 192–201
Wang C, Middel A, Myint S W, Kaplan S, Brazel A J, Lukasczyk J. 2018. Assessing local climate zones in arid cities: The case of Phoenix, Arizona and Las Vegas, Nevada. ISPRS J Photogramm Remote Sens, 141: 59–71
Wen D, Huang X, Liu H, Liao W, Zhang L. 2017. Semantic classification of urban trees using very high resolution satellite imagery. IEEE J Sel Top Appl Earth Observ Remote Sens, 10: 1413–1424
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
Weng Q. 2012. Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sens Environ, 117: 34–49
Xie C, Huang X, Zeng W, Fang X. 2016. A novel water index for urban high-resolution eight-band WorldView-2 imagery. Int J Digital Earth, 9: 925–941
Yoo H Y, Lee K, Kwon B D. 2009. Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery. Int J Remote Sens, 30: 6219–6239
Yu X, Zhang B Q, Li Q, Chen J. 2016. A method characterizing urban expansion based on land cover map at 30 m resolution. Sci China Earth Sci, 59: 1738–1744
Zhang C, Sargent I, Pan X, Li H, Gardiner A, Hare J, Atkinson P M. 2019. Joint Deep Learning for land cover and land use classification. Remote Sens Environ, 221: 173–187
Zhang L, Huang X, Huang B, Li P. 2006. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. IEEE Trans Geosci Remote Sens, 44: 2950–2961
Zhang L, Zhang L, Tao D, Huang X. 2013. A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images. ISPRS J Photogramm Remote Sens, 83: 30–39
Zhang T, Huang X. 2018. Monitoring of urban impervious surfaces using time series of high-resolution remote sensing images in rapidly urbanized areas: A case study of Shenzhen. IEEE J Sel Top Appl Earth Observ Remote Sens, 11: 2692–2708
Zhang T, Huang X, Wen D, Li J. 2017. Urban building density estimation from high-resolution imagery using multiple features and support vector regression. IEEE J Sel Top Appl Earth Observ Remote Sens, 10: 3265–3280
Zhang X, Du S. 2015. A Linear Dirichlet Mixture Model for decomposing scenes: Application to analyzing urban functional zonings. Remote Sens Environ, 169: 37–49
Zhang Y, Zhang H, Lin H. 2014. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. Remote Sens Environ, 141: 155–167
Zhou P, Cheng G, Liu Z, Bu S, Hu X. 2016. Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. Multidim Syst Sign Process, 27: 925–944
Zhu X X, Tuia D, Mou L, Xia G S, Zhang L, Xu F, Fraundorfer F. 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci Remote Sens Mag, 5: 8–36

(Responsible editor: Xin LI)