Computers, Environment and Urban Systems 78 (2019) 101374

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Functional urban land use recognition integrating multi-source geospatial data and cross-correlations

Yatao Zhang[a], Qingquan Li[a,b,d], Wei Tu[b,c,d,*], Ke Mai[a], Yao Yao[e], Yiyong Chen[c]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[b] Guandgong Key Laboratory of Urban Informatics, Shenzhen Key Laboratory of Spatial Smart Sensing and Services and Research Institute of Smart Cities, Shenzhen University, Shenzhen 518060, China
[c] Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China
[d] Key Laboratory for Geo-Environment Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation, Shenzhen University, Shenzhen 518060, China
[e] School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, Hubei, China

ARTICLE INFO

Keywords:
Cross-correlations
CC-FLU model
Functional urban land use
Multi-source geospatial data

ABSTRACT

The significance of urban function recognition has stimulated the need for multi-source geospatial data fusion, especially the fusion between remote sensing images and spatiotemporal big data. In previous studies, the natural correspondence across multi-source geospatial data has often been ignored in the description of one object, which would influence the performance of data fusion. Therefore, this study introduces the cross-correlations mechanism to achieve the natural correspondence by taking remote sensing images, point of interest (POI), and real-time social media users as an example. It proposes a new cross-correlations based functional urban land use (CC-FLU) model to infer urban functions. The presented model extracts physical and human semantic features from multi-source geospatial data, then maps them to their common subspaces to obtain their cross-correlations respectively. These semantic features and their cross-correlations are integrated together to classify urban functions. An experiment in Shenzhen, China was implemented to evaluate the performance of the presented model at a fine scale. The results show that the proposed CC-FLU model achieved a better performance than previous methods, yielding OA and Kappa values of 0.851 and 0.812, respectively. The results of the presented approach outperform those of methods using one single source geospatial data. The results demonstrate that the skilled information of each type of geospatial data is fully melted into the data fusion model, and simultaneously achieves the natural correspondence across multi-source geospatial data. Moreover, this study resolves the possible disadvantages of models using one single source data and fusion methods by sequentially concatenating multi-source features. The results will benefit urban planners and urban policy-makers.

1. Introduction

Urbanization has inevitably continued to increase around the world; in 2018, 55% of the global population resided in urban areas, and this urbanization has greatly impacted humans' living environments and urban land use (Hsieh, 2014; McDonald, Marcotullio, & G U Neralp, 2013; Sapena & Ruiz, 2018; *United Nations Secretariat,* 2018). As the most populous country in the world, China has experienced unprecedented urbanization in a very short period of time (Guan et al., 2018). During this process, the spatial layout of urban functions has become a vital issue that is intensively related to urban vitality, traffic, and transportation (Abdullahi et al., 2015; Dovey & Pafka, 2017). Hence, it is essential to effectively and efficiently collect functional urban land use information.

High spatial resolution (HSR) remote sensing images are an important data source and tool for classifying urban land use, which can provide the physical information of geographical objects, and are widely used in object-based scene classification models (Forestier et al., 2012; Shahriari & Bergevin, 2017). Some features, such as spectral and textural features, have constituted the footstone of urban land use recognition (Blaschke et al., 2014; Hu & Wang, 2013). However, these basic features are mostly used to depict one land patch based on its physical properties, thus making it difficult to distinguish objects with the same physical attributes but different functional attributes in high-

* Corresponding author at: Room 1402, Science and Technology Building, Shenzhen University, No. 3688 Nanhai Avenue, Nanshan, Shenzhen 518060, Guangdong, China.
E-mail addresses: tuwei@szu.edu.cn, tuwei@mit.edu (W. Tu).

https://doi.org/10.1016/j.compenvurbsys.2019.101374
Received 20 September 2018; Received in revised form 19 June 2019; Accepted 24 July 2019
0198-9715/

density cities (Cao et al., 2018; Liu et al., 2017; Ríos & Muñoz, 2017), such as Shenzhen, London, and New York. For example, grid-layout buildings with rectangular boundaries in industrial parks are classified as working functions, which are different from residential buildings with similar spectral and texture features in commercial areas. In fact, functional urban land use is not only associated with physical feature spaces but is also influenced by human feature spaces.

Fusing HSR images with new data sources is an alternative approach to recognizing functional urban land use in high-density cities. With the emergence of new urban data, e.g., point of interest (POI) data, smart card data, taxi trajectory data, mobile phone data, social media data, street view images, etc., many new approaches have been developed to understand urban systems (Liu et al., 2015; Chen et al., 2017; Ríos & Muñoz, 2017; Xu, Belyi, Bojic, & Ratti, 2018; Zhang et al., 2017; Zhong et al., 2014). These urban data can reflect the temporal changes and spatial patterns of human flows, material flows and information flows (Liu et al., 2015), thus are widely used to carry out urban function recognition; for example, Pei et al. (2014) uncovered the urban land use using time-series mobile phone user data; Tu et al. (2017) coupled mobile phone data and social media data to infer urban functions and uncovered their hourly dynamics; Xing and Meng (2018) integrated the landscape metrics and socioeconomic features extracted from crowd-sourced data to recognize urban functions. Above studies suggest that each type of geospatial data has its own merits, and a promising solution to functional urban land use recognition is to fuse them together.

However, each type of geospatial data also has its own demerits, specifically that the information obtained from a single data source may be sparse, uncertain and incomplete and have varying noise (Baltru, Aitis, Ahuja, & Morency, 2019; Tu et al., 2018b; Wu et al., 2014; Wang et al., 2015; Zhang, Xu, Tu, & Ratti, 2018). This has raised doubts about whether the natural correspondence can be explicitly guaranteed if one object is simultaneously depicted by multiple features from multi-source geospatial data (Pereira et al., 2014; Rasiwasia et al., 2010; Zheng, 2015). Here, the natural correspondence refers to the state that the multiple descriptions of one object refined from multi-source data can achieve the consistency, which makes sure that the information provided by each data source can be fully utilized in the data fusion process (Rasiwasia et al., 2010; Zheng, 2015). Compared with previous studies without the consideration of the natural correspondence, it can help to abstract the most important information across multiple feature spaces and eliminate possible noises, and then improve the performance of the models. The scenario involving remote sensing images and human activity data is given as an example. Considering the ubiquitous phenomenon of the same object with different spectrums occurring in remote sensing data (Liu et al., 2017) and the influence of the emergency events in human activity data (Zheng et al., 2016), the two obtained information from them may not be accurate. In this case, the natural correspondence demands for a state of consistency between them, which can exclude the noises caused by their demerits and take full use of their merits, and then increase the recognition accuracy.

However, most previous studies sequentially embedded the features extracted from different sources in the same spatial units (Liu et al., 2017; Tu et al., 2017) but often ignored the natural correspondence across them. Thus, a following question is how to establish this correspondence. One solution is to utilize the subspace learning method which links multiple feature spaces to achieve the goal, such as canonical correlation analysis (CCA) (Hardoon, Szedmak, & Shawe-Taylor, 2004; Rasiwasia et al., 2010; Zheng, 2015). CCA aims to refine the original feature spaces into the intermediate isomorphic feature spaces, then maximize the correlations across them (Baltru et al., 2019; Pereira et al., 2014). In this way, the demerits of each geospatial data are decreased through their correlations and their own merits are fully utilized through the combinations. Here, this study uses the cross-correlations to depict this refining process due to its relationship with the correlations across multi-feature spaces. Thus, a functional land use recognition framework combining the cross-correlations (CC-FLU) is established based on this proposed multi-source geospatial data fusion model. In detail, physical features from HSR images and human features from POIs and social media users are extracted separately. Then, the semantics mined by the probabilistic topic model (PTM) and their cross-correlations mined by CCA are integrated to unearth urban functional land use through a random forest (RF) algorithm. In general, this proposed CC-FLU model takes full use of the skilled information in multiple feature spaces and resolves the possible issue of bias in a single data source and the demerits of previous data fusion models through cross-correlations. An experiment in Shenzhen, China illustrates the advantages of our CC-FLU model.

The rest of this paper is organized as follows. The study area and data sets in this study are described in Section 2. Section 3 provides an explicit description of the framework of the presented CC-FLU model, which consists of semantic feature construction, cross-correlation learning, and functional urban land use recognition. The experimental results and discussion are reported in Sections 4 and 5, respectively. Finally, Section 6 concludes the study.

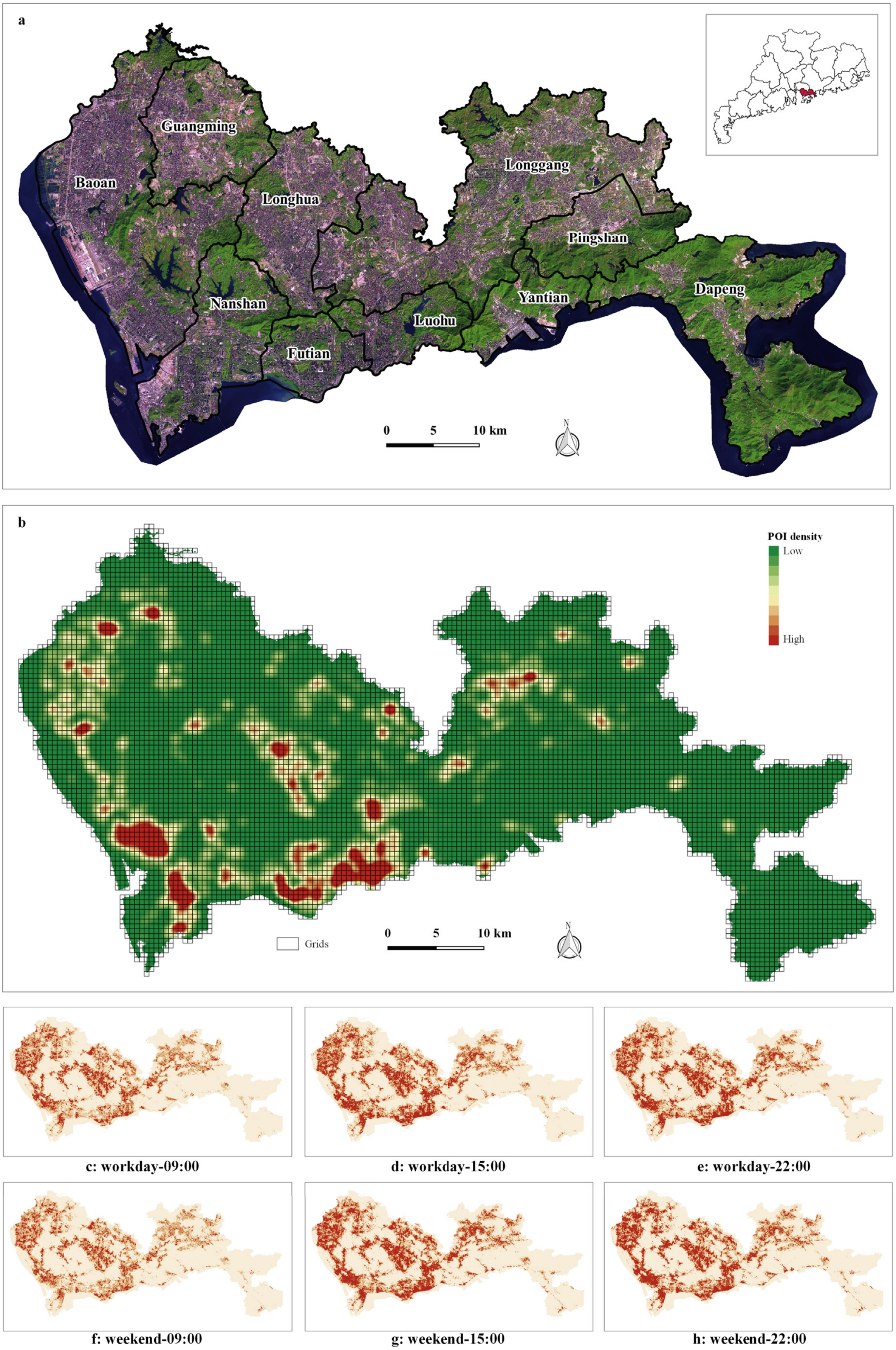
2. Study area and datasets

This study was conducted in Shenzhen, China, which is located to the north of the Pearl River Delta. As the first special economic zone of China, Shenzhen has experienced unprecedented urbanization in the past forty years and has become one of the largest global metropolises, with a population of more than 10 million people (*United Nations Secretariat,* 2018). Monitoring the functional land use of this global city is essential to achieving its sustainable development. Fig. 1(a) displays the location and boundary of Shenzhen. To provide a fine-grained urban functional map, rather than using cadastral plots or traffic analysis zones, we divided Shenzhen into 500-m grids according to the previous studies (Guan & Rowe, 2016; Tu et al., 2018a), with a total number of 8295 grids, defined as 8295 land patches, as shown in Fig. 1(b).

High spatial resolution (HSR) remote sensing images, point of interest (POI) data and real-time Tencent users (RTU) were utilized to implement the presented CC-FLU model, which will be introduced in Section 3.

- The HSR image in Fig. 1(a) was obtained using the pan-sharpening fusion method based on the multispectral and panchromatic SPOT-5 images collected on 30 November 2013. This image has four spectral images and 37,368 × 19,440 pixels, with a spatial resolution of 2.5 m per pixel (Tu et al., 2018a).
- A POI refers to a geographical point with a property label and the position. The POI data used in this study is obtained from Baidu map, the biggest map service provider in China, which has been widely used in urban studies (Yao et al., 2016). For all the POI data in 2016, several types with great benefits on functional urban land use recognition were selected, with a total number of 156,303, including residential communities, commercial sites, industrial facilities, entertainment facilities, medical facilities, landscape sites, and education facilities. Fig. 1(b) represents a kernel density raster map of all of the POIs used here.
- The RTU is a raster dataset with a spatial resolution of 25 m collected in 2016 by Tencent, which is one of the largest internet companies in the world. The RTU data measure the hourly numbers of phone users who use Tencent applications, such as QQ and WeChat. Fig. 1(c–h) show the RTU data obtained in 3 typical time intervals on a workday and weekend.

There are significant differences in the spatial distributions of POIs and RTUs among grids, which implies that these urban land patches have different functions. Note that due to the limitations of data availability, the time of HSR images is different from the other two



(caption on next page)

Fig. 1. Study area and the used datasets. (a) Study area and HSR image. (b) The kernel density of POIs. (c) RTU density at 09:00 on a workday. (d) RTU density at 15:00 on a workday. (e) RTU density at 22:00 on a workday. (f) RTU density at 09:00 on a weekend. (g) RTU density at 15:00 on a weekend. (h) RTU density at 22:00 on a weekend. RTU: real-time Tencent users.

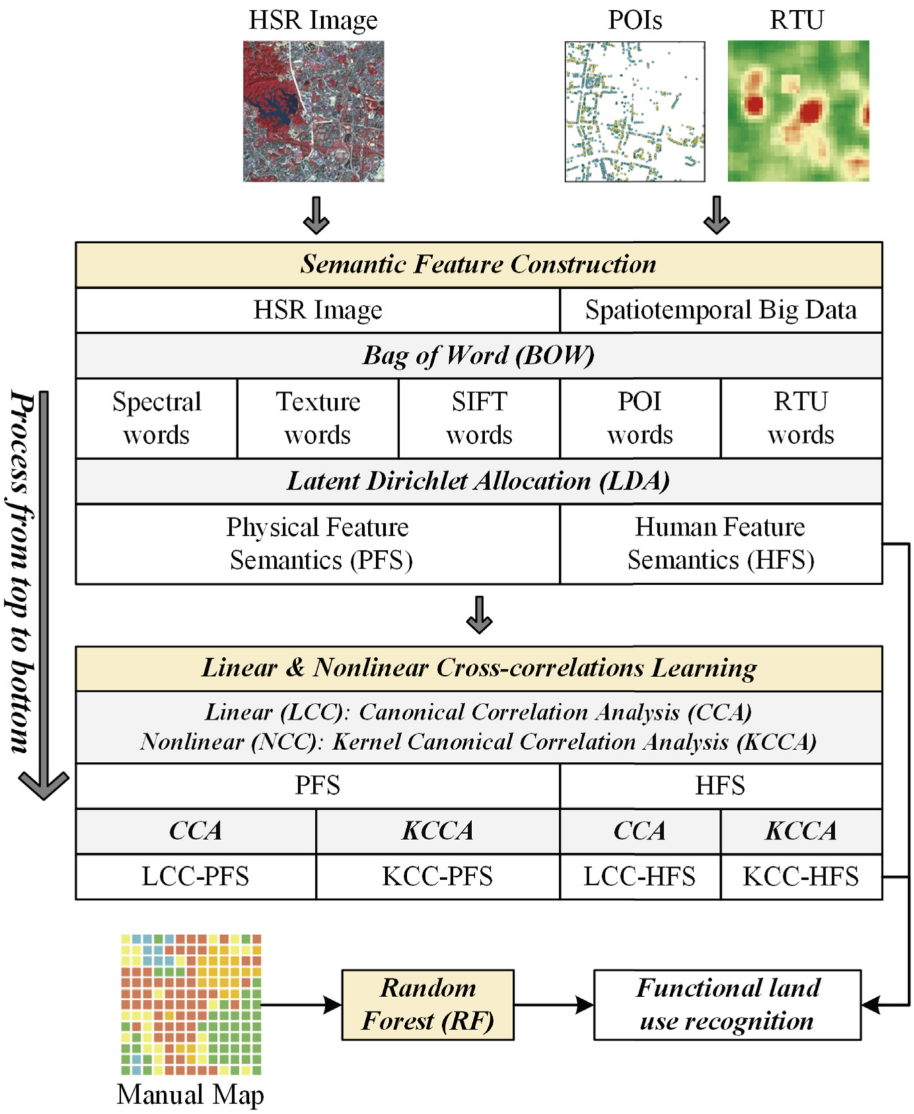


Fig. 2. Framework of the CC-FLU model.

datasets, which may have an effect on this study. But Shenzhen experienced a fast urban expansion from 1978 to 2010; after 2010, the urbanization in Shenzhen stepped into a relatively stable stage (Fei & Zhao, 2019). Thus, in the short interval from 2013 to 2016, features extracted from the HSR image changed little, and do not significantly affect the evaluation of the presented CC-FLU model.

3. Functional land use recognition through cross-correlations

The main goal of the CC-FLU model presented here is to combine the semantic information extracted from multi-source geospatial data and build their natural correspondence to classify functional urban land use. Fig. 2 presents the framework of the CC-FLU model. This model consists of three parts: semantic feature construction, cross-correlation learning, and functional urban land use recognition. In the first part, multi-type features are extracted into the bag-of-words (BOW) format (Csurka et al., 2004); then the latent Dirichlet allocation (LDA) model (Blei, Ng, & Jordan, 2003) is used to mine the semantic topics of each-source geospatial data. In the second part, canonical correlation analysis (CCA) and kernel CCA (KCCA) are utilized to determine the linear and nonlinear cross-correlations (Hardoon et al., 2004) across multi-semantic spaces, respectively. In the third part, the random forest (RF) algorithm is employed to classify functional land use based on semantic features and their cross-correlations.

3.1. Semantic feature construction via probability topic model

Feature construction plays a vital role in object-based recognition (Lillywhite et al., 2013). Because the BOW model has received much attention in object-based scene classification, probabilistic topic models, especially LDA, have been widely applied in related studies (Zhong, Zhu, & Zhang, 2015; Zhou, Zhou, & Hu, 2013). Here, the spatial features extracted from multi-source geospatial data are generated into BOW dictionaries, and they are then inputted to the LDA model to unearth the hidden semantic topics of land patches.

Spatial features, including spectral features, texture features, and scale-invariant-feature-transform (SIFT) features, are widely applied in land use recognition (Wu, Zhang, & Zhang, 2016; Zhang & Du, 2015;

Zhong et al., 2015). Given an HSR image with l bands, the spatial features are extracted within a moving window, in which the size of the window is set as 25×25 pixels, with 15 overlapping pixels, based on previous studies (Liu et al., 2017). Each spectral descriptor is defined as the mean and standard deviation (*std*) of the spectral features in one moving window (Lienou, Maitre, & Datcu, 2010), and it can be expressed as:

$$\text{spectral features} = \{mean_1, std_1, ..., mean_l, std_l\}.$$

The gray-level co-occurrence matrix (GLCM) (Mohanaiah, Sathyanarayana, & GuruKumar, 2013) is used to depict the texture feature. Four Haralick's feature statistics of GLCM in a moving window are selected to determine the texture descriptor, including contrast, energy, correlation and homogeneity, and the texture descriptor can be expressed as:

$$\text{texture features} = \{con_1, ene_1, cor_1, hom_1, ..., con_l, ene_l, cor_l, hom_l\}.$$

In terms of the SIFT descriptor, a 128-dimensional vector is extracted in each band, which is expressed as $Sift_i = \{sift_1, \ldots, sift_{128}\}$. The complete SIFT descriptor consists of the SIFT points in all bands and can be expressed as:

$$SIFT = \{Sift_1, ..., sift_l\}.$$

In all of these spatial descriptors, the clustering method is necessary to generate their visual words in BOW dictionaries. In this study, the K-Means method is used to group these spatial descriptors into different clustering numbers; then, Davies-Bouldin index (DBI) (Davies & Bouldin, 1979) is used to score the clustering results based on their various numbers. The lower the DBI value is, the better the clustering result is. The clustering centers of the best clustering result are stored as visual words in the BOW dictionary for their spectral, texture and SIFT features.

The visual words for the POIs and RTU can be determined directly. The POI visual words are equal to the categories of POIs. The RTU words are defined as the users' density by hours, noting that the hours on workdays and weekends should be counted separately due to the different human patterns on workdays and weekends. Finally, five types of visual words in BOW dictionaries are defined, including three types of physical feature words, namely, spectral, texture and SIFT words, and two types of human-related words, namely, POI and RTU words.

The LDA model has been successfully used to extract hidden semantic topics. Assuming that all land patches are the documents of a corpus, the functions in each land patch are regarded as a mixture of latent abstract topics, and the visual words in the BOW dictionary are defined as the words of a vocabulary. Then, the LDA model provides explicit representations of documents through the mixture of latent topics, and each topic is characterized by a distribution of words (Blei et al., 2003). In each grid, we can obtain five different LDA descriptions from the spectral, texture, SIFT, POI and RTU features. For each LDA description, the documents are generated by counting the distribution frequency of visual words, and they are input to the LDA model to unearth the latent topics in each land patch. The details of using the LDA model in land use applications have been discussed in previous studies (Blei et al., 2003; Liu et al., 2017; Yuan, Zheng, & Xie, 2012; Zhong et al., 2015). Finally, this procedure outputs five types of semantic descriptions, including three types of physical feature semantics (PFS), namely, spectral, texture and SIFT semantics, and two types of human feature semantics (HFS), namely, POI and RTU semantics, which correspond to five types of spatial features. Each type of LDA description has its own feature space due to its various feature sources.

3.2. Cross-correlation learning via subspace learning methods

As one single data source has its own merits and demerits, the purpose of multi-source geospatial data fusion is to emphasize the merits of a single data source and decrease its demerits, and then achieve the state of natural correspondence across multi-data sources. One feasible method is to draw the cross-correlations across multiple feature spaces into a data fusion model (Pereira et al., 2014; Rasiwasia et al., 2010; Zheng, 2015). Here, the cross-correlations of the PFSs and HFSs are combined with their semantics to recognize urban functional land use, which constitutes the kernel of the presented CC-FLU model.

Assume that all land patches in the study area are expressed as $\mathscr{D} = \{\mathscr{D}_1, ..., \mathscr{D}_{|\mathscr{D}|}\}$, where each element $\mathscr{D}_i$ represents the set of all descriptions in the i-th land patch, and comprises more than one description. For simplicity, we consider the case in which one land patch consists of two types of semantic descriptions, i.e., $\mathscr{D}_i = (X_i, Y_i)$, where they are represented as vectors in the feature spaces $\mathscr{R}^X$ and $\mathscr{R}^Y$, respectively. Because X_i and Y_i co-occur to describe the same object but in different feature spaces, one efficient solution to decreasing the influence of their own demerits is to map them into two intermediate isomorphic spaces respectively ($\mathscr{U}^X$, $\mathscr{U}^Y$) by subspace learning (Rasiwasia et al., 2010; Wei et al., 2017; Zheng, 2015). This process contains two mappings:

$$\mathscr{M}_X: \mathscr{R}^X \to \mathscr{U}^X,\ \mathscr{M}_Y: \mathscr{R}^Y \to \mathscr{U}^Y \tag{1}$$

This solution demands a joint dimensionality reduction model that takes effects on both two different feature spaces ($\mathscr{R}^X$, $\mathscr{R}^Y$), meanwhile builds the high correlations between the mapping isomorphic spaces ($\mathscr{U}^X$, $\mathscr{U}^Y$), which can be satisfied by canonical correlation analysis (CCA) (Hardoon et al., 2004). CCA is a subspace learning method, which maps the two original feature spaces into two maximally-correlated isomorphic spaces, and makes sure that the distance between the two refined descriptions of one object can be minimized and that their correlations can be maximized (Hardoon et al., 2004; Wei et al., 2017). Then, the demerits or noises in each type of geospatial data can be excluded to some extent.

Given two directions $\omega_x \in \mathscr{R}^X$ and $\omega_y \in \mathscr{R}^Y$, then CCA seeks the directions along which they are maximally correlated,

$$\max_{\omega_x \neq 0, \omega_y \neq 0} \frac{{\omega_x}' \Sigma_{XY} \omega_y}{\sqrt{{\omega_x}' \Sigma_{XX} \omega_x} \sqrt{{\omega_y}' \Sigma_{YY} \omega_y}} \tag{2}$$

where Σ_{XX} and Σ_{YY} are the covariance matrices for the descriptions $\{X_1, ..., X_{|\mathscr{D}|}\}$ and $\{Y_1, ..., Y_{|\mathscr{D}|}\}$, respectively, and $\Sigma_{XY} = {\Sigma_{XY}}'$ is the cross-covariance matrix between them. The first d canonical components $\{\omega_{x,k}\}_{k=1}^d$ and $\{\omega_{y,k}\}_{k=1}^d$ become the basis for mapping $\mathscr{R}^X$ and $\mathscr{R}^Y$ on the subspaces $\mathscr{U}^X$ and $\mathscr{U}^Y$, respectively (Pereira et al., 2014). For the two original semantic descriptions in $\mathscr{D}_i = (X_i, Y_i)$, X_i in $\mathscr{R}^X$ space is mapped into the P_x in $\mathscr{U}^X$ space based on $\{\omega_{x,k}\}_{k=1}^d$, and Y_i in $\mathscr{R}^Y$ space is mapped into the P_y in $\mathscr{U}^y$ space based on $\{\omega_{y,k}\}_{k=1}^d$. Given a space $\mathscr{U}$ overlapped by $\mathscr{U}^X$ and $\mathscr{U}^y$, then P_x and P_y can be regarded as two coordinates in this overlapped space (Rasiwasia et al., 2010), which represent the linear cross-correlations.

Kernel canonical correlation analysis (KCCA) is a nonlinear extension of CCA that maximally correlates nonlinear projections and is restricted to reproducing kernel Hilbert spaces with certain kernels (Andrew et al., 2013; Pereira et al., 2014), with two nonlinear mappings into high-dimensional spaces:

$$\phi_X: \mathscr{R}^X \to \mathscr{F}^X,\ \phi_Y: \mathscr{R}^Y \to \mathscr{F}^Y \tag{3}$$

These two transformations ϕ_X and ϕ_Y are implemented by two kernel functions, i.e., $\mathscr{K}_X(.,.)$ and $\mathscr{K}_Y(.,.)$, and they are expressed as

$$\mathscr{K}_X(X_i, X_j) = \langle \phi_X(X_i), \phi_X(X_j) \rangle,\ \mathscr{K}_Y(Y_i, Y_j) = \langle \phi_Y(Y_i), \phi_Y(Y_j) \rangle, \tag{4}$$

in the format of inner products respectively. In detail, KCCA seeks the directions $f_x \in \mathscr{F}^X$ and $f_y \in \mathscr{F}^Y$, along which they are maximally correlated in nonlinear common spaces:

$$\max_{\alpha_x \neq 0, \alpha_y \neq 0} \frac{{\alpha_x}' \mathscr{K}_X \mathscr{K}_Y \alpha_y}{V(\alpha_x, \mathscr{K}_X) V(\alpha_y, \mathscr{K}_Y)}, \tag{5}$$

where $V(\alpha, \mathscr{K}) = \sqrt{(1 - \kappa)\alpha' \mathscr{K}^2 \alpha + \kappa \alpha' \mathscr{K} \alpha}$ is a regularization parameter, $\ell \in [0,1]$, and $\mathcal{K}_X$ and $\mathcal{K}_Y$ are the kernel matrices of the representations in X and Y, respectively. Then, f_x and f_y can be obtained through the weight vectors α_x and α_y. The first d canonical components $\{f_{x,\ k}\}_{k=1}{}^d$ and $\{f_{y,\ k}\}_{k=1}{}^d$ become the basis for mapping $\mathscr{R}^X$ and $\mathscr{R}^Y$ on the nonlinear subspaces $\mathscr{F}^X$ and $\mathscr{F}^Y$, respectively (Hardoon et al., 2004; Pereira et al., 2014). The nonlinear cross-correlations can be obtained with the same way as that is introduced in CCA.

3.3. Functional urban land use recognition

The goal of this step is to combine the outputs of the first two parts together to classify the functional land use. In the first part, two types of semantic features, namely, PFS and HFS, are outputted respectively. In previous studies, the PFS was widely used in scene classification models to recognize land use based on remote sensing images; the HFS constituted the basis of the model of using human activity semantics to sense urban functions based on spatiotemporal big data; their combination formed another typical model that fuses PFS and HFS together. These three models are vital parts of our CC-FLU model and are also good examples with which to test our CC-FLU model. In the second part, CCA and KCCA (Hardoon et al., 2004) are separately utilized to output the linear and nonlinear cross-correlations across multiple semantic feature spaces. Then, the semantic information and their cross-correlations make up the presented CC-FLU model. In this way, the aforementioned disadvantages of a single data source and previous data fusion models can be improved.

In addition, one classifier is needed to implement the classification task. The random forest (RF) algorithm is chosen due to its outstanding performance, which performs better than other 178 classifiers, as determined by performing a number of trials on 121 datasets (Fern A Ndez-Delgado et al., 2014). The performance of the model is estimated based on its overall accuracy (OA) and Kappa coefficient. Moreover, RF has the ability to evaluate the importance of features (Deng & Runger, 2013), which is used to rank and select the features in this study.

4. Result

4.1. Experiment settings

The presented CC-FLU model was implemented using C++ on Windows 7 (x64). Several open-source C/C++ libraries were employed to process multi-source geospatial data and carry out machine learning methods, including GDAL (http://www.gdal.org/), OpenCV (https://opencv.org/) and Shark (http://image.diku.dk/shark/). The codes of LDA are from the widely used GibbsLDA++ model (http://gibbslda.sourceforge.net/).

To examine the CC-FLU model, we designed nine different examples to classify functional land use. The details of these nine examples are shown in Fig. 3. In the first three examples (***A***, ***B*** and ***C***), the semantic features from the HSR images (PFS), POI and RTU (HFS), and their combination are utilized to carry out functional land use classification tasks. In the next two examples (***D*** and ***E***), the linear (LCC-PFS) and nonlinear (KCC-PFS) cross-correlations for PFS are the inputs, and the one that exhibits a better performance will be used in case ***H***. Similarly, ***F*** and ***G*** are executed with the linear (LCC-HFS) and nonlinear (KCC-HFS) cross-correlations for HFS, respectively, and the better one will be used in case ***H***. Then, our CC-FLU model (***I***) is carried out through the combination of case ***C*** and ***H***. To avoid the possible influence of the curse of dimensionality and abundant redundant features, the importance of features is first ranked by the RF algorithm, and then the top fifty features are selected as the input of ***I***. In this way, the effects of cross-correlations on the recognition task can be further showed.

During the implementation of these trials, 465 land patches were randomly selected from 8295 land patches to label their functional land use categories, including residential land (RES), industrial land (IND), commercial land (COM), public management and service land (PUB), green and forest land (GEN), and water body (WAT). In the labelling process, remote sensing images, street view images, and urban planning land data were simultaneously used to distinguish each grid's ground truth via manual interpretation. Then, if one grid involves more than two land use types, the land use covering the largest area is identified as this grid's label. Importantly, mixed land use is a complicated and important issue in urban land use recognition (Zhang & Du, 2015), but the gist of this paper is to introduce a geospatial data fusion model involving the natural correspondence across multiple data sources. Thus, a simple method determining its actual land use with the largest area is adopted here. To evaluate the accuracy of the results obtained in different examples (***A*** to ***I***), the RF algorithm was run 100 times. In each run, all labeled samples were randomly divided into training and validation datasets, and the former occupied approximately 75% of all labeled samples. Then, for each case (***A*** to ***I***), the average OA and Kappa coefficient values were calculated; these are shown in Table 1. The confusion matrix closest to the average OA is presented in Fig. 5, where each number is normalized by its row and column simultaneously and visualized using colour depth.

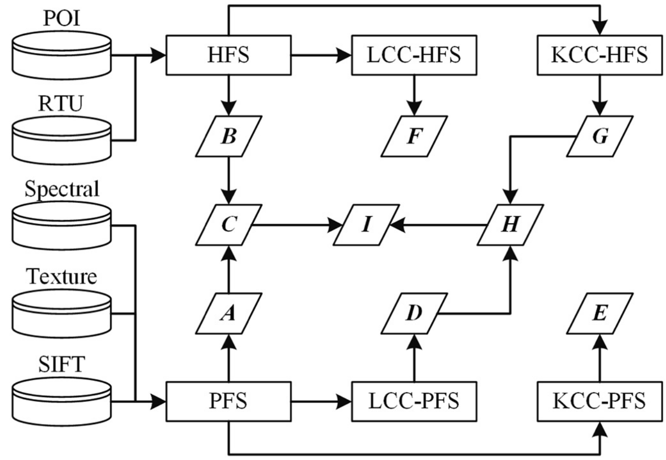


Fig. 3. Diagram of the feature designs in nine examples.

4.2. Urban functional land use recognition results

The trails from ***A*** to ***I*** can be divided into three hierarchies. The first hierarchy (***A*** to ***C***) settles the solution using physical features, human features and their combination. The second hierarchy (***D*** to ***H***) tests the effects of the cross-correlations across multiple feature spaces on functional land use recognition. The third hierarchy (***I***), namely, our CC-FLU model, is a combination of the first and second hierarchies. The accuracies of these three hierarchies are shown in Table 1, and our CC-FLU model (***I***) obtains the highest OA and Kappa coefficient values (OA-0.851, Kappa-0.812). In addition, the confusion matrix in Fig. 5(***I***) shows that this CC-FLU model performs well in distinguishing each type

Table 1
Urban functional land use classification accuracy of different examples.

Id.	Process	PFS			HFS		OA	Kappa
		Spectral	Texture	SIFT	POI	RTU		
A	LDA	√	√	√			0.693	0.615
B	LDA				√	√	0.649	0.557
C	***A*** + ***B***	√	√	√	√	√	0.824	0.778
D	CCA	√	√	√			0.711	0.638
E	KCCA	√	√	√			0.675	0.593
F	CCA				√	√	0.605	0.501
G	KCCA				√	√	0.623	0.524
H	***D*** + ***G***	√	√	√	√	√	0.817	0.773
I	***C*** + ***H***	√	√	√	√	√	***0.851***	***0.812***

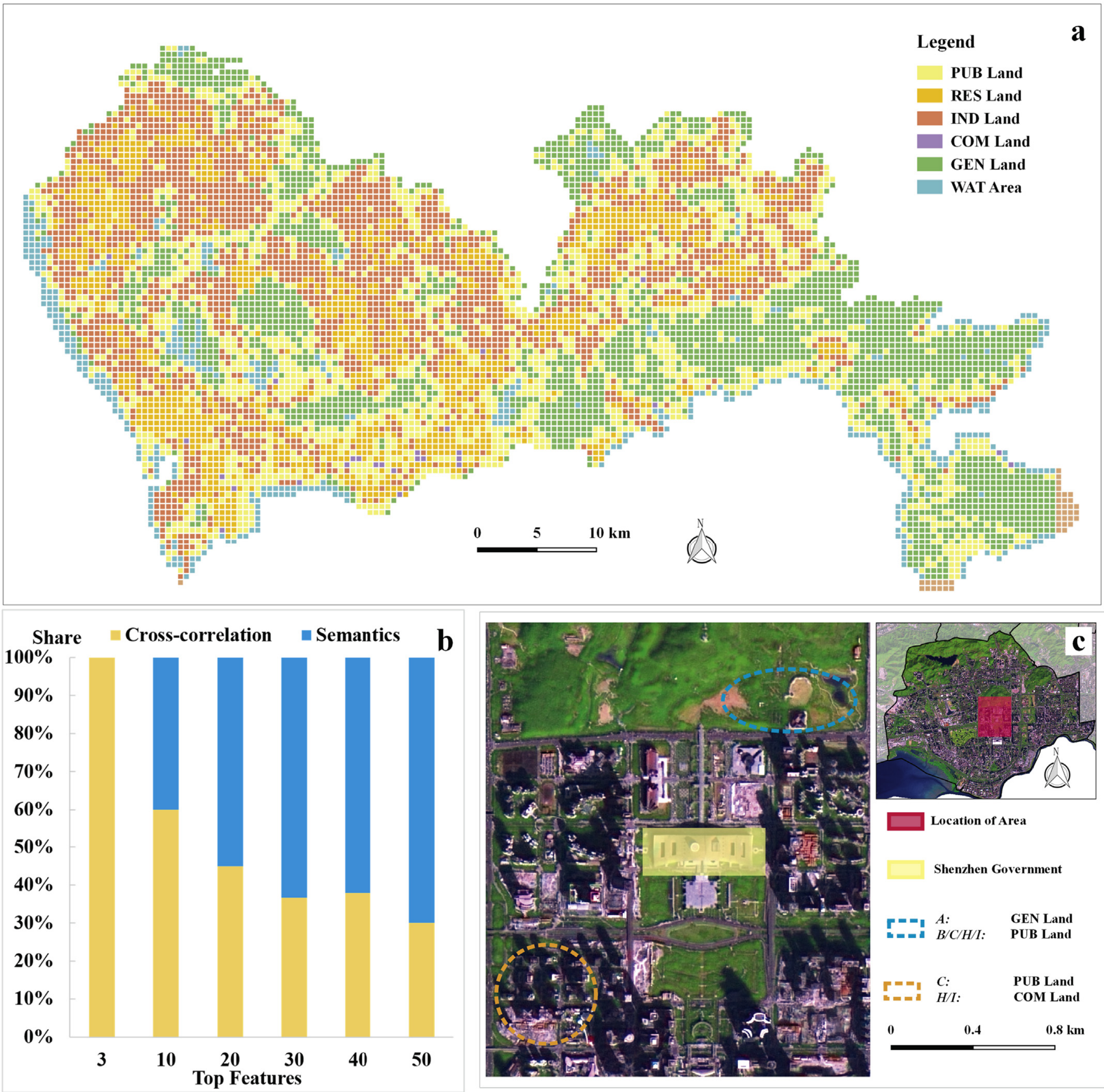


Fig. 4. Recognition results of the CC-FLU model. (a) Shenzhen functional land use map obtained by the CC-FLU model. The small brown area in the southeast corner is the absent area due to missing data. (b) The importance evaluation of the top-fifty features used in the CC-FLU model. (c) The details of results comparison by several different models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of functional land use.

Fig. 4 presents the detailed results of our CC-FLU model. As shown in the recognition map of Fig. 4(a), the geographical distribution of land use in Shenzhen is variable and complicated. A great deal of GEN land is distributed in the southeast, where are the natural-reserved area and the tourist destination. The industrial land (IND), public management and service land (PUB) and residential land (RES) occupy the widest areas of the built-up environment of Shenzhen. As urban functions hold great influences on the urban vitality, traffic and transportation (Abdullahi et al., 2015; Dovey & Pafka, 2017), this produced functional urban land use map in high-density Asia cities will benefit urban planning and management. For example, city planners would evaluate the current urban land usage status and provide useful insights into the policy-making for urban renewal.

Fig. 4(b) provides the importance evaluation of features used in the CC-FLU model. The result shows that the top three important features are all obtained from the cross-correlations, and this share in the top ten features is 60%. Then, as the ranking moves backward, the share of features extracted from PFS and HFS is increasing, and achieves the highest in the top fifty features with 70%. This demonstrates the great effects of introducing the natural correspondence on functional urban land use recognition.

Fig. 4(c) shows some details of the comparison of partial land use models in this study around the Shenzhen government, which is one of

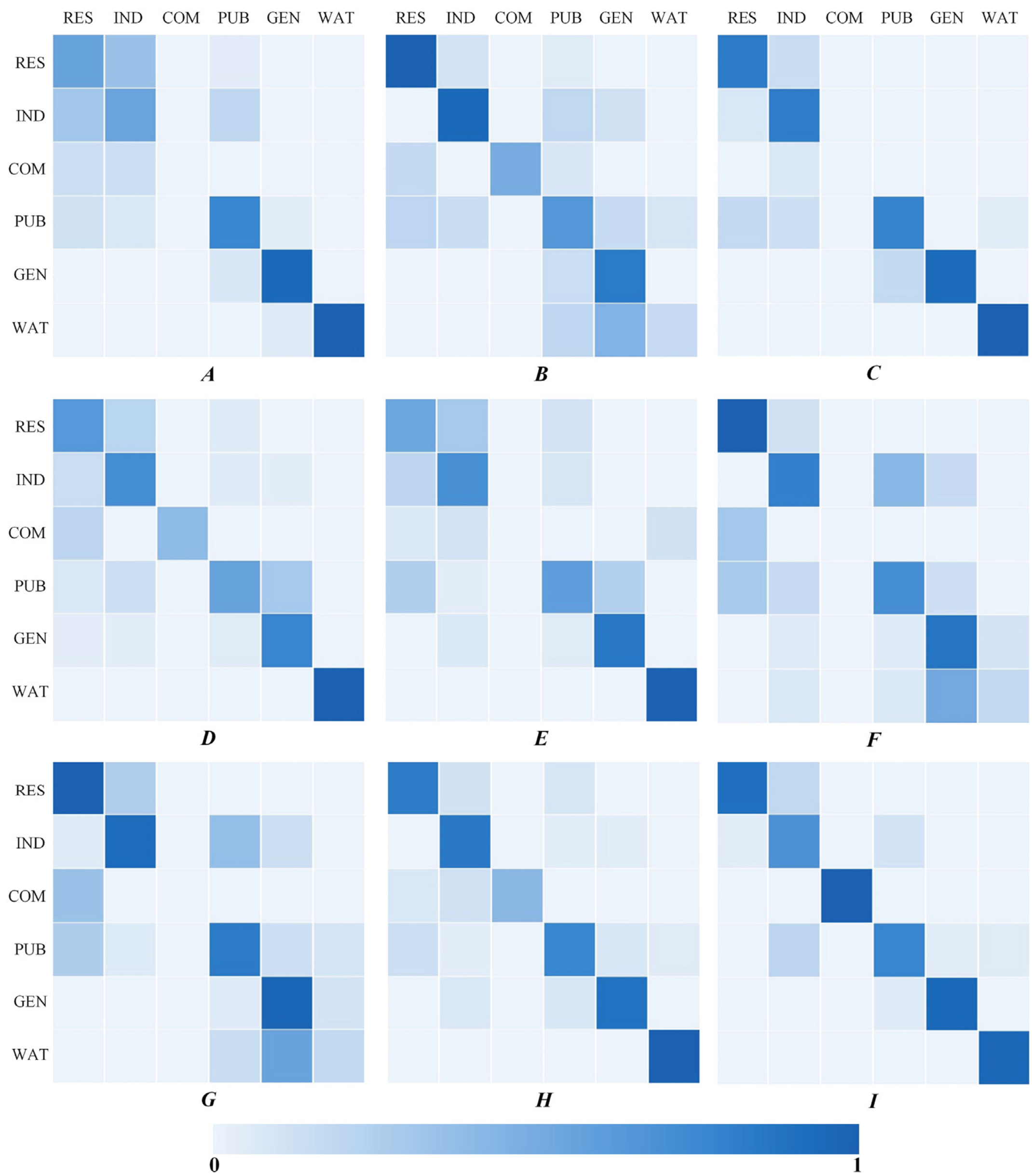


Fig. 5. Confusion matrices of different examples (***A*** to ***I***).

the most prosperous areas of Shenzhen. In the upper circle, the actual land use property is PUB land, which comprises an art museum and a memorial garden in a leisure park. However, two different land use properties were obtained through these models. In case ***A***, using PFS, this circle was recognized as GEN land, mostly due to its similar physical features as green and forest land. In case ***B***, using HFS, this circle was classified as PUB land based on the judgment of place semantics and human behavior. In the fusion models that contain HFS (***C***, ***H***, ***I***), we obtained the right result and thus illustrated the advantages of multi-source data fusion. The point is that each geospatial data type has its unique features, and correctly choosing and fusing the fitting data would greatly benefit the accomplishment for certain recognition task.

In the lower circle of Fig. 4(c), the actual land use property is COM land, gathering several commercial and financial buildings, and a shopping park. However, through the recognition result obtained from ***C***, this area is classified as a PUB land, and a few clues can be obtained from the confusion matrix of Fig. 5(***C***), which suggests that case ***C*** fails in the recognition of COM land. In case ***H***/***I*** when the cross-correlations across several feature spaces are utilized to distinguish this area, a right type is obtained, and the confusion matrices in Fig. 5(***H***/***I***) also prove that these two cases hold a good discernment capacity in the COM land. In this way, the state among multi-source geospatial data achieves the natural correspondence through the cross-correlations, which can help to extract some more efficient information comparing with previous models in functional urban land use recognition. The produced functional land use map is important for city planners to evaluate current urban development and provide advices for urban running.

4.3. The performance of multi-source data fusion

The differences among cases ***A*** (PFS), ***B*** (HFS) and ***C*** (PFS, PHS) shown in Table 1 suggest the performance of multi-source data fusion with different data groups in the first hierarchy. The comparison of OA and Kappa coefficient values reveals that ***C*** performs much better than ***A*** and ***B*** and that ***A*** is slightly better than ***B*** because human semantic features are spare at water area.

However, more details can be found in the confusion matrices shown in Fig. 5 (***A***, ***B***, ***C***). Comparing ***A*** with ***B*** reveals that ***A*** is excellent in distinguishing the land of PUB, GEN and WAT land but performs poorly on RES, IND and COM land, where the landscape is spatially complex but supports daily human activities. It can be explained that the features in case ***A*** are all physical feature semantics, which allow it to depict the physical properties of land patches well but cannot acquire abundant socioeconomic features. In contrast, ***B*** performs well in the recognition of RES and IND land but is not skilled in distinguishing WAT land where there are very few POIs and RTU. Also, POI data provide the rough land use attributes of certain areas, and RTU data reflect the temporal variations in human activity. Both of them are related to socioeconomic issues and human mobility. Therefore, they allow ***B*** to recognize RES and IND land well but cause it to have difficulty in recognizing land patches with physical land use properties, where there are fewer human activities.

Case ***C*** represents an effective solution to combine physical feature semantics and human activity semantics. In Table 1, ***C*** yields high OA (0.824) and Kappa (0.778) values, and it also exhibits a very good discernment capacity in the recognition of RES, IND, PUB, GEN and WAT land in Fig. 5(***C***). In this combination, case ***C*** almost combines the advantages of ***A*** and ***B*** and performs relatively well but exhibits shortcomings in the recognition of COM land.

4.4. The performance of cross-correlations

The second and third hierarchies refer to the framework of using cross-correlations to recognize functional urban land use. The second hierarchy is designed to first test the recognition performance of the cross-correlations on functional urban land use, and compare the difference of linear and nonlinear cross-correlations, and then exhibit the advantages of the cross-correlations in urban land use recognition. Following the above two hierarchies, a selected combination of the best performed semantics group in the first hierarchy and the best performed cross-correlations group in the second hierarchy, is employed in the third hierarchy to recognize functional urban land use.

In the inner comparison of the second hierarchy in Table 1, we find that the linear cross-correlations perform well in PFS, while the nonlinear cross-correlations perform well in HFS. Essentially, even though the spectral, texture and SIFT semantics are from different feature spaces, they are all extracted from the same HSR image, which endows PFS with a better correspondence using linear rather than nonlinear cross-correlations on this recognition task, especially in the recognition of COM land. However, for two HFSs, it has been implied that nonlinear projections exist between the place properties (POI) and population density (RTU) in geographical space (Yao et al., 2017). The correspondence provided by the nonlinear cross-correlations holds a better discernment capacity on functional urban land use recognition than the linear. Thus, a combination of two cross-correlations groups with the better discernment capacity in PFS (case ***D***) and HFS (case ***G***) produces case ***H***, and yields a better performed result than ***D*** and ***G*** (OA-0.817, Kappa-0.773).

When comparing the first and second hierarchies through the OA values, Kappa values and confusion matrices in Table 1 and Fig. 5, we find that both the semantic information and the cross-correlations provide extra clues with which to recognize functional land use that the other does not provide. First, the comparison of ***A*** and ***D*** shows that ***D*** holds a higher accuracy than ***A***, revealing that the cross-correlations got from the PFS endows the classifier with a better discernment ability in the land use recognition, especially in the COM land. Second, in the comparison of ***B*** and ***G***, ***B*** attains a higher accuracy, revealing that ***B*** provides more information with which to recognize functional land use than ***G***. Third, in the comparison of ***C*** and ***H***, ***C*** is the one with better OA and Kappa values. However, if we examine the confusion matrices, we can find that this result does not reflect the absolute superiority of ***C*** relative to ***H***. Compared to ***C***, ***H*** shows a balanced discernment capacity for every functional land use type, which ***C*** does not have.

Therefore, a better solution is to combine semantic information and their cross-correlations together and then select the best important features, which represents our CC-FLU model (***I***). ***I*** obtains the highest accuracy based on its OA (0.851) and Kappa (0.812) values, and the confusion matrix shown in Fig. 5(***I***) demonstrates that the CC-FLU model distinguishes every type of functional land use well. These results prove the superiority of the CC-FLU model over the first two hierarchies through the combination of their best important features. In general, this CC-FLU model represents a new feature selection pattern of multi-source geospatial data fusion model with the involvement of the natural correspondence, and obtains a better performed recognition result.

4.5. Parameter sensitivity analysis

Several parameters of the CC-FLU model may influence the discrimination accuracy, such as the numbers of spectral, GLCM and SIFT visual words and the numbers of LDA topics for the five used features. In the process of generating BOW, identifying the visual words and their counts is of great importance (Hardoon et al., 2004; Zhong et al., 2015). Unlike the POI and RTU data, whose words and word counts are fixed due to their data type, we utilize the K-Means method to cluster the spectral, GLCM and SIFT features, and the DBI is used to score the clustering results with various numbers. In Fig. 6(a), the DBIs decrease as the clustering numbers increase, but the speed of decrease tends to slow down and even stabilize. According to the previous study (Csurka et al., 2004), the clustering number of 300 is pinpointed as the word number where the tendency of DBI scores begins to become gentler. In terms of the texture (GLCM) and spectral features in Fig. 6(b), the DBIs achieve better scores when the clustering numbers are 4 and 3,

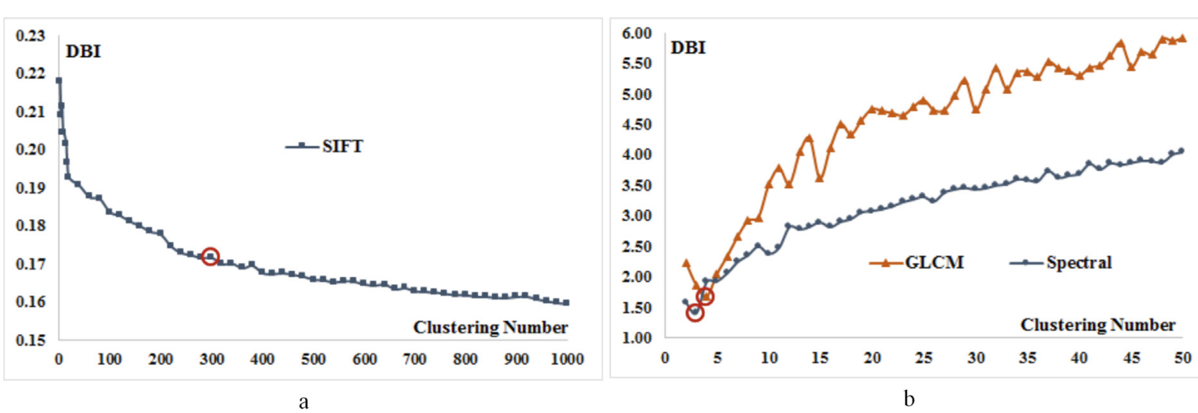


Fig. 6. DBI of K-means results for different features in different clustering numbers: (a) SIFT: the red circle represents the clustering number of 300; (b) GLCM and Spectral: the red circles represent the clustering numbers with the lowest DBI, namely, 4 and 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

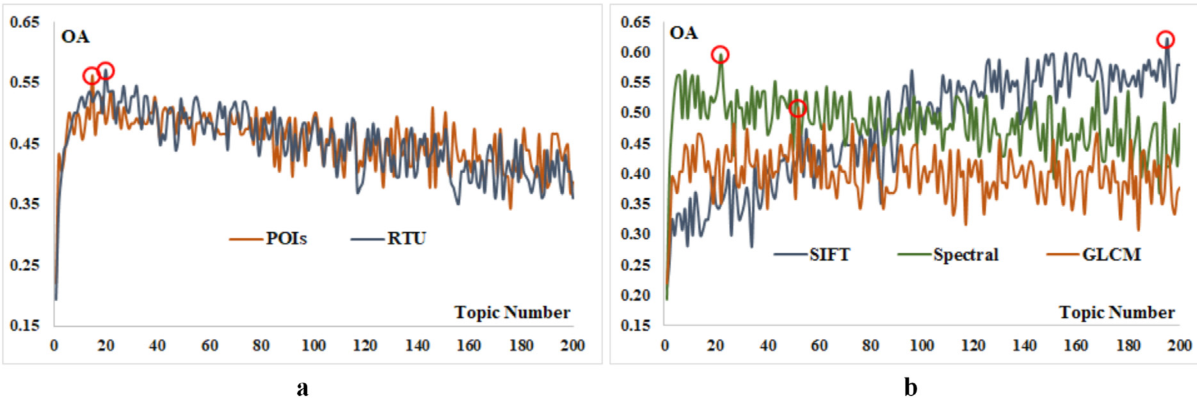


Fig. 7. RF classification accuracies for single features with different LDA topic numbers. (a) POIs and RTU: when the topic numbers are 15 and 20, these two features, respectively, yield their highest OA values. (b) SIFT, GLCM and spectral: when the topic numbers are 195, 52 and 22, these three features, respectively, yield their highest OA values.

respectively.

The identification of LDA topic numbers represents another difficult question. We solve this problem by comparing the RF classification accuracy results of different LDA topic numbers, and the training and validation datasets are the same as those discussed in Section 4.1. As shown in Fig. 7(a), the shapes of the two dotted lines representing the OAs of the POI and RTU semantic features fluctuate as the LDA topic numbers increase, but the overall trends are both peak-shaped. The OAs achieve their peaks when the LDA topic numbers are 15 and 20 for POIs and RTU, respectively. In terms of the features extracted from HSR images of Fig. 7(b), the overall trends of OAs using GLCM and spectral semantic features both fall as the topic numbers increase, and the highest values occur at the points of 52 and 22, respectively. For the SIFT semantic features, the OA values increase with some fluctuation as the LDA topic numbers increase, and they reach their highest value when the topic number is 195. Specifically, because the best topic number of SIFT is far greater than the best topic numbers of spectral and texture features, we fuse the spectral and texture semantic features together to calculate their cross-correlations with SIFT semantic features in our CC-FLU model.

5. Discussion

Multi-source geospatial data fusion is an important issue in functional urban land use recognition. On the basis of scene classification, this study proposed a functional land use recognition model (CC-FLU) integrating the semantic information from multi-source geospatial data and their cross-correlations. This CC-FLU model represents a new attempt to make full use of the information in multi-source geospatial data and hidden in their common subspaces. Our experiments verified that the CC-FLU model achieved a better performance, with OA and Kappa values of 0.851 and 0.812, respectively, than those in previous studies, such as scene classification methods using probabilistic topic models and HSR images (Zhong et al., 2015), urban function sensing models using spatiotemporal big data (Tu et al., 2018a). However, some issues need to be discussed further.

The first concerns the selection of multi-source geospatial data. In this study, three types of geospatial data were used, including data representing physical features (HSR image), static POI labels, and dynamic social media users, and two types of semantics were generated, including physical feature semantics and human feature semantics. However, this model does not have limitations associated with its data sources. Other features or geospatial datasets that can reveal the functions of land patches can also be used, such as nighttime light data (Ma et al., 2015), mobile phone records (Pei et al., 2014; Tu et al., 2017), and taxi trajectories (Yuan et al., 2012). They can be abstracted as other semantics and input into our model, perhaps in the format of place semantics or spatiotemporal interactive semantics.

The second concerns the scale problem of multi-source geospatial data fusion. The scale issues of geography significantly affect geographical phenomena and processes (Li & Cai, 2005). Due to the purpose of this study focusing on the test of the CC-FLU model, a fine-scale grid division of 500 m is directly utilized. However, the scale differences between data sources in multi-source geospatial data fusion have an impact. For example, one POI refers to a point with a rough land use property and coordinate (Yuan et al., 2012), while the actual cover of a POI is a region or perhaps a building, courtyard or community. Its actual cover is not equal in different POI types, and its corresponding area in the HSR image is variant. In further studies, the geographical scale matching of different data sources should be developed to obtain more precise recognition results.

The third issue is how to identify the natural correspondence across multiple feature spaces obtained from multi-source geospatial data. This study proposed a solution of the combination of cross-correlations across multi-semantic feature spaces through the projection on their intermediate isomorphic subspaces (Pereira et al., 2014). Here, the extraction of cross-correlations involves two feature spaces. If three or more feature spaces are available, the method of canonical correlations can handle the problem by generalization (Tenenhaus & Tenenhaus, 2011). In addition, deep learning has recently been a popular tool with which to recognize land use (Marmanis et al., 2016; Zhang, Zhang, & Du, 2016), but determining how to integrate multi-source geospatial data into deep learning must be explored in further studies of urban functions.

6. Conclusion

Land use classification mainly relies on remote sensing images. Recently, the emergence of spatiotemporal big data has transited land function recognition in the physical dimension to those in the socio-economic or human activity dimension (Xin and Meng, 2018), thus making it an inevitable trend to fuse multi-source geospatial data efficiently and effectively. To solve the possible biased problem of one single data source, this study proposed the CC-FLU model to recognize functional urban land use through the natural correspondence across multi-source data. The experiment in Shenzhen demonstrates that the presented approach works well, yielding OA and Kappa values of 0.851 and 0.812, respectively. This produced functional urban land use map in high-density Asia cities enables urban planners to be aware of urban development and provide useful insights on the policy-making for urban renewal in the future.

The main contributions of this study are two-fold. First, the natural correspondence across multi-source geospatial data is achieved through the cross-correlations extracted respectively from physical feature semantics and human feature semantics, which is a new approach to fusing multi-source geospatial data. Second, based on the proposed natural correspondence, a general framework is designed to more accurately recognize functional urban land use combining the semantics across multiple feature spaces and their cross-correlations, which can be further developed to integrate other urban data.

In general, this study presents a new way to combining multi-source geospatial data to recognize functional urban land use. In future work, more models referring to multi-source geospatial data fusion will take the natural correspondence into account, and the feasibility and advantages of this approach will be tested in more applications.

Acknowledgments

This work was jointly supported by the National Natural Science Foundation of China (#71961137003), and the Basic Research Program of Shenzhen Science and Technology Innovation Committee (No. JCJY 201803053125113883, No. JCYJ20170412105839889). We would like to thanks Prof. Christopher Pettit and two anonymous reviewers for their constructive comments.

References

Abdullahi, S., et al. (2015). GIS-based modeling for the spatial measurement and evaluation of mixed land use development for a compact city. *Mapping Sciences and Remote Sensing, 52*(1), 18–39.
Andrew, G., et al. (2013). Deep canonical correlation analysis. *International conference on machine learning* (pp. 1247–1255). .
Baltru, V. S., Aitis, T., Ahuja, C., & Morency, L. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443.
Blaschke, T., et al. (2014). Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing, 87*, 180–191.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(1), 993–1022.
Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., ... Qiu, G. (2018). Integrating aerial and street view images for urban land use classification. *Remote Sensing, 10*, 1553.
Chen, Y., et al. (2017). Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method. *Landscape and Urban Planning, 160*, 48–60.
Csurka, G., et al. (2004). *Visual categorization with bags of keypoints. Workshop on statistical learning in computer vision, ECCV. Prague.* 1–22.
Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*(2), 224–227.
Deng, H., & Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition, 46*(12), 3483–3489.
Dovey, K., & Pafka, E. (2017). What is functional mix? An assemblage approach. *Planning Theory & Practice, 18*(2), 249–267.
Fei, W., & Zhao, S. (2019). Urban land expansion in China's six megacities from 1978 to 2015. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv. 2019.02.008.
Fern A Ndez-Delgado, M., et al. (2014). Do we need hundreds of classifiers to solve real world classification problems. *Journal of Machine Learning Research, 15*(1), 3133–3181.
Forestier, G., et al. (2012). Knowledge-based region labeling for remote sensing image interpretation. *Computers, Environment and Urban Systems, 36*(5), 470–480.
Guan, C., & Rowe, P. G. (2016). The concept of urban intensity and China's townization policy: Cases from Zhejiang Province. *Cities, 55*, 22–41.
Guan, X., et al. (2018). Assessment on the urbanization strategy in China: Achievements, challenges and reflections. *Habitat International, 71*, 97–109.
Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Computation, 16*(12), 2639–2664.
Hsieh, S. C. (2014). Analyzing urbanization data using rural-urban interaction model and logistic growth model. *Computers, Environment and Urban Systems, 45*(4), 89–100.
Hu, S., & Wang, L. (2013). Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing, 34*(3), 790–803.
Li, S., & Cai, Y. (2005). Some scaling issues of geography. *Geographical Research, 24*(1), 11–18.
Lienou, M., Maitre, H., & Datcu, M. (2010). Semantic annotation of satellite images using latent Dirichlet allocation. *IEEE Geoscience and Remote Sensing Letters, 7*(1), 28–32.
Lillywhite, K., et al. (2013). A feature construction method for general object recognition. *Pattern Recognition, 46*(12), 3300–3314.
Liu, X., et al. (2017). Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science, 31*(8), 1675–1696.
Liu, Y., et al. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*(3), 512–530.
Ma, T., et al. (2015). Night-time light derived estimation of spatio-temporal characteristics of urbanization dynamics using DMSP/OLS satellite data. *Remote Sensing of Environment, 158*, 453–464.
Marmanis, D., et al. (2016). Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters, 13*(1), 105–109.
McDonald, R. I., Marcotullio, P. J., & G U Neralp, B. (2013). Urbanization and global trends in biodiversity and ecosystem services. *Urbanization, Biodiversity and Ecosystem Services: Challenges and Opportunities* (pp. 31–52). Springer.
Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International Journal of Scientific and Research Publications, 3*(5), 1.
Pei, T., et al. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science, 28*(9), 1988–2007.
Pereira, J. C., et al. (2014). On the role of correlation and abstraction in cross-modal multimedia retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(3), 521–535.
Rasiwasia, N., et al. (2010). *A new approach to cross-modal multimedia retrieval.* ACM251–260.
Ríos, S. A., & Muñoz, R. (2017). Land use detection with cell phone data using topic models: Case Santiago, Chile. *Computers, Environment and Urban Systems, 61*, 39–48.
Sapena, M., & Ruiz, L.Á. (2018). Analysis of land use/land cover spatio-temporal metrics and population dynamics for urban growth characterization. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.08.001.
Shahriari, M., & Bergevin, R. (2017). Land-use scene classification: A comparative study on bag of visual word framework. *Multimedia Tools and Applications, 76*(21), 1–17.
Tenenhaus, A., & Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika, 76*(2), 257.
Tu, W., et al. (2017). Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science, 31*(12), 2331–2358.
Tu, W., et al. (2018a). Portraying urban functional zones by coupling remote sensing imagery and human sensing data. *Remote Sensing, 10*(1), 141.
Tu, W., et al. (2018b). Spatial variations in urban public ridership derived from GPS trajectories and smart card data. *Journal of Transport Geography, 69*, 45–57.
United NationsWorld urbanization prospects, the 2018 revision. Population division, Department of Economic and Social Affairs*United Nations Secretariat*.
Wang, W., et al. (2015). On deep multi-view representation learning. *International conference on machine learning* (pp. 1083–1092). .
Wei, Y., et al. (2017). Cross-modal retrieval with CNN visual features: A new baseline. *IEEE Transactions on Cybernetics, 47*(2), 449–460.
Wu, C., Zhang, L., & Zhang, L. (2016). A scene change detection framework for multi-temporal very high resolution remote sensing images. *Signal Processing, 124*, 184–197.
Wu, X., et al. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26*(1), 97–107.
Xing, H., & Meng, Y. (2018). Integrating landscape metrics and socioeconomic features for urban functional region classification. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.06.005.
Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Computers, Environment and Urban Systems, 72*, 51–67.
Yao, Y., et al. (2016). Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science, 31*(4), 825–848.
Yao, Y., et al. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science, 31*(6), 1220–1244.
Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 186–194). .
Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine, 4*(2), 22–40.
Zhang, W., et al. (2017). Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. *Computers, Environment and Urban Systems, 64*, 215–228.
Zhang, X., & Du, S. (2015). A linear Dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings. *Remote Sensing of Environment, 169*, 37–49.
Zhang, X., Xu, Y., Tu, W., & Ratti, C. (2018). Do different datasets tell the same story about urban mobility — A comparative study of public transit and taxi usage. *Journal of Transport Geography, 70*, 78–90.
Zheng, X., et al. (2016). Crowdsourcing based description of urban emergency events using social media big data. *IEEE Transactions on Cloud Computing, 99*, 1.
Zheng, Y. (2015). Methodologies for cross-domain data fusion: An overview. *IEEE Transactions on Big Data, 1*(1), 16–34.
Zhong, C., et al. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems, 48*, 124–137.
Zhong, Y., Zhu, Q., & Zhang, L. (2015). Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *IEEE Geoscience and Remote Sensing, 53*(11), 6207–6222.
Zhou, L., Zhou, Z., & Hu, D. (2013). Scene classification using a multi-resolution bag-of-features model. *Pattern Recognition, 46*(1), 424–433.