ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Integrating landscape metrics and socioeconomic features for urban functional region classification

Hanfa Xing[a,b], Yuan Meng[b,*]

[a] School of Geography, South China Normal University, Guangzhou, Guangdong, China
[b] College of Geography and Environment, Shandong Normal University, Jinan, Shandong, China

ARTICLE INFO

Keywords:
Urban functional region
Classification
Landscape metrics
Crowdsourced data
Socioeconomic features

ABSTRACT

Urban functional region classification plays an important role in urban planning, resource management and sustainable development. While previous studies have solely focused on delineating regions by urban morphology or deriving functional types from human activities, the sufficient depiction and integration of these characteristics are seldom discussed. This paper inherently considers both urban morphology and human activities to classify urban functional regions. Building-based and region-based landscape metrics derived from building-level blocks are delineated to measure urban morphology, while socioeconomic features are extracted from crowdsourced data related to human activities using a topic model and semantic scaling method. These characteristics are then fused by applying random forest to measure different functions. This approach is tested in Futian district, Shenzhen, Guangdong. The result shows an overall classification accuracy of 0.818, approximately 0.15 higher than that utilising only landscape metrics or only socioeconomic features. This result indicates the effectiveness of the delineated characteristics to depict urban landscapes and socioeconomic information and the reliability of integrating these features for urban functional region classification.

1. Introduction

Urban functional regions refer to the areas in which human activities occur, including residential, commercial, and industrial land. They serve as an essential indicator in designing urban environments, managing natural resources, and detecting rapid urbanisation (Anderson, 1976; Treitz, Howarth, & Gong, 1992; Wentz, Stefanov, Gries, & Hope, 2006). Classifying urban functional regions is crucial for uncovering the characteristics of urban cities, which comprise not only urban regions delineated from physical properties but also socioeconomic functions driven by human activities (Pei et al., 2014; Xiuyuan, Zhang, Du, & Wang, 2017).

Significant progress has been made in sensing physical properties to depict urban forms. Conventionally, this relies heavily on remote sensing images, where spectral and textural characteristics are extracted for presenting urban land use information (Lu & Weng, 2006; Weng, 2012). Current analyses of functional region detection focus on three levels of perspectives, namely, pixel, object and scene-based classification (Gong & HOWARTH, 1990; Ma et al., 2017; Xiuyuan, Zhang & Du, 2015). Although the pixel and object-based classification approaches have gained popularity in deriving land cover features, they fail to capture detailed functional information because of the complex urban form (Hu, Yang, Li, & Gong, 2016). On the other hand, scene-based classification shows advantages in identifying urban functional patterns (Xiuyuan, Zhang & Du, 2015). However, it can only decompose and measure the whole picture of the mixed scenes, whereas detailed urban morphology, which is often strongly correlated with the functional type, cannot be captured (Vanderhaegen & Canters, 2017).

In fact, remotely sensed images can to some extent reveal the morphology of urban regions, which is defined as the presence, size, shape and spatial distribution among urban landscapes, including different constructed and open spaces. Quantifying landscapes with effective metrics may help improve the delineation of urban morphology. The relationship between urban landscapes and functions have been discussed in many studies (Baus, Kováč, Pauditšová, Kohutková, & Komorník, 2014; Li, Peng, Yanxu, & Yi'na, 2017; Van de Voorde et al., 2016). Typically, building-level blocks are one of the vital aspects in functional regions (Steiniger, Lange, Burghardt, & Weibel, 2008). Many studies have investigated the potential of buildings with different metrics in urban patterns analysis. Wu et al. (2018) proposed an Extended Minimum Spanning Tree method to describe both spatial proximity and building characteristics to discover local urban patterns. In addition, Chen, Xu, and Devereux (2014) and Liu, Hu, and Li (2017) have designed 2- and 3-dimensional scales of landscape metrics to effectively

* Corresponding author.
E-mail addresses: xinghanfa@sdnu.edu.cn (H. Xing), mengyuan@stu.sdnu.edu.cn (Y. Meng).

https://doi.org/10.1016/j.compenvurbsys.2018.06.005
Received 5 February 2018; Received in revised form 12 April 2018; Accepted 15 June 2018

Please cite this article as: Xing, H., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.06.005

reflect building structures. The landscape metrics proposed for buildings enable further investigations into functional regions. For example, urban block perimeters vary among different functional types: industrial and public regions usually compromise a low density of large individual buildings, while residential regions often contain continuous blocks with small areas, which leads to a much larger value of the perimeter. In fact, an attempt has been made to distinguish different functional regions with building-based metrics (Vanderhaegen & Canters, 2017), proving the ability to describe urban landscapes and to detect functional regions. Despite the achievement of depicting the ground components in urban regions, the functional types that are highly influenced by the interaction between environmental landscape and human activities have not been further explored (Tu et al., 2017).

The massive amounts of crowdsourced data, such as points of interest (POIs), social media data, trajectories and mobile phone signals potentially provide information about human activities and socioeconomic information, which have been investigated to indicate urban functions in many studies (Yu et al., 2015; Zheng, Capra, Wolfson, & Yang, 2014). Yuan, Zheng, and Xie (2012) employed POIs and human mobility to infer functional regions by identifying the distributions of mobility patterns. Gao, Janowicz, and Couclelis (2017) analysed the cooccurrence patterns of POI types to discover functional regions at a semantic level. Semantic information in crowdsourced data can reflect socioeconomic features such as commercial, industrial and residential information, which are considered significant indicators of urban patterns. Yao, Li et al. (2017) proposed a POI-based framework to extract semantic information using a Word2Vec model for sensing functional types. Lansley and Longley (2016) introduced the latent Dirichlet allocation (LDA) topic model to mine functional distributions using Twitter messages. However, due to the spatial heterogeneity of crowdsourced data (Fonte, Bastin, See, Foody, & Lupia, 2015), regions containing less human activities can hardly be differentiated without identifying urban morphology (e.g., people seldom post activity information or generate points of interest in undeveloped open regions); thus, capturing only socioeconomic features cannot meet the requirement of effectively detecting urban functions.

Therefore, combining both urban morphology and human activities has the potential to effectively reveal urban functions. Previous studies have attempted to fuse remote sensing images and crowdsourced data to monitor land cover/land use applications. Liu et al. (2017) proposed a land use word dictionary by extracting natural features from high spatial resolution images and semantic features from social media data. Hu et al. (2016) generated urban functional regions based on the spatial distribution of POIs and on spectral information from satellite images. Moreover, a recent attempt has been made to measure both landscape patterns and human activities from remote sensing images and cell phone data (Tu et al., 2018). Since they delineated functional zones based on regular grids with 2000-m resolution using only four landscape metrics, the urban morphology, which is highly related to the community or building scale, cannot be captured.

Although these studies have integrated both physical properties and socioeconomic information, which result in significant improvements of the functional region classification, challenges still remain in the following aspects. First, they simply extracted and fused a variety of natural features from remote sensing images and failed to quantitively measure the landscape to describe the detailed urban morphology. Second, as crowdsourced data are usually generated irregularly with spatial heterogeneity, effectively quantifying indicators is required to relate the unorganised information with certain functional types. Third, the integration of urban landscapes and socioeconomic information to better delineate the functional regions remains undiscussed.

Faced with the above issues, this paper proposes a functional region classification approach by integrating landscape metrics and socioeconomic features, which includes three major steps. In the first step, landscape metrics are adopted to capture urban physical properties. Particularly, metrics with both building level and functional region level are considered in order to obtain comprehensive descriptions of urban landscapes at a fine scale. In the second step, a variety of semantic information from crowdsourced data is automatically calculated for characterising socioeconomic patterns. Faced with the spatial heterogeneity of crowdsourced data, semantic scaling is proposed to normalise semantic information and extract socioeconomic features. Finally, by utilising both landscape metrics and socioeconomic features, random forest is applied to gain better insight into the diversity of functional types.

The major advantage of the methodology is that it inherently integrates landscape and socioeconomic characteristics. In particular, by calculating landscape metrics from building-level blocks and extracting function-related topics via crowdsourced data, the presented approach not only delineates the underlying urban morphological and social properties of urban forms but also provides an initial insight into identifying different functional types. To validate the presented approach, an analysis of variance is proposed to evaluate the contribution of each landscape metric and socioeconomic feature to functional region classification. Furthermore, comparative experiments are proposed utilising solely landscape metrics and socioeconomic features to prove our hypothesis.

The contributions of our work involve three main aspects. First, urban regions are delineated by extracting physical properties from a landscape perspective. We quantify the spatial morphology of urban regions by constructing building-based and region-based metrics. Second, functional types are determined by socioeconomic features derived from crowdsourced data. By calculating different topics using the topic model and a semantic scaling method, socioeconomic information can be viewed from the perspective of human activities and shows potential to be applied as one of the effective data sources for urban function arrangement. Third, we inherently consider landscape metrics and socioeconomic features to automatically identify different urban functions. As most studies mainly focused solely on urban landscape or social information, the combination of both can give a better understanding of functional region classification.

The remainder of this paper is organised as follows. In Section 2, the methodology for functional region classification is presented considering both landscape metrics and socioeconomic features. Section 3 introduces the study area and experimental data. The classification results are presented in Section 4. The utilisation of experimental data, the selection of both landscape metrics and socioeconomic features, and the identification of mixed function regions are discussed in Section 5. Section 6 concludes the presented work and discusses future studies.

2. Methodology

2.1. Overview

The framework of urban functional region classification is displayed in Fig. 1, including three major steps. In the first step, we utilise building-level blocks to delineate the urban morphology (Section 2.2). In particular, the basic measurement of building-level blocks consists of four indices including the building area, edge, height and structure (Section 2.2.1). These indices are combined for calculating building-based metrics and region-based metrics (Section 2.2.2). The constructed landscape metrics are used to delineate physical properties from the landscape perspective, capture urban morphology information, and further provide essential natural characteristics for functional region classification.

In the second step, crowdsourced data are used to extract socioeconomic information (we choose POIs as crowdsourced data in the following experiment) (Section 2.3). First, using an LDA (latent Dirichlet allocation) topic model, we mathematically extract topics containing words with probabilities from crowdsourced data. Specifically, the number of topics is determined by a perplexity algorithm, with a lower perplexity value representing a more appropriate topic