

Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data



Xiuyuan Zhang^a, Shihong Du^{a,*}, Qiao Wang^b

^a Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China

^b Satellite Environment Center, Ministry of Environmental Protection, Beijing 100094, China

ARTICLE INFO

Article history:

Received 8 March 2017

Received in revised form 8 August 2017

Accepted 12 September 2017

Available online 17 September 2017

Keywords:

Urban functional zone

Image classification

Hierarchical semantic cognition

Point of interest

ABSTRACT

As the basic units of urban areas, functional zones are essential for city planning and management, but functional-zone maps are hardly available in most cities, as traditional urban investigations focus mainly on land-cover objects instead of functional zones. As a result, an automatic/semi-automatic method for mapping urban functional zones is highly required. Hierarchical semantic cognition (HSC) is presented in this study, and serves as a general cognition structure for recognizing urban functional zones. Unlike traditional classification methods, the HSC relies on geographic cognition and considers four semantic layers, i.e., visual features, object categories, spatial object patterns, and zone functions, as well as their hierarchical relations. Here, we used HSC to classify functional zones in Beijing with a very-high-resolution (VHR) satellite image and point-of-interest (POI) data. Experimental results indicate that this method can produce more accurate results than Support Vector Machine (SVM) and Latent Dirichlet Allocation (LDA) with a larger overall accuracy of 90.8%. Additionally, the contributions of diverse semantic layers are quantified: the object-category layer is the most important and makes 54% contribution to functional-zone classification; while, other semantic layers are less important but their contributions cannot be ignored. Consequently, the presented HSC is effective in classifying urban functional zones, and can further support urban planning and management.

© 2017 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

The “functional zone” is a cultural concept that describes human activities in a certain area (Hu et al., 2016). Each functional zone is spatially aggregated by diverse geographic objects, and semantically abstracted from land uses (Zhang and Du, 2015). Functional zones are often employed as the basic units for urban planning (Matsuoka and Kaplan, 2008; Montanges et al., 2015), and greatly impact urban transportation, resource management, and factory relocation (Heiden et al., 2012; Shin, 2009; Zhao and Lü, 2009).

However, functional-zone maps are hardly available, as most urban investigations focus on geographic objects instead of large-scale functional zones. Functional zones are not only spatially larger than objects but also semantically different from objects. For example, a building belongs to land-cover objects, while a residential district is related to a functional zone. Accordingly, the two

kinds of units are located at different semantic layers, and thus functional zones cannot be classified by traditional object-based methods. In addition, the widely used commercial-map servers, e.g., Google Map and Baidu Map, can solely provide points of interest rather than functional zones; thus, functional zones are unavailable through commercial map servers. Therefore, we need to develop new techniques to analyze urban functional zones and generate their maps.

Recent years have witnessed the development of high-quality remote sensing images, especially for VHR satellite images which hold certain advantages in representing functional zones, because of the large spatial coverage, detailed information, and wide availability (Lienou et al., 2010). In order to map functional zones with VHR images, three challenging issues have to be resolved, i.e., zone segmentation, feature representation, and functional classification (Zhang et al., 2015). For zone segmentation, it cannot be achieved by classical object segmentation (Baatz and Schape, 2000; Drăguț et al., 2014), even by using large scale parameters, as functional zones have substantial discontinuities in the visual cues of spectrums, edges, and textures (Pertuz et al., 2015). Instead, recent

* Corresponding author at: Peking University, Beijing 100871, China.

E-mail address: dshgis@hotmail.com (S. Du).

studies tended to use blocks to spatially delineate functional zones (Heiden et al., 2012; Hu et al., 2016; Montanges et al., 2015; Zhang and Du, 2015), because blocks always serve as the basic units of urban sprawl and renewal, and they have clear functional meanings (Neis and Zipf, 2012; Zheng et al., 2016), e.g., residential, commercial, and industrial districts (Fig. 1). Accordingly, this segmentation method will be applied to this study, and each block is regarded as a functional zone.

For feature representations, visual ones including spectral, geometrical and textural features were widely used at the earlier phase of zone classification (Zhang et al., 2015). Steeves et al. (2004) first used colors and textures to characterize functional zones and further distinguish their categories. They believed that per-pixel visual features are effective enough for representing zones, and zone classification can operate independently of object perception. This opinion is supported by some other studies. For example, Rhinane et al. (2011) directly used the visual features defined in ENVI software to classify shanty towns in urban areas, and Wen et al. (2016) extracted spectral features to detect the temporal changes in urban functional zones. On the other hand, recent work has noticed that per-pixel, or even local-image-patch, features have few semantic meanings (Li et al., 2010; Serrano et al., 2004); while objects as the basic elements of zones play an important role in measuring zone features. Accordingly, object categories and spatial object patterns are widely used to characterize functional zones (Benz et al., 2004; Burnett and Blaschke, 2003; Yao et al., 2012).

Although there are so many features ranging from per-pixel to object-based ones, they are still unable to characterize functional zones comprehensively, as their hierarchical relations are totally ignored, which does harm to their representative abilities (Burnett and Blaschke, 2003; Farabet et al., 2013). For example, per-pixel visual features are basic cues for recognizing object categories, and then object categories are fundamental to distinguishing zone functions. Accordingly, the hierarchical relations from visual features to zone functions should be considered in zone representations.

For zone classification, it was usually regarded as a computer-vision issue, i.e., scene classification, in which functional zones are represented by image scenes and are finally processed by classifiers (Boutell et al., 2004; Farahzadeh et al., 2015). Preliminary scene classifications (Fig. 2) came from 2002, and concentrated on recognizing living scenes, such as indoor/outdoor images, by using support vector machine (SVM) and artificial neural network (ANN) (Payne and Singh, 2005; Serrano et al., 2002; Yan et al., 2003). These classifiers directly use visual features to distinguish diverse image scenes, and generate outstanding recognition results, which contributes to image searching on the internet (Farabet et al., 2013). However, these classifiers are weak in handling heterogeneous image scenes (Kontschieder et al., 2014; Lienou et al., 2010; Oquab et al., 2014).

To resolve this issue, topic models were introduced and they used potential semantics as cues to recognize scene categories (Cao and Fei-Fei, 2007). Experimental results indicated that topic models, e.g., probabilistic latent semantic analysis (pLSA) and latent Dirichlet allocation (LDA), produced more accurate classification results than traditional classifiers for most image scenes (Bosch et al., 2006; Zhong et al., 2015). However, the uncertain parameters in these topic models, e.g., the number of potential semantics, can significantly impact classification results (Blei et al., 2003; Wang and Grimson, 2008). Accordingly, Li et al. (2010) used object semantics to replace potential ones, and presented a novel scene classifier, object bank (Fig. 2c). This method does not only generate more accurate classification results, but also has stronger robustness and interpretability. Nevertheless, functional-zone classification with VHR satellite images is rather than a computer-vision issue, and is much more complicate than traditional image-scene classifications, owing to its ambiguity and variability (Song et al., 2015; Zhong et al., 2015).

From the view of geographic cognition, the two issues in mapping functional zones, i.e., feature representation and zone classification, are still unresolved (Cohen, 2000; Wang et al., 2010). Taking the three zones in Fig. 3 as examples, all of them are composed of buildings, vegetation, and impervious layer, and their distinguishability at different semantic layers, i.e., visual features, object categories, spatial object patterns, and zone functions, are discussed respectively. At visual-feature layer, spectral features are extracted to characterize these zones. As the three zones consist of similar objects, their visual features show no difference, and thus cannot distinguish their categories. At the second layer, object categories and their proportions are used to represent these zones, which have low distinguishability for the three zones. Spatial object patterns at the third layer characterize the structures of the three zones. As reported in Fig. 3, the three zones have different spatial object patterns, which can be respectively defined as “evenly distribution”, “random distribution”, and “aggregation”. Accordingly, these zones can be easily distinguished by spatial object patterns. Finally, at the forth layers, the three zones can be described as “a villa district”, “a residential zone”, and “a shantytown” respectively. Accordingly, spatial object patterns are essential to distinguish zone functions. Nevertheless, visual features and object categories cannot be ignored, because object categories are fundamental to measuring spatial object patterns, while visual features are basic cues to recognize object categories (Li et al., 2010). Consequently, the four semantic layers should be fully considered in functional-zone classification.

In summary, the existing studies on functional-zone mapping were weak in feature representation and zone classification, as they totally ignore the hierarchical structure from visual features to zone functions, and are inconsistent with the geographic-cognition process. On the contrary, hierarchical semantic cognition (HSC) presented in this study integrates the four semantic layers together and considers their hierarchical relations. Four contribu-



Fig. 1. Examples of urban functional zones in satellite images.

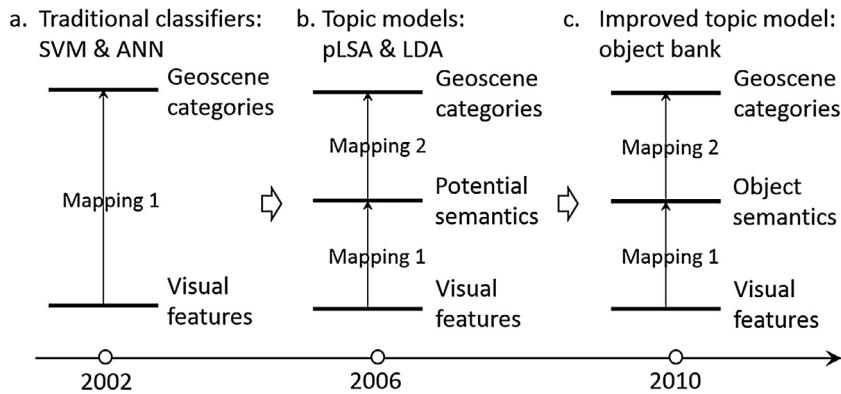
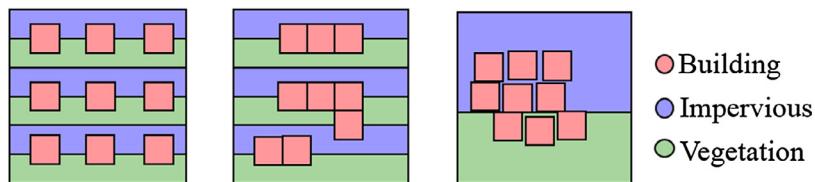
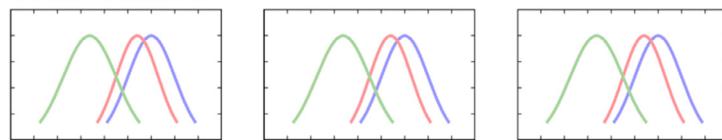


Fig. 2. History of image scene classification.

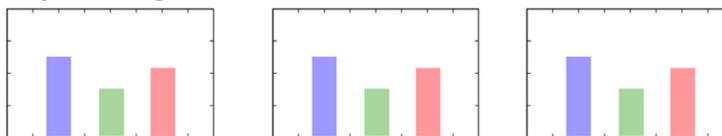
Concept graphs of three urban zones:



1. Visual features:



2. Object categories:



3. Spatial object patterns:



4. Zone functions:

“Villa district” “Residential zone” “Shanty town”

Fig. 3. The representations of three zones at four semantic layers. The three zones are totally composed of buildings, vegetation, and impervious objects which have similar proportions but different spatial patterns.

Table 1

The terminologies frequently used in this study.

Terminologies	Meanings
Functional zones	Spatially continuous and non-overlapping regions, with each one spatially composed of diverse objects, and semantically represented by a functional category
Zone classification	Labeling each zone with a functional category. In this study, the category system consists of six urban functions, i.e., commercial zones, parks, campuses, residential districts, shanty towns, and industrial zones
Semantic layers	Different layers for storing information of pixels, objects, object relations, and zones, respectively. They have hierarchical relations from finest units to coarsest ones, both spatially and semantically
Hierarchical semantic cognition (HSC)	A general cognition method for functional-zone analysis. It essentially integrates multiple semantic layers as well as their relationship into a hierarchical structure to characterize and recognize functional zones

tions have been made: (1) a novel hierarchical feature representation and the corresponding classification method for urban functional zones are first proposed; (2) POIs are fused with VHR images to provide object information, including object categories and spatial object patterns, for zone classification; (3) the importance of diverse semantic layers in HSC are quantitatively measured, and then attached with a geographic explanation; and (4) this study finally proves that HSC is more effective than single-layer cognition in mapping urban functional zones. Accordingly, the four contributions are totally novel and crucial to functional-zone studies. The novel definitions used in this study are outlined in Table 1.

2. Hierarchical semantic cognition

For convenience, the notations and terminologies used in hierarchical semantic cognition (HSC) are demonstrated in Section 2.1. Then, HSC's mathematical representation and reasoning process are presented in Section 2.2.

2.1. Notation and terminology

HSC is a mathematical model simulating the cognitive process for functional zones (Fig. 3), and considers the four semantic layers: visual features, object categories, spatial object patterns, and zone functions. The notations used at different layers are explained as follows.

- At visual-feature layer, each object is characterized by a feature vector \vec{w} with N elements including spectral, textural, and geometrical features. These features will be defined in Section 3.1, and are critical to recognize object categories;
- At object-category layer, all objects are categorized into K classes, and the k -th ($1 \leq k \leq K$) class is denoted by c_k ;

- At spatial-object-pattern layer, patterns of different kinds of objects are quantitatively measured by spatial statistic methods, and salient patterns are generated by clustering. Accordingly, the number of salient patterns is represented by J , and the j -th ($1 \leq j \leq J$) salient pattern is named as s_j ; and
- At zone-function layer, zones are classified into M functional categories, with f_i denoting the i -th ($1 \leq i \leq M$) category. In addition, the zone under classification is named as z .

2.2. Mathematical representation of HSC

HSC (Fig. 4) is a three-level Bayesian model with each level characterizing a relation between two semantic layers by using conditional probabilities. For example, the first level models the relationship between spatial object patterns and zone functions. As shown in Fig. 4, $p(s_j|f_i)$ denotes the probability of the spatial object pattern s_j appearing in the zones of f_i . Here, three kinds of spatial object patterns appear in f_i , and their probabilities are denoted as $p(s_1|f_i)$, $p(s_2|f_i)$, and $p(s_3|f_i)$ respectively. Then, $p(c_k|f_i, s_j)$ models the relationship between spatial object patterns and object categories. Taking $p(c_1|f_i, s_1)$ as an example, it refers to the proportion of c_1 in s_1 of f_i zones. For different spatial object patterns, the proportions of diverse objects are usually different. Finally, $p(\vec{w}|f_i, s_j, c_k)$ measures the probability distribution of c_k 's visual features in s_j of f_i . In the same spatial object pattern, different objects often have different distributions of visual features, e.g., $p(\vec{w}|f_i, s_1, c_1)$ and $p(\vec{w}|f_i, s_1, c_2)$; while in different spatial object patterns, the same kind of objects are allowed to have slight differences in their visual features' distributions, e.g., $p(\vec{w}|f_i, s_1, c_1)$ and $p(\vec{w}|f_i, s_2, c_1)$. Accordingly, four semantic layers and their relations are totally considered in HSC, and three conditional probabilities, i.e., $p(s_j|f_i)$, $p(c_k|f_i, s_j)$, and $p(\vec{w}|f_i, s_j, c_k)$, are used to characterize functional zones, which is the basic cue of zone classification.

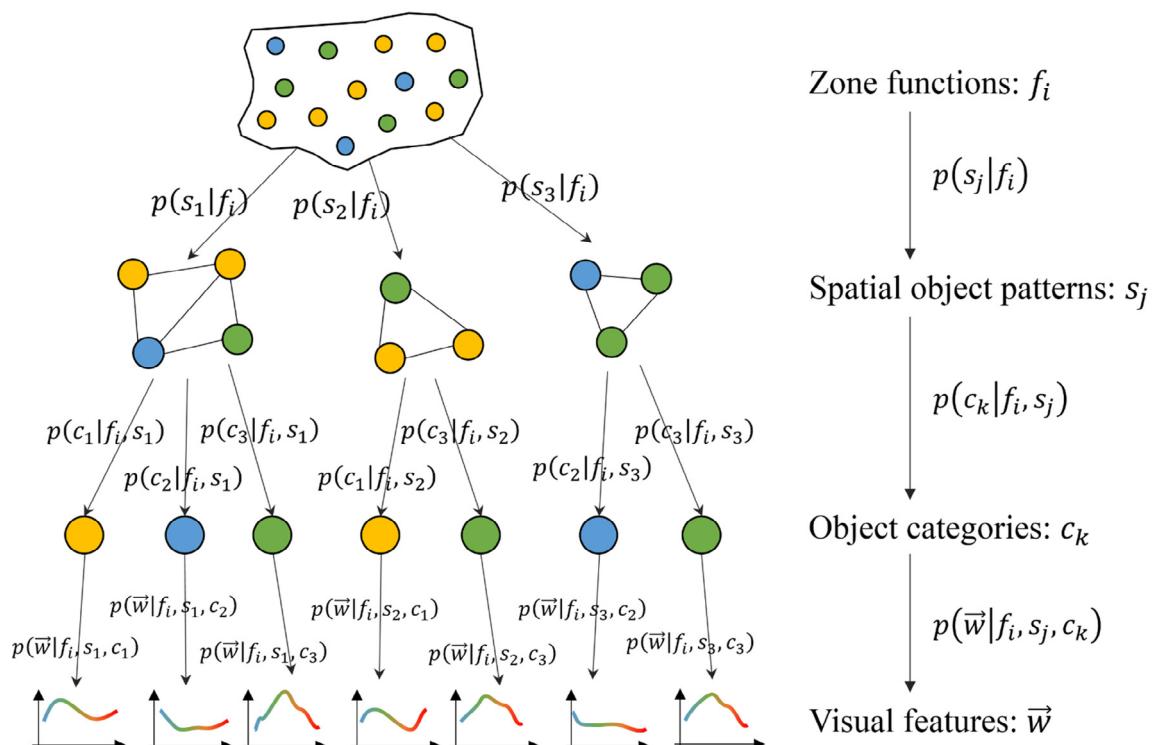


Fig. 4. Concept graph of hierarchical semantic cognition (HSC). The circles with different colors represent different kinds of objects, and the curves refer to their diverse visual features. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The corresponding classification method is demonstrated as follows.

In geophysical world, a zone z is a combination of diverse functions, thus its visual features $p(\vec{w}, z)$ can be represented as a linear combination of different categories' features $p(\vec{w}|f_i)$:

$$p(\vec{w}, z) = \sum_{i=1}^M p(\vec{w}, f_i, z) = \sum_{i=1}^M p(z) * p(\vec{w}, f_i | z) = p(g) \sum_{i=1}^M p(\vec{w}|f_i) * p(f_i | z) \quad (1)$$

where $p(f_i | z)$ represents the probability that the zone z belongs to the functional category f_i , and it finally determines the classification result. In order to get $p(f_i | z)$, other parameters, i.e., $p(\vec{w}|f_i)$, have to be calculated first. $p(\vec{w}|f_i)$ refers to the visual features inside f_i , and it is hard to estimate, as visual features are variant in f_i . For example, the f_i in Fig. 4 is composed of several spatial object patterns with different visual features. Accordingly, $p(\vec{w}|f_i)$ is mixed and has to be decomposed by considering diverse spatial object patterns, as shown in Eq. (2).

$$p(\vec{w}|f_i) = \sum_{j=1}^J p(\vec{w}, s_j | f_i) = \sum_{j=1}^J p(\vec{w}|f_i, s_j) * p(s_j | f_i) \quad (2)$$

In Eq. (2), $p(s_j | f_i)$ refers to the probability of j -th spatial object pattern occurring in f_i , and $p(\vec{w}|f_i, s_j)$ denotes the visual features in s_j of f_i . However, $p(\vec{w}|f_i, s_j)$ is still mixed as diverse objects with variant features co-occur in s_j ; thus, it should be further decomposed with respect to different object categories:

$$p(\vec{w}|f_i, s_j) = \sum_{k=1}^K p(\vec{w}, c_k | f_i, s_j) = \sum_{k=1}^K p(\vec{w}|f_i, s_j, c_k) * p(c_k | f_i, s_j) \quad (3)$$

where $p(c_k | f_i, s_j)$ represents the frequency of c_k in s_j of f_i . $p(\vec{w}|f_i, s_j, c_k)$ refers to visual features of c_k , and it is relative pure with low variation compared to $p(\vec{w}|f_i)$ and $p(\vec{w}|f_i, s_j)$. Accordingly, $p(\vec{w}, z)$ is decomposed into four semantic layers with a hierarchical structure (Fig. 4) and finally represented by Eq. (4).

$$p(\vec{w}, z) = p(z) * \left\{ \sum_{i=1}^M p(f_i | z) * \left[\sum_{j=1}^J p(s_j | f_i) * \sum_{k=1}^K p(\vec{w}|f_i, s_j, c_k) * p(c_k | f_i, s_j) \right] \right\} \quad (4)$$

Apart from Eq. (4), $p(\vec{w}, z)$ has another representation form, as reported in Eq. (5).

$$p(\vec{w}, z) = p(z) * p(\vec{w}|z) = p(z) * \sum_{j=1}^J p(s_j | z) * \sum_{k=1}^K p(\vec{w}|s_j, c_k, z) * p(c_k | s_j, z) \quad (5)$$

where $p(s_j | z)$ denotes the probability that z 's spatial object pattern belongs to s_j . $p(c_k | s_j, z)$ refers to the c_k 's proportion in z , and its features are represented by $p(\vec{w}|s_j, c_k, z)$. Accurately, z has only one kind of spatial object pattern which is denoted as s_t ($1 \leq t \leq J$), and thus $p(s_j | z) = 0$, when $j \neq t$. Accordingly, The Eq. (5) can be simplified as follow:

$$p(\vec{w}, z) = p(z) * p(s_t | z) * \sum_{k=1}^K p(\vec{w}|s_t, c_k, z) * p(c_k | s_t, z) \quad (6)$$

Combining the two equations of Eqs. (4) and (6), the Eq. (7) finally models the relationship between the under-classification zone z and diverse functions f_i , which is so called HSC. In HSC, $p(f_i | z)$ determines classification result of z and needs to be finally calculated; while, other parameters can be obtained from data. The HSC essentially employs hierarchical structures to characterize both z and $\{f_i | 1 \leq i \leq M\}$, and it is the first model which can math-

ematically model the cognitive process for functional zones where four semantic layers are integrated together and their hierarchical relations are also modeled.

$$\begin{aligned} p(s_t | z) * \sum_{k=1}^K p(\vec{w}|s_t, c_k, z) * p(c_k | s_t, z) \\ = \sum_{i=1}^M p(f_i | z) * \sum_{j=1}^J p(s_j | f_i) * \sum_{k=1}^K p(\vec{w}|f_i, s_j, c_k) * p(c_k | f_i, s_j) \end{aligned} \quad (7)$$

3. Methodology

In this section, HSC will be used to classify urban functional zones, and four semantic layers are respectively considered (Fig. 5). First, every zone is segmented into objects, and each object is characterized by multiple features, including spectral, textural, and geometrical ones. Then, these objects are classified with using these features and POI samples, and their categories can be obtained. Third, spatial object patterns in functional zones are measured by G functions considering diverse objects, and the salient patterns are detected by clustering. Finally, functional zones are classified based on HSC (Fig. 5).

3.1. Extracting visual features

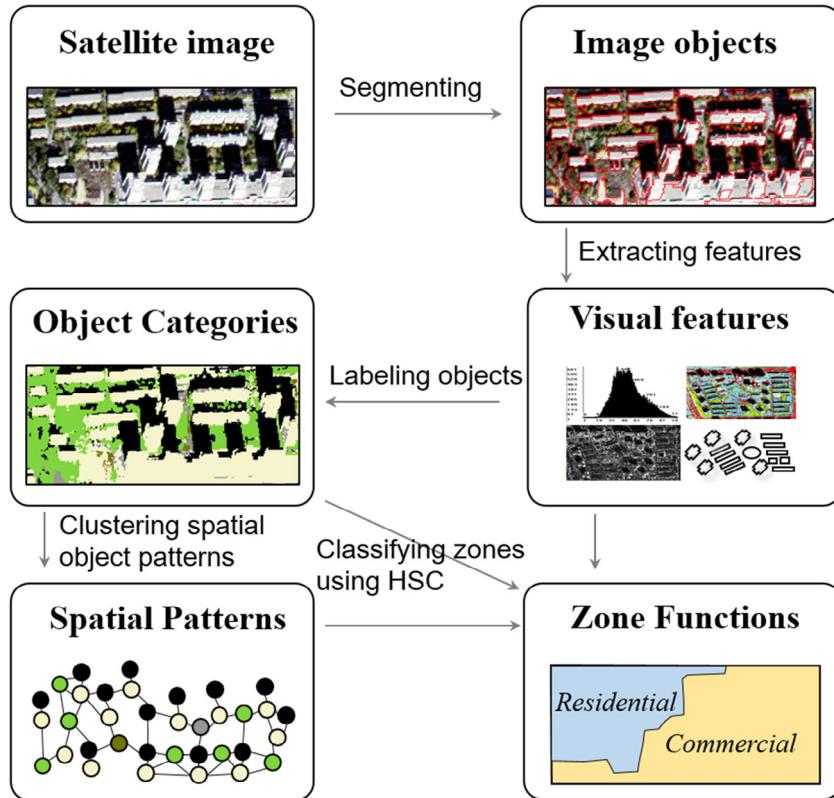
At the bottom semantic layer, multiple visual features are extracted to characterize geographic objects. We first segment remote sensing images into objects using multiresolution segmentation method (Baatz and Schape, 2000), which experts in delineating objects at different scales. Then, three kinds of object features are extracted to characterize objects (Table 2). These features are effective to measure objects from different aspects, and have been widely used in most object-based image analyses (Cleve et al., 2008; Laliberte et al., 2004; Peña-Barragán et al., 2011; Stow et al., 2008).

3.2. Labeling object categories

Objects serve as the basic components of zones, and their categories are fundamental to recognizing zone functions. However, objects always have different category systems which mainly fall into two types: land covers and land uses. This study focuses on functional analysis, and thus land uses are considered here (Heiden et al., 2012), but accurate land-use information is hardly available in remote sensing images. To resolve this issue, point of interest (POI) data are used to label land-use categories to objects.

POI data are easily available in most volunteered geographic information platform, and allow us to learn object categories based on human cognition at spatial, temporal, and semantic granularity (Hu et al., 2016; McKenzie et al., 2014). They often include educational institutes, commercial points, residential buildings, parks, companies, and public services, which are different from land covers, e.g., vegetation, soil, water, and buildings; thus, they are fundamental to analyzing functional categories of urban zones. However, the qualities of POIs vary among categories. For example, the commercial POIs are significantly more than other types, as POI data are produced by human activities, and focus mainly on commercial institutes more than others. It will lead to an unbalanced distribution of POIs, and thus the original POIs cannot represent geographic objects well with respect to volumes and spatial distributions. Accordingly, POI data should be modified to be best consistent with the ground truth. To achieve this purpose, we use the POIs as training samples to build classifier for labeling object categories.

Random forest (RF), an ensemble classifier (Breiman, 2001), is used here to label object categories, as it is effective to distinguish

**Fig. 5.** Procedure of functional-zone classification by using HSC.**Table 2**
Visual features for characterizing objects.

Types	Names	Meanings
Spectral	[Mean]	Average spectrum of pixels
	[Std. Dev.]	Gray standard deviation of pixels in an object
	[Skewness]	Skewness of spectral histogram
	[Contrast]	The average differences between border pixels and their neighborhood
	[Border]	Normalized Difference Vegetation Index
	[NDBI]	Normalized Difference Building Index
Textural	[GLDV]	The vector composed of diagonal elements of GLCM
	[Homogeneity]	The homogeneity derived from GLCM
	[Dissimilarity]	The heterogeneity parameters derived from GLCM
	[Entropy]	Information entropy derived from GLCM
	[Correlation]	Correlation of pixels which is derived from GLCM
Geometrical	[Area]	The number of pixels within image objects
	[Length/Width]	Length-width ratio of the object's MBR
	[Eclipse Fit]	The fitting degree of eclipse fit
	[Main Direction]	Eigenvectors of covariance matrix
	[Shape Index]	The ratio of perimeter to four times side length

diverse land-use objects, which has been experimentally validated by Du et al. (2015). RF is composed of multiple decision trees, and uses their votes to decide the final classification results (Breiman, 2001). RF however can generate some misclassification results for outliers, which can further impact zone classifications. To resolve this issue, the unreliable classification results will be

detected and further abandoned. The reliability of a classification result is defined as follow:

$$R(\text{Object}_t) = \frac{\max(vote_k(\text{Object}_t))}{\text{TreeNumber}} \quad (1 \leq k \leq K) \quad (8)$$

where $R(\text{Object}_t)$ refers to the reliability of Object_t 's classification result, and TreeNumber represents the number of decision trees. $vote_k(\text{Object}_t)$ denotes the votes of k -th object category. Consequently, the 80% most reliable classification results will be kept, and others will be abandoned. This method is effective to generate object categories which are important in functional-zone classification.

3.3. Clustering spatial object patterns

Spatial object pattern aims to measure spatial distributions of objects considering their spatial relations (Ri and Yao, 2015; Yang et al., 2013). In this study, objects are represented by their centroids, and thus their spatial relations can be measured by centroid distances. As a result, G curves can be used to represent spatial object patterns (Gatrell et al., 1996). G curve measures the cumulative distribution of nearest-neighbor distances (NND), i.e., the distance from one object to its nearest neighbor (Haase, 1995). In practice, G curve is able to distinguish aggregation, random distribution, and evenly spaced patterns (Diggle et al., 1976) which are respectively shown in Fig. 6(a)–(c). The G curve of aggregation pattern soars within a short distance (Fig. 6a); while that of evenly spaced pattern first rises slowly, and then rapidly (Fig. 6c). Random distribution is related to an evenly rising G curve which is shown in Fig. 6(b).

In practice, a G curve is used to measure a kind of objects' spatial patterns; while, different kinds of objects often have different spatial patterns, and thus are related to different G curves. Suppos-

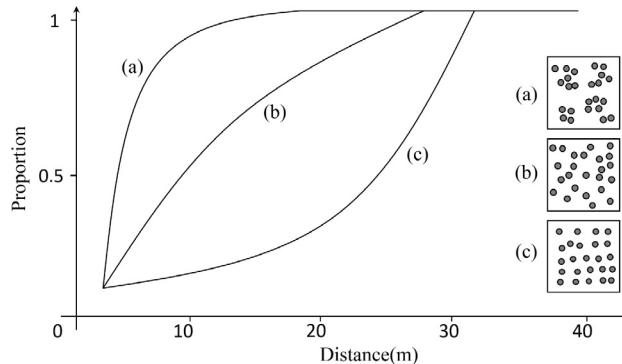


Fig. 6. G curves for characterizing three spatial patterns: (a) aggregation, (b) random distribution, and (c) evenly spaced pattern.

ing that the G curve of k -th class of objects (c_k) in v -th zone (z_v) is denoted by $\vec{G}(c_k, z_v)$, then $\vec{G}(z_v) = \{\vec{G}(c_1, z_v), \vec{G}(c_2, z_v), \dots, \vec{G}(c_K, z_v)\}$ can quantitatively measure the spatial patterns of diverse objects in the zone z_v .

HSC needs the salient spatial-object-patterns. Accordingly, ISODATA algorithm (Memarsadeghi et al., 2007) is used to automatically cluster spatial object patterns based on $\vec{G}(z_v)$. ISODATA needs little prior information and experts in detecting salient clusters (Memarsadeghi et al., 2007); thus, it can generate salient spatial-object-patterns which are represented by clustering centers. This method can automatically and accurately generate salient spatial-object-patterns for zone classification.

3.4. Classifying functional zones with HSC

In terms of the three steps (Sections 3.1–3.3), the three semantic layers, i.e., visual features, object categories, and spatial object patterns, can be established. They will be integrated into a HSC frame to characterize zones, and further to classify their functional categories (Fig. 4).

To achieve this purpose, we need to establish the HSC and then estimate its parameters. $p(f_i|z)$ in Eq. (7) represents the probability of zone z belonging to the functional category f_i . In order to calculate $p(f_i|z)$, other parameters in HSC should be obtained. The three parameters, i.e., $p(s_j|f_i)$, $p(c_k|f_i, s_j)$, and $p(\vec{w}|f_i, s_j, c_k)$, respectively denote the three relations between the four semantic layers. For example, $p(s_j|f_i)$ represents the relationship between s_j and f_i , where the spatial object pattern s_j is measured in Section 3.3, and the zone function f_i is defined in training samples. Likewise, other two parameters, i.e., $p(c_k|f_i, s_j)$, and $p(\vec{w}|f_i, s_j, c_k)$, can be obtained from object categories (Section 3.2) and visual features (Section 3.1). Additionally, other parameters in HSC, i.e., $p(s_t|z)$, $p(c_k|s_t, z)$, and $p(\vec{w}|s_t, c_k, z)$, mathematically model the hierarchical representation of the under-classification zone z , and can be directly extracted from z . $p(s_t|z)$ refers to the probability that z 's spatial object pattern belongs to s_t , which is represented by the distance from z to s_t 's cluster center in feature space. Additionally, $p(c_k|s_t, z)$ represents the proportion of c_k in z ; while, $p(\vec{w}|s_t, c_k, z)$ denotes the visual features of c_k in z . As all of those parameters are calculated, $p(f_i|z)$ as the only unknown one can be finally estimated by using Original Least Square (Waseem et al., 2013). As a result, z is labeled with $\text{Category}(z) = \text{argmax}_{i=1}^M p(f_i|z)$.

The classification method essentially uses a hierarchical structure to characterize zones. This structure contains the information at four semantic layers as well as their mapping relations, and is employed as the cue to recognize zone functions.

4. Study area and data collection

This study focuses on urban functional zones, and thus a site covering 67.1 km^2 urban area (Fig. 7a) in Beijing, China, is chosen which spreads from developing (Fig. 7b) to developed (Fig. 7c and d) conditions. As the capital city and cultural center of China, Beijing is composed of diverse archaic and modern zones with different architectural styles, and its environmental as well as social issues have recently raised worldwide concerns. To study this area, we should investigate the functional zones there and generate a functional-zone map. The functional zones in this city are greatly heterogeneous, and are difficult to be accurately classified, which challenges our methods.

Here, a QuickBird image covering the study area is employed. The QuickBird image was acquired in Mar., 2002, and its multi-spectral bands are merged with the panchromatic band to produce the pan-sharpened image of 0.61 m resolution with four bands (Rahmani et al., 2010). The corresponding road vectors are used to delineate functional zones, and each zone is spatially represented by a block. Consequently, 617 zones are generated (Fig. 8). Additionally, 24824 POIs are used to label land-use categories to objects. These POIs are collected in 2002 and they are sorted into six classes, i.e., commercial points (7977), companies (3043), public services (3157), scenic spots (2075), residential buildings (5112), and educational institutes (3460). As shown in Fig. 8, the used image and vector data are well integrated.

5. Experimental results

In this section, we used the proposed methods (Section 3) to classify urban functional zones in Beijing, and the results of each step are reported and analyzed as follows.

5.1. Results of object categories

First, the multiresolution segmentation method (MRS; Baatz and Schape, 2000) is used to segment the QuickBird image into multi-scale objects, where the shape parameter is set as 0.1 and the compactness set as 0.5. Among the multiple scales, an appropriate scale of 60 is selected based on the estimation tool of scale parameter (ESP; Drăguț et al., 2010, 2014). As a result, 116201 objects are generated. For each object, its visual features including spectral, textural, and geometrical ones (Table 2) are extracted. Then, the POIs in Fig. 8 are used to classify land-use categories of objects. For the objects which are overlapped with POIs, they are labeled by the corresponding POI's category, and if an object contains more than one POIs, the object will be labeled by the most frequent POI's category. The labeled objects are employed to train a random forest for recognizing other objects. The overall classification accuracy is 85.3%. As demonstrated in Section 3.2, only 80% classification results with high reliabilities (Eq. (8)) will be kept. Accordingly, 92,961 objects are kept and their categories are used to building the object-category layer in HSC. For convenient visual comparison, the kept classification results are represented by the semantic points shown in Fig. 9(b).

As shown in Fig. 9, the original POI data are too sparse to represent ground objects. For example, many objects have no corresponding POI (Fig. 9c). It is because that POIs are manually produced and focus on commercial institutes, and thus they are inconsistent with geographic objects. On the contrary, the generated semantic points (Fig. 9b) were relatively dense, and more importantly they have consistent distributions with geographic objects, which is shown in Fig. 9(d). Accordingly, the generated semantic points are more appropriate than original POI data to represent object categories. Furthermore, the numbers of generated

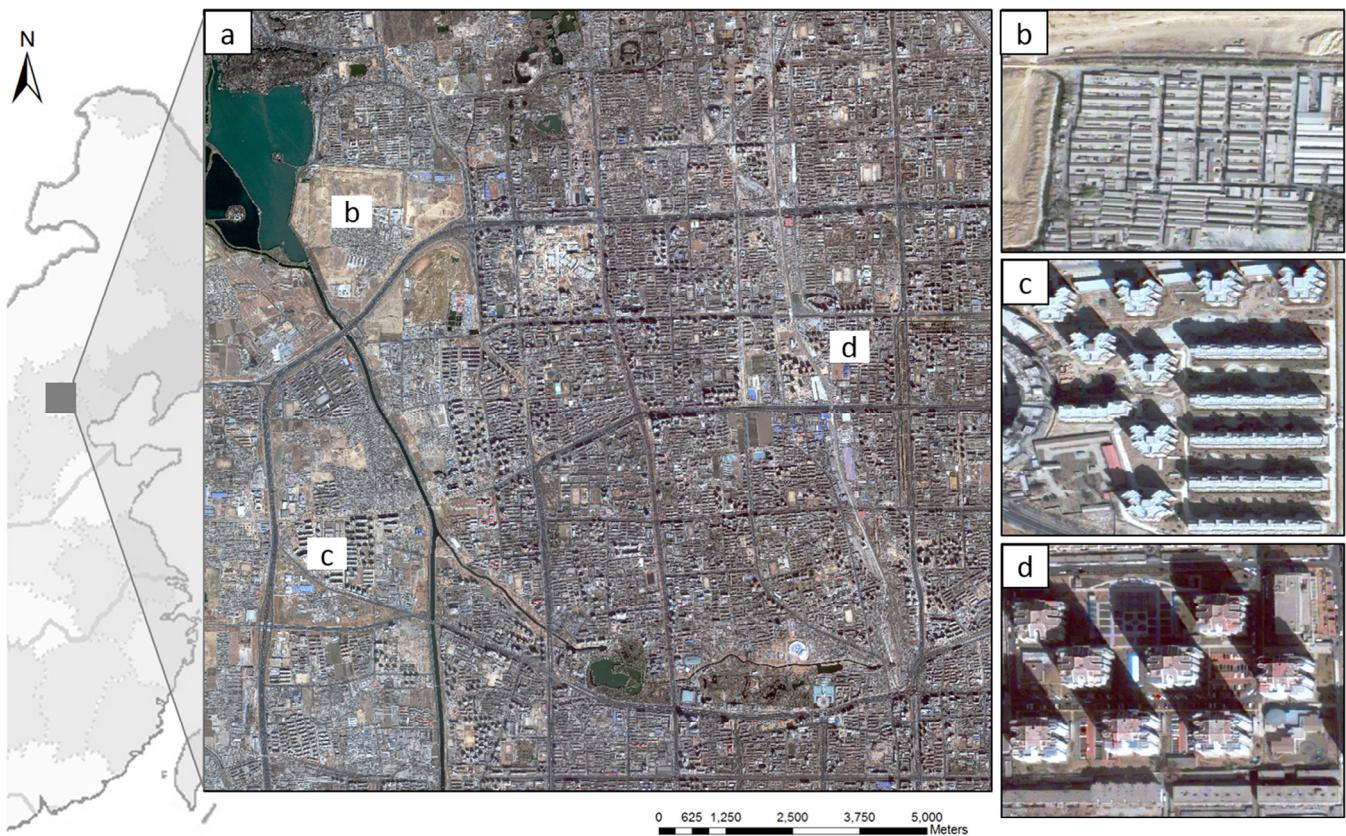


Fig. 7. The study area in (a) Beijing, China (QuickBird image in band combination 3/2/1, true color). The three sub-regions (b)–(d) indicate the heterogeneity of functional zones in the study area. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

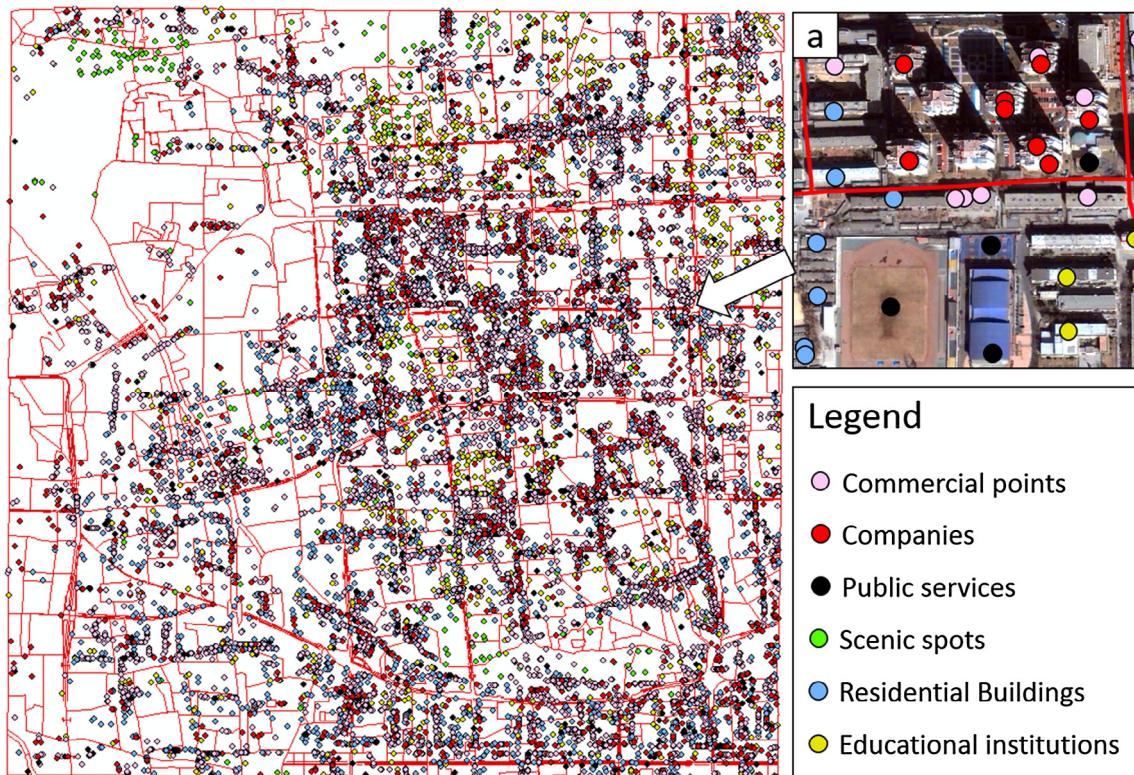


Fig. 8. Vector data including road lines and POIs. The sub-image (a) demonstrates the relationship between QuickBird image and vector data.

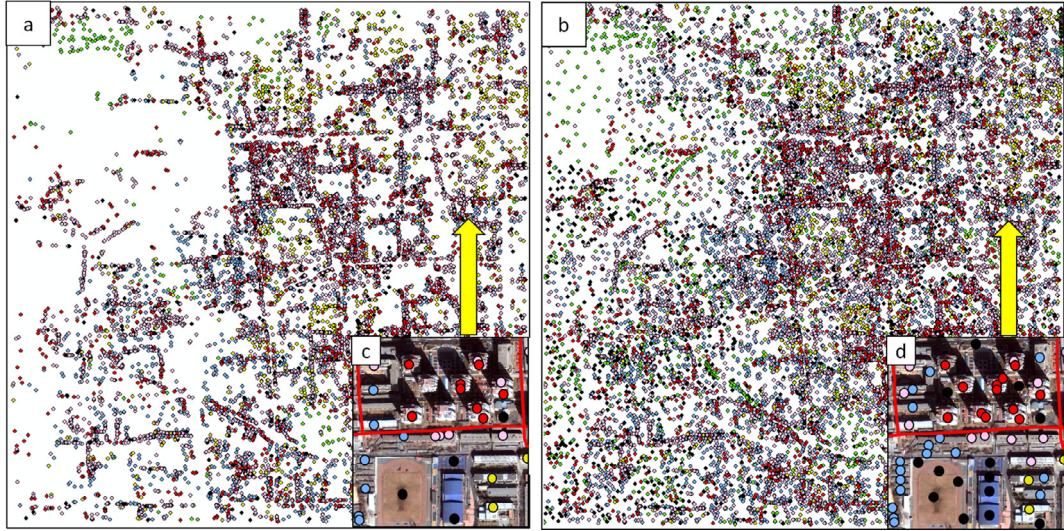


Fig. 9. A comparison of (a) original POIs and (b) the generated semantic points. A sub-region was chosen to show the detailed differences between (c) and (d).

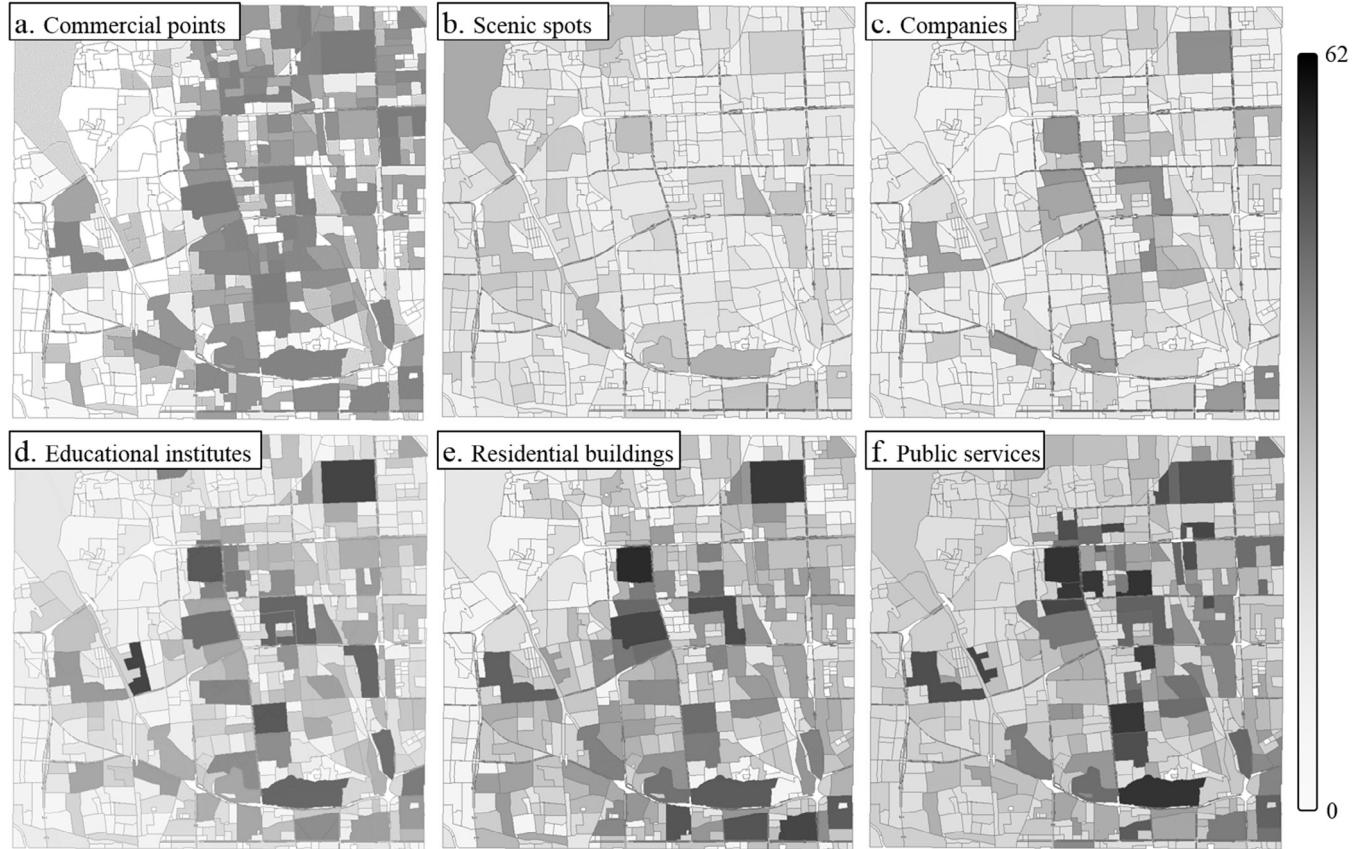


Fig. 10. Numbers of diverse land-use objects inside functional zones. The numbers are limited within the scale [0, 62], and stretched into [0, 255] as the gray levels in the sub-figures.

semantic points in each zone are measured (Fig. 10). For the same kind of semantic points, their density varies from zone to zone. Taking the commercial points as an example, we found that they are dense on the right part of the study area, but sparse on the left. Additionally, within a zone, the densities of different kinds of points are various. For example, there are many scenic spots in the Summer Palace (the largest zone in the study area), but few companies.

The generated object categories play an important role in functional-zone classification, and they construct the object-category layer in HSC.

5.2. Results of spatial object patterns

As demonstrated in Section 3.3, a G curve characterize a kind of objects' spatial pattern, and multiple G curves can measure diverse

objects' patterns. Here, the spatial object patterns in a functional zone will be characterized by six G curves, as six kinds of objects are considered. The salient object patterns are clustered by ISO-DATA based on G curves (Section 3.3), and four salient patterns are generated which are shown in Fig. 11(a)–(d).

Generally, the convex G curves represent the aggregated patterns; while concave ones denote evenly distributed objects (Gatrell et al., 1996). For the first type of spatial patterns (Fig. 11a), residential buildings are evenly spaced, but other objects are aggregated. For the second type (Fig. 11b), the residential buildings and public services are related to aggregation distributions, and others are randomly distributed. Additionally, the spatial object patterns shown in Fig. 11(c) and (d) are much more complex than the previous two types. In Fig. 11(c), scenic spots and public services are densely aggregated, educational institutes are evenly spaced, other objects are randomly distributed; while in Fig. 11(d), companies, residential, and educational objects are aggregated, scenic spots are randomly distributed, and public services and commercial points are evenly spaced. These four types are significantly different from each other, and thus are effective to distinguish diverse zone functions.

5.3. Results of zone classification

For the final step of zone classification, HSC is implemented by C++ and it is used to sort the 617 urban zones into the six functional categories: commercial zones, parks, campuses, residential districts, shanty towns, and industrial zones. Fig. 12(c) reports the classification results. To validate the effectiveness of the pro-

posed methods, we compared HSC with two classical methods, i.e., Support Vector Machine (SVM; Suykens and Vandewalle, 1999) and Latent Dirichlet Allocation (LDA; Blei et al., 2003). Differently, SVM incorporates visual features into one layer (Table 2), while LDA encodes both visual features and object categories into two layers.

The three methods SVM, LDA, and HSC are trained by the same samples. Their classification results are respectively shown in Fig. 12(a)–(c). Based on the visual interpretations in Fig. 12(d)–(f), we found that HSC produced the most accurate results. For example, Fig. 12(d) displays a residential district which is correctly recognized by HSC, but misclassified as shanty towns by SVM and LDA. Likewise, the zone in Fig. 12(e) contains many educational institutes, including two colleges, two elementary schools, and a daycare, and thus this zone should be a "campus", instead of a "residential district" labeled by SVM or a "shanty town" by LDA; while, HSC makes the correct classification, as it uses generated object categories, e.g., educational institutes, to make recognition. Additionally, SVM and LDA make wrong classification results for Fig. 12(f), and they both classify the zone as a "shanty town", because it covers many soils and is similar in visual features with "shanty towns". Actually, it is an under-construction commercial zone, and it is difficult to be accurately recognized. Nevertheless, HSC considers spatial object patterns, such as the relationship between shadows and buildings, thus it can distinguish the zone from "shanty towns", and correctly classify its function.

In spite of the good visual performances of HSC, the three methods are further quantitatively compared with respect to their overall accuracies. The 617 zones are manually recognized based on the

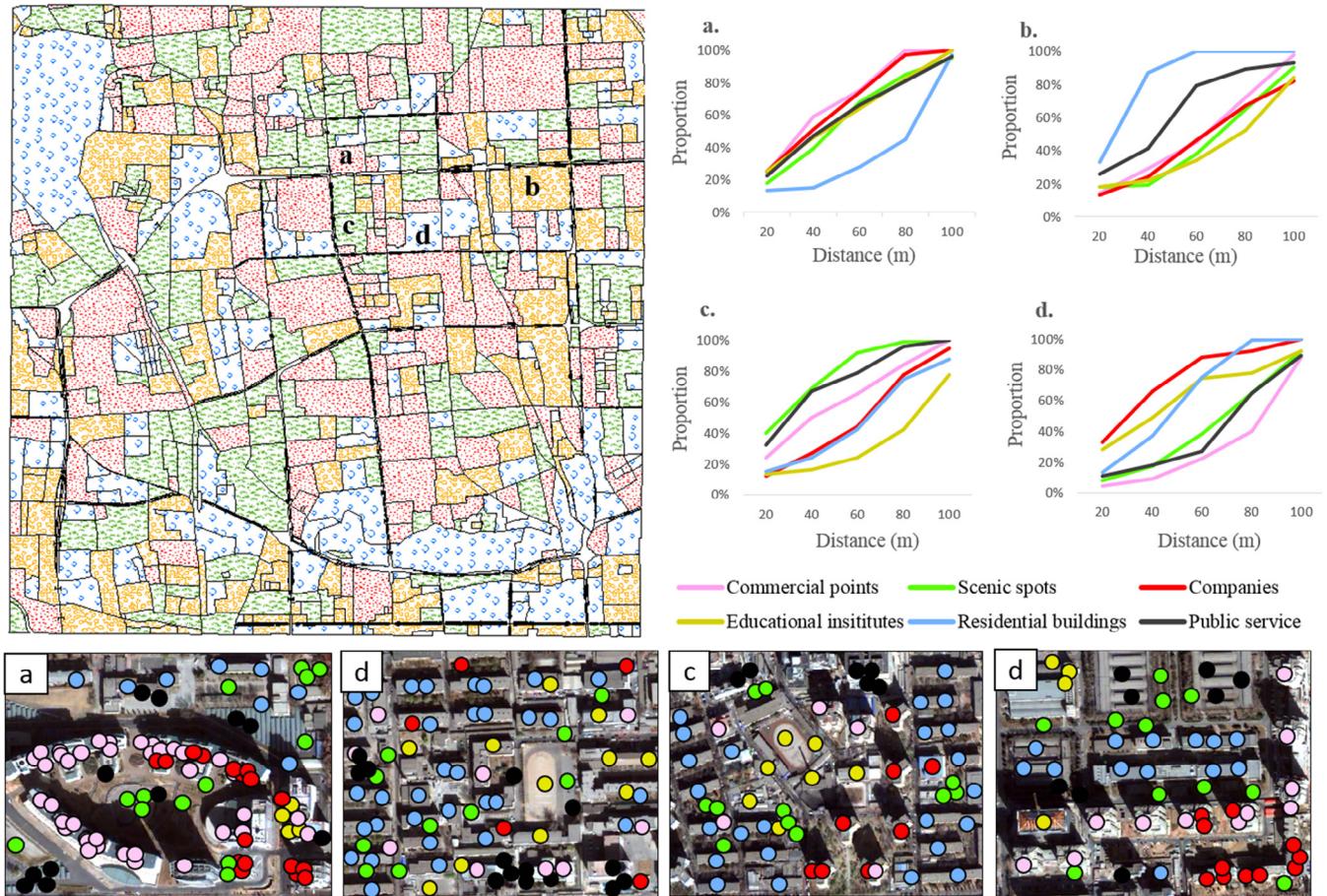


Fig. 11. Four types of spatial object patterns. Correspondingly, four functional zones (a)–(d) are selected as examples with their G curves reported at the top right corner.

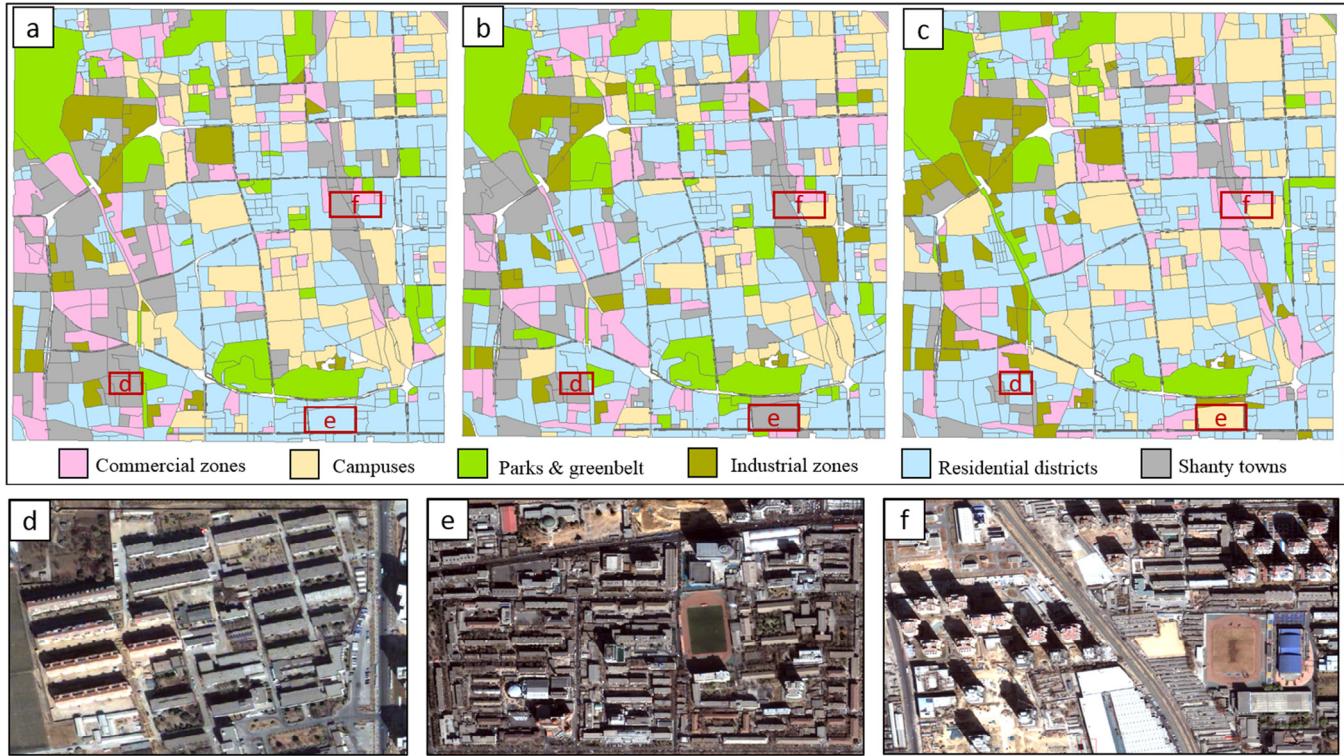


Fig. 12. Three functional-zone classification results by using (a) SVM, (b) LDA, and (c) HSC. Three regions (d)–(f) are chosen to visually compare the three results.

QuickBird image and POIs, and the results are considered as the ground truth to evaluate the performances of three classification methods; thus, the classification accuracies are evaluated based on 2002-information. As reported in Table 3, HSC produces the most accurate results for most categories. Taking “commercial zones” as an example, SVM and LDA produce poor classification results with small producer’s accuracies (70.3% and 75.2%); while, HSC shows excellent performance in recognizing commercial zones with a large accuracy of 91.2%, as it considers object categories, e.g., commercial points and companies, which can significantly improve the differentiation of commercial zones from other categories. However, HSC is weaker than SVM only in recognizing “industrial zones”, as POI data contains little information on industrial objects and thus the proposed methods contribute little to industrial zones’ classification. Nevertheless, HSC outperforms SVM and LDA in urban zone classification with a larger overall accuracy (90.8%). By reference to the classification results, it can be concluded that the HSC is more effective than SVM and LDA in recognizing urban functional zones.

5.4. Results of urban structure analysis

As summarized above, it has been validated that our methods are effective to classify urban functional zones and generate more

accurate results than classical methods. Furthermore, the generated functional-zone map can be used to analyze this city’s functional structures.

The proportions of diverse functional zones are measured and shown in Fig. 13. At that time, residential districts in Beijing accounted for more than 1/3 urban area. It is because that the city had a large population which needed so many residential districts to hold. Additionally, there were educational zones which accounted for 18.62%. Parks and commercial zones respectively covered 15.30% and 12.17% land in this city; while, the industrial zones and shanty towns were minimum with smallest proportions, 8.26% and 7.40%. Accordingly, it can be concluded that Beijing had a large residential area, and attached great importance to educational development. Few industrial zones and shanty towns indicated the high-urbanization level of this city. However, Beijing’s commerce at that time was underdeveloped, with respect to the proportion of commercial zones.

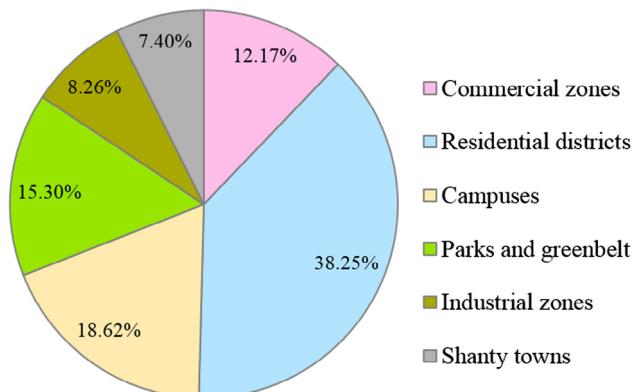


Fig. 13. Proportions of diverse functional zones in the study area.

Table 3
Classification accuracies of the three classifiers.

	SVM (%)	LDA (%)	HSC (%)
Commercial zones	70.3	75.2	91.2
Residential districts	81.5	85.0	92.5
Campuses	79.7	88.2	90.0
Parks and greenbelt	91.1	96.7	96.7
Industrial zones	83.3	81.5	82.7
Shanty towns	79.4	79.8	83.2
Overall accuracies	80.3	83.8	90.8

6. Discussions

6.1. A comparison of hierarchical and non-hierarchical methods for functional zones

It has been verified that the HSC outperformed SVM and LDA in classifying urban functional zones, but it is still unknown whether multiple features or the hierarchical-cognition structure makes the difference. To answer this question, two cognition structures, i.e., a hierarchical one and a non-hierarchical one, are compared.

The hierarchical semantic cognition (HSC) insists that hierarchical relations between semantic layers are important for zone classification and should be considered in feature representations and classification models (Cohen, 2000; Wang et al., 2010). On the contrary, many existing studies on image scene classification concentrated on non-hierarchical semantic cognitions (nHSC) including RF, SVM, and neural network, and they argued that the multiple features, instead of their hierarchical relations, make differences in zone classification. Accordingly, nHSC usually employed a super feature vector composed of multiple features to characterize functional zones but overlooked the hierarchical constraints and the relationships among features (Farabet et al., 2013; Yao et al., 2016; Zhang et al., 2015; Zhao and Du, 2016).

The two cognition structures, i.e., HSC and nHSC, are compared by reference to their urban-zone-classification results in Beijing. For the HSC, its classification process and results are reported in Section 5. For the nHSC, we similarly extract multiple features at different semantic layers (e.g., visual features, frequencies of object categories, and G curves of spatial object patterns) to characterize each zone and integrate them into a high-dimension feature vector without distinguishing their roles; then, the extracted feature vectors are used to train a SVM classifier which is regarded as a nHSC model (Suykens and Vandewalle, 1999). The two structures' classification results are reported in Table 4. For all categories, HSC produces more accurate results, especially for commercial zones and residential districts, their accuracies are 7.5% and 8.1% higher than those generated by nHSC. This phenomenon can be explained from three aspects. First, the used feature vector in nHSC is very large in volume, as it is composed of multiple features at different semantic layers, and the high-dimension features do harm to zone classification because of "curse of dimensionality" (Indyk and Motwani, 1998); while, features in HSC are represented by a finite number of conditional probabilities and thus are much smaller in volume than the feature vectors in nHSC. Second, the correlations between features are totally ignored by nHSC, but are correctly modeled as the hierarchical relationship by HSC. Third, default values in nHSC, i.e., some features with no value, can significantly influence classification results. For example, a park has no commercial objects, and the G curve of commercial objects will be set as default values. In the same case, HSC can effectively resolve the default values, as it uses conditional probability of salient clusters instead of directly using G-curve values to represent spatial object patterns. Accordingly, it can be concluded that HSC outperforms nHSC.

Table 4
A comparison of HSC and nHSC according to their classification accuracies.

	HSC (%)	nHSC (%)
Commercial zones	91.2	83.7
Residential districts	92.5	84.4
Campuses	90.0	87.1
Parks and greenbelt	96.7	94.3
Industrial zones	82.7	80.8
Shanty towns	83.2	80.6
Overall accuracies	90.8	85.2

6.2. Contributions of different semantic layers to zone classification

In this section, we will discuss the contributions of the three semantic layers (i.e., visual features, object categories, and spatial object patterns) to urban zone classification.

First, visual features are fundamental to characterizing objects, and are important for distinguishing functional zones (Lillesand et al., 2014). For example, the features of the buildings in Fig. 7 (b)–(d) are totally different. Then, object categories are also important for zone classification (Li et al., 2010), as their proportions vary from zone to zone. For example, the zone of Fig. 12(e) is a campus and composed of 61.7% educational institutes, 22.8% residential buildings, and 15.5% other objects; while, the one in Fig. 12(f) is a commercial zone and it consists of 35.7% commercial points, 23.7% companies, 19.1% residential buildings, and 16.5% others. Third, spatial object patterns aim to measure spatial information of objects, and they can enhance the differences between different zones. For example, the companies are aggregated in commercial zones but evenly spaced in residential districts.

As demonstrated above, these semantic layers play different roles in functional-zone classification. Furthermore, to quantitatively measure their contributions, we present an evaluation method based on HSC. Supposing that spatial-object-pattern layer is under evaluation, $p(s_t|z)$ and $p(s_j|f_i)$ in Eq. (7) are set as 1. In this case, spatial object patterns make no difference to zone classifications, which can cause accuracy decrease compared to original HSC's classification results. The decrease degree is regarded as the contribution of spatial-object-pattern layer. Likewise, while evaluating object-category layer, both $p(c_k|s_t, z)$ and $p(c_k|f_i, s_j)$ can be set as 1. $p(\tilde{w}|s_t, c_k, z)$ and $p(\tilde{w}|f_i, s_j, c_k)$ will be set as 1 to evaluate visual features' contribution. Accordingly, the contributions of the three semantic layers are measured and reported in Table 5. For campuses, residential districts, commercial zones, and shanty towns, object categories contribute most to their classification results, as the used objects, e.g., commercial points, companies, residential buildings, and educational institutes, make sense to distinguish these zone functions. On the other hand, visual features are more important for recognizing other two categories, especially for industrial zones. It is because that there is no POI data related to industrial objects, and thus object categories and spatial object patterns make less contributions. Although spatial object patterns make the least contribution, their effects cannot be ignored. For example, they are effective in distinguishing residential districts from shanty towns, as both of them have the same components, residential buildings, but are different in spatial building patterns. As demonstrated above, the semantic layers play different but totally important roles in zone classifications.

6.3. Pros and cons of HSC

We believe that there can be and will be more methods to use diverse-semantic-layer information to classify functional zones,

Table 5
Contributions of different semantic layers to zone classification. The bold numbers represent the greatest contribution for each function.

	Visual features (%)	Object categories (%)	Spatial patterns (%)
Campuses	11	72	17
Residential districts	12	54	34
Commercial zones	5	91	4
Shanty towns	40	49	11
Parks and greenbelt	52	39	9
Industrial zones	72	21	7
Total	33	54	13

but we regard the HSC as an effective one from the following three aspects. First, HSC adopts a hierarchical structure to characterize every functional zone, instead of simply combining multiple features into one layer. HSC not only considers the four semantic layers, but also measures their hierarchical relations which are important for zone classification (Wang et al., 2010). Then, HSC fuses POIs and VHR images together to represent functional zones, instead of extracting features respectively from these data, which can provide more accurate land-use information for analyzing urban zones. Third, every parameter in HSC has a clear geographic meaning, and thus the whole classification process has a strong interpretability, which is lacked in traditional classifiers, e.g., SVM and deep learning (Yan et al., 2003; Zhao and Du, 2016; Zhou et al., 2014). Accordingly, HSC is effective in functional-zone classification.

In addition, HSC is not only a classifier for urban functional zones, but also serves as a general cognition structure for other image analysis tasks and landscape investigations. First, HSC can decompose images into four semantic layers, and it is able to analyze image representations, components, structures, and semantics in a consistent cognition system. Then, from the landscape-ecology point of view, HSC can couple landscape compositions, structures, and functions together (Fu et al., 2011). Landscape compositions are characterized by object categories, structures can be measured by spatial object patterns, and functions are represented by functional categories; thus, landscape compositions, structures, and functions can be best represented in the HSC. Third, HSC serves as a skeleton frame for fusing multisource geographic data (Gahegan and Ehlers, 2000). Visual feature layer is extracted from remote sensing images, object categories and spatial object patterns are generated based on POIs, and for functional-zone boundaries, they are delineated by block polygons. Accordingly, remote sensing images, POIs, and polygon vectors can be totally integrated and expressed in the HSC frame.

HSC also has some limitations. First, the spatial object patterns in this study are represented by spatial point patterns and measured by G curves. This measurement ignores object shapes and sizes which however are also important for characterizing spatial object patterns (Vaduva et al., 2013). Second, although HSC itself is independent of object-based classification methods, HSC's inputs (object categories) are obtained based on object-classification methods, and thus the classification errors will further impact functional-zone-recognition results. Third, the spatial relations between functional zones are ignored in HSC, but they are important to recognize heterogeneous urban zones (Zhang et al., 2015). Consequently, these three issues need to be further studied.

6.4. HSC's extensibility: other usages

As demonstrated above, the HSC serves as a general cognition structure for urban functional zones, and we have provided one of its usages in Section 3, but it does have different forms and usages. In this section, we will show you the HSC's extensibility with respect to its default inputs.

HSC used a satellite image and POIs to classify functional zones, but it can still work if one kind of input data is unavailable. For the satellite image, it provides visual features, which are important for recognizing industrial zones and parks (Table 5). If the satellite image is missing, the $p(\vec{w}|s_t, c_k, z)$ and $p(\vec{w}|f_i, s_j, c_k)$ in HSC will be set as 1. As a result, the visual-feature layer and its relationship with object categories will be totally ignored. In this case, HSC's classification accuracy will decrease, but it still can work for generating classification results.

POIs data can provide object information for recognizing functional zones. If they are unavailable, the $p(s_t|z)$, $p(s_j|f_i)$, $p(c_k|s_t, z)$,

and $p(c_k|f_i, s_j)$ will be totally set as 1. Accordingly, both the object categories and spatial object patterns will be ignored, which can lead to awful classification results as the object information is critical to recognize most kinds of functional zones (Table 5). To resolve this issue, the missing POIs can be replaced by land-cover maps, and object information can be extracted from land-cover maps. Therefore, the HSC combining satellite images and land-cover maps are also able to classify functional zones.

Apart from recognizing functional zones, HSC can also be applied to other applications, such as object-based land cover mapping. Most land cover mapping methods are bottom-up processes which use visual features to distinguish object categories (Cleve et al., 2008; Laliberte et al., 2004; Peña-Barragán et al., 2011; Stow et al., 2008), but they ignore top-down prior information provided by spatial object patterns and zone functions. The top-down prior information is also important for object recognition, as Zhang and Du (2016) demonstrated that local information measured by spatial object patterns and zone functions can complement individual information for object classification. For example, the object (Fig. 14a) with such visual features can be a shadow or a water, but it should be the former one, considering its surrounding buildings and the local function, i.e., a residential district.

Accordingly, local information is useful but always ignored in most object classifications. To resolve this issue, Johnson and Xie (2013) employed super-object features as local information to improve object classifications, but they ignored the hierarchical relationships between local and individual features. Some other studies (Duchenne et al., 2011; Ri and Yao, 2015) established Bayesian network to model local relations between objects, and achieved some improvements. However, zone functions at the highest semantic layer are totally overlooked in existing object studies. HSC is a hierarchical Bayesian model, and can represent four semantic layers as well as their mapping relations. Accordingly, HSC can be used to classify objects, if other three layers have been obtained. Here, the parameter of $p(c_k|s_t, \vec{w}, z)$ is to be resolved, and other parameters can be learned from object training samples.

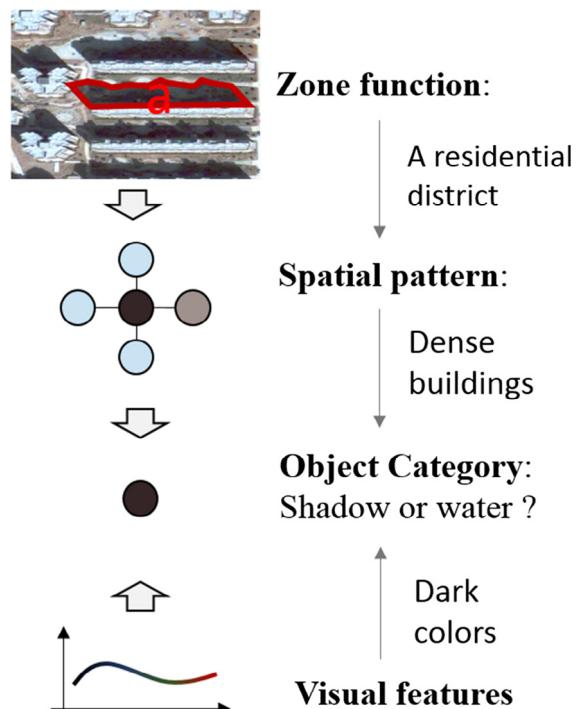


Fig. 14. Bottom-up and top-down processes for object recognition considering zone functions, spatial object patterns, and visual features.

HSC does not only consider the bottom-up relationship between visual features and object categories, but also employs the top-down prior information including zone functions and spatial object patterns. In theory, HSC can produce more accurate classification results for objects.

In summary, the proposed HSC has excellent extensibility for different conditions and applications. Its inputs can be easily replaced and its outputs can be freely designed.

7. Conclusions

In order to accurately map urban functional zones, this study presents a novel classification method, i.e., hierarchical semantic cognition (HSC). HSC essentially use a hierarchical structure instead of linear feature vectors to characterize and classify functional zones. To our best knowledge, the HSC is the first model which can integrate four semantic layers together and consider their hierarchical relations.

In experiment, HSC was used to classify urban functional zones in Beijing, and three conclusions have been drawn. First of all, HSC outperforms traditional classifiers, i.e., SVM and LDA, in zone categorization, with the overall accuracy improved by more than 7% (Section 5.3), and the hierarchical structure makes the difference (Section 6.2). Then, object categories are most important in HSC for zone classification with a greatest contribution degree of 54% (Section 6.1). In addition, visual features and spatial object patterns make 33% and 13% contributions to zone classification respectively. Third, Beijing in 2002 had a large population according to the large residential area, and it was located at a high-urbanization level with a few industrial zones and shanty towns.

Three limitations of HSC are pointed out in Section 6.3, and will be the focuses of our future researches. In addition, HSC's potential applications, e.g., object recognition, can be further studied.

Acknowledgments

The work presented in this paper is supported by the National Natural Science Foundation of China (No. 41471315). The comments from four anonymous reviewers are greatly appreciated.

References

- Baatz, M., Schape, A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. *Angew. Geogr. Informationsverarbeitung* 12, 12–23.
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Rem. Sens.* 58 (3), 239–258.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.
- Bosch, A., Zisserman, A., Muñoz, X., 2006. Scene classification via pLSA. In: European Conference on Computer Vision. Springer, Berlin Heidelberg, pp. 517–530.
- Boutell, M.R., Luo, J., Shen, X., Brown, C.M., 2004. Learning multi-label scene classification. *Pattern Recogn.* 37 (9), 1757–1771.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecol. Model.* 168 (3), 233–249.
- Cao, L., Fei-Fei, L., 2007. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. In: 2007 IEEE 11th International Conference on Computer Vision. IEEE, pp. 1–8.
- Cleve, C., Kelly, M., Kearns, F.R., Moritz, M., 2008. Classification of the wildland–urban interface: a comparison of pixel-and object-based classifications using high-resolution aerial photography. *Comput. Environ. Urban Syst.* 32, 317–326.
- Cohen, G., 2000. Hierarchical models in cognition: do they have psychological reality? *Eur. J. Cogn. Psychol.* 12 (1), 1–36.
- Diggle, P.J., Besag, J., Gleaves, J.T., 1976. Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, 659–667.
- Drăguț, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *Int. J. Geogr. Inform. Sci.* 24 (6), 859–871.
- Drăguț, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogr. Rem. Sens.* 88, 119–127.
- Du, S., Zhang, F., Zhang, X., 2015. Semantic classification of urban buildings combining VHR image and GIS data: an improved random forest approach. *ISPRS J. Photogr. Rem. Sens.* 105, 107–119.
- Duchenne, O., Joulin, A., Ponce, J., 2011. A graph-matching kernel for object categorization. In: 2011 International Conference on Computer Vision. IEEE, pp. 1792–1799.
- Farabet, C., Couprie, C., Najman, L., LeCun, Y., 2013. Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8), 1915–1929.
- Farahzadeh, E., Cham, T.J., Li, W., 2015. Semantic and spatial content fusion for scene recognition. In: New Development in Robot Vision. Springer, Berlin Heidelberg, pp. 33–53.
- Fu, B., Liang, D., Nan, L., 2011. Landscape ecology: coupling of pattern, process, and scale. *Chin. Geogr. Sci.* 21 (4), 385–391.
- Gahegan, M., Ehlers, M., 2000. A framework for the modelling of uncertainty between remote sensing and geographic information systems. *ISPRS J. Photogr. Rem. Sens.* 55 (3), 176–188.
- Gatrell, A.C., Bailey, T.C., Diggle, P.J., Rowlingson, B.S., 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Trans. Inst. Brit. Geogr.*, 256–274.
- Haase, P., 1995. Spatial pattern analysis in ecology based on Ripley's K-function: introduction and methods of edge correction. *J. Veg. Sci.* 6 (4), 575–582.
- Heiden, U., Heldens, W., Roessner, S., Segl, K., Esch, T., Mueller, A., 2012. Urban structure type characterization using hyperspectral remote sensing and height information. *Landscape Urban Planning* 105 (4), 361–375.
- Hu, T., Yang, J., Li, X., Gong, P., 2016. Mapping urban land use by using Landsat images and open social data. *Rem. Sens.* 8 (2), 151.
- Indyk, P., Motwani, R., 1998. Approximate nearest neighbors: towards removing the curse of dimensionality. In: Thirtieth ACM Symposium on Theory of Computing. ACM.
- Johnson, B., Xie, Z., 2013. Classifying a high resolution image of an urban area using super-object information. *ISPRS J. Photogramm. Rem. Sens.* 83, 40–49.
- Kontschieder, P., Bulo, S.R., Pelillo, M., Bischof, H., 2014. Structured labels in random forests for semantic labelling and object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (10), 2104–2116.
- Laliberte, A.S., Rango, A., Havstad, K.M., Paris, J.F., Beck, R.F., McNeely, R., Gonzalez, A.L., 2004. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico. *Remote Sens. Environ.* 93, 198–210.
- Li, L.J., Su, H., Xing, E.P., Fei-Fei, L., 2010. Object bank: a high-level image representation for scene classification & semantic feature sparsification. In: Advances in Neural Information Processing Systems, pp. 1378–1386.
- Lillesand, T., Kiefer, R.W., Chipman, J., 2014. *Remote Sensing and Image Interpretation*. John Wiley & Sons.
- Lienou, M., Maître, H., Datcu, M., 2010. Semantic annotation of satellite images using latent dirichlet allocation. *IEEE Geosci. Remote Sens. Lett.* 7 (1), 28–32.
- Matsuoka, R.H., Kaplan, R., 2008. People needs in the urban landscape: analysis of landscape and urban planning contributions. *Landscape Urban Planning* 84 (1), 7–19.
- McKenzie, G., Janowicz, K., Gao, S., Yang, J.A., Hu, Y., 2014. POI pulse: a multi-granular, semantic signatures-based information observatory for the interactive visualization of big geosocial data. *Cartographica* 50 (2), 71–85.
- Memarsadeghi, N., Mount, D.M., Netanyahu, N.S., Le Moigne, J., 2007. A fast implementation of the ISODATA clustering algorithm. *Int. J. Comput. Geom. Appl.* 17 (01), 71–103.
- Montanges, A.P., Moser, G., Taubenböck, H., Wurm, M., Tuia, D., 2015. Classification of urban structural types with multisource data and structured models. In: 2015 Joint Urban Remote Sensing Event (JURSE). IEEE, pp. 1–4.
- Neis, P., Zipf, A., 2012. Analyzing the contributor activity of a volunteered geographic information project—the case of OpenStreetMap. *ISPRS Int. J. Geo-Inform.* 1 (2), 146–165.
- Oquab, M., Bottou, L., Laptev, I., Sivic, J., 2014. Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1717–1724.
- Payne, A., Singh, S., 2005. Indoor vs. outdoor scene classification in digital photographs. *Pattern Recogn.* 38 (10), 1533–1545.
- Peña-Barragán, J.M., Ngugi, M.K., Plant, R.E., Six, J., 2011. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sens. Environ.* 115 (6), 1301–1316.
- Pertuz, S., Garcia, M.A., Puig, D., 2015. Focus-aided scene segmentation. *Comput. Vis. Image Underst.* 133, 66–75.
- Rahmani, S., Strait, M., Merkurjev, D., Moeller, M., Wittman, T., 2010. An adaptive IHS pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.* 7 (4), 746–750.
- Rhinane, H., Hilali, A., Berrada, A., Hakdaoui, M., 2011. Detecting slums from SPOT data in Casablanca Morocco using an object based approach. *J. Geogr. Inform. Syst.* 3 (03), 217.
- Ri, C.Y., Yao, M., 2015. Bayesian network based semantic image classification with attributed relational graph. *Multimedia Tools Appl.* 74 (13), 4965–4986.
- Serrano, N., Savakis, A., Luo, J., 2002. A computationally efficient approach to indoor/outdoor scene classification. *Pattern Recognition*, 2002. Proceedings. 16th International Conference on, vol. 4. IEEE, pp. 146–149.
- Serrano, N., Savakis, A.E., Luo, J., 2004. Improved scene classification using efficient low-level features and semantic cues. *Pattern Recogn.* 37 (9), 1773–1784.
- Shin, H.B., 2009. Residential redevelopment and the entrepreneurial local state: the implications of Beijing's shifting emphasis on urban redevelopment policies. *Urban Stud.* 46 (13), 2815–2839.

- Song, W., Wen, D., Wang, K., Liu, T., Zang, M., 2015. Satellite image scene classification using spatial information. In: Sixth International Conference on Graphic and Image Processing (ICGIP 2014). International Society for Optics and Photonics, pp. 94431K–94431K.
- Steeves, J.K., Humphrey, G.K., Culham, J.C., Menon, R.S., Milner, A.D., Goodale, M.A., 2004. Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J. Cogn. Neurosci.* 16 (6), 955–965.
- Stow, D., Hamada, Y., Coulter, L., Anguelova, Z., 2008. Monitoring shrubland habitat changes through object-based change identification with airborne multispectral imagery. *Remote Sens. Environ.* 112, 1051–1061.
- Suykens, J.A., Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural Process. Lett.* 9 (3), 293–300.
- Vaduva, C., Gavat, I., Datcu, M., 2013. Latent Dirichlet allocation for spatial analysis of satellite images. *IEEE Trans. Geosci. Remote Sens.* 51 (5), 2770–2786.
- Wang, J., Chen, Y., He, T., Lv, C., Liu, A., 2010. Application of geographic semantic cognition approach in land type classification using Hyperion image: a case study in China. *Int. J. Appl. Earth Obs. Geoinf.* 12, S212–S222.
- Wang, X., Grimson, E., 2008. Spatial latent dirichlet allocation. In: Advances in Neural Information Processing Systems, pp. 1577–1584.
- Waseem, M., Muhammad, A.N., Khalida, I.N., 2013. Decomposition method for solving system of linear equations. *Eng. Math. Lett.* 2, 34–41.
- 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 (1), 609–625.
- Yan, R., Liu, Y., Jin, R., Hauptmann, A., 2003. On predicting rare classes with SVM ensembles in scene classification. *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International Conference on*, vol. 3. IEEE, pp. III–21.
- Yang, G., Pu, R., Zhang, J., Zhao, C., Feng, H., Wang, J., 2013. Remote sensing of seasonal variability of fractional vegetation cover and its object-based spatial pattern analysis over mountain areas. *ISPRS J. Photogramm. Rem. Sens.* 77 (3), 79–93.
- Yao, J., Fidler, S., Urtasun, R., 2012. Describing the scene as a whole: joint object detection, scene classification and semantic segmentation. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, pp. 702–709.
- Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2016. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inform. Sci.*, 1–24.
- Zhang, X., Du, S., 2015. A linear Dirichlet mixture model for decomposing scenes: application to analyzing urban functional zonings. *Rem. Sens. Environ.* 169, 37–49.
- Zhang, X., Du, S., 2016. Learning selfhood scales for urban land cover mapping with very-high-resolution VHR images. *Rem. Sens. Environ.* 178, 172–190.
- Zhang, X., Du, S., Wang, Y.-C., 2015. Semantic classification of heterogeneous urban scenes using intrascene feature similarity and interscene semantic dependency. *IEEE J. Sel. Top. Appl. Earth Observations Rem. Sens.* 8, 2005–2014.
- Zhao, P., Lü, B., 2009. Transportation implications of metropolitan spatial planning in mega-city Beijing. *Int. Dev. Plan. Rev.* 31 (3), 235–261.
- Zhao, W., Du, S., 2016. Scene classification using multi-scale deeply described visual words. *Int. J. Rem. Sens.* 37 (17), 4119–4131.
- Zheng, X., Wang, Y., Gan, M., Zhang, J., Teng, L., Wang, K., Shen, Z., Zhang, L., 2016. Discrimination of settlement and industrial area using landscape metrics in rural region. *Rem. Sens.* 8 (10), 845.
- Zhong, Y., Cui, M., Zhu, Q., Zhang, L., 2015. Scene classification based on multifeature probabilistic latent semantic analysis for high spatial resolution remote sensing images. *J. Appl. Rem. Sens.* 9 (1), 6207–6222.
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems, pp. 487–495.