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

Representing place locales using scene elements

Fan Zhang[a,b], Ding Zhang[b], Yu Liu[a], Hui Lin[b,c,*]

[a] Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China
[b] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China
[c] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Scene ontology
Quantitative representation
Street-level images
Deep learning
Place formalization

ABSTRACT

Locale is one of the basic elements of place, referring to the physical settings and visual appearance of a place. Understanding and representing a locale is of great importance in terms of human perception and human activity. However, taking a quantitative measurement of the visual appearance of urban environment has proven to be challenging because visual information is inherently ambiguous and semantically impoverished. To mitigate this issue, this paper employs street-level images as the proxy for urban physical appearance, utilizes the recently developed image semantic segmentation techniques to parse an urban scene into scene elements, and proposes a framework for locale representation using scene elements. The framework is composed of two major components: *street scene ontology* and *street visual descriptor*, which are aimed at street scene qualitative understanding and quantitative representation respectively. A case study is developed to demonstrate the application and advantage of the street scene ontology and street visual descriptor. A series of quantitative analyses demonstrates the ability and great potential of the framework for investigating the connections between place and other socioeconomic factors.

1. Introduction

Place, as one of the most important and fundamental concepts in the research of geography, has been defined as "…spatial locations that have been given meaning by human experiences" (Tuan, 1977). Recently, with the development of Geographic Information Science (GIScience), place formalization in the digital world has been proposed, offering a way to explore connections between the informal world of human cognition and the formal world of digitally represented geography (Goodchild, 2011).

Location, locale, and sense of place have been accepted as three basic elements of a place (Agnew, 2011; Routledge, 1993). In particular, location refers to in which a scene or activity is located (Agnew, 1987); locale indicates the physical settings where everyday-life activities take place, including visible and tangible aspects of a place such as buildings, streets, parks, etc. (Cresswell, 2014); and sense of place refers to the human experience and nebulous meanings associated with a place (Agnew, 2011; Cresswell, 2014). Research efforts have been made to formalize a place in terms of two aspects: location, namely, to represent a vague cognitive region of a place (Gao et al., 2017; Montello, Friedman, & Phillips, 2014), and the sense of place, which mines place semantics regarding human activities and perceptions (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009; Li & Goodchild, 2012; Rattenbury & Naaman, 2009). However, how to formalize the concept of place with respect to locale and how to build a quantitative representation of locale remain unclear. Although several attempts have been made (Doersch, Singh, Gupta, Sivic, & Efros, 2012; Quercia, O'Hare, & Cramer, 2014; Rundle, Bader, Richards, Neckerman, & Teitler, 2011) to analyze the physical appearance of an urban space by photos, the research scale and throughput have been limited by the accessibility of qualified data and the appropriate tools to process it (Salesses, Schechtner, & Hidalgo, 2013).

Recently, the emergence of big data has resulted in a massive number of geo-tagged datasets - especially street-level imagery, which covers general landscape of an urban area and can represent the physical appearance of an urban space completely (Gebru et al., 2017; Li, Ratti, & Seiferling, 2018), providing new opportunities to help formalize the concept of place with more dimensions and in a scalable manner (Goodchild, 2015; Liu et al., 2015; MacEachren, 2017; Rattenbury & Naaman, 2009). On the other hand, enabled by the proliferation of computer vision and deep learning techniques, it has been proven possible to acquire the semantic information of every single pixel in a natural image with high accuracy (Zhao, Shi, Qi, Wang, & Jia, 2017), thus improving our ability to semantically understand scene content (LeCun, Bengio, & Hinton, 2015).

Drawing upon the review of place formalization and new methods

* Corresponding author at: Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China.
E-mail address: huilin@cuhk.edu.hk (H. Lin).

https://doi.org/10.1016/j.compenvurbsys.2018.05.005
Received 4 August 2017; Received in revised form 15 May 2018; Accepted 16 May 2018

in deep learning, the purpose of this study is to formalize the concept of place in terms of locale - the physical appearance of place. We employ an image semantic segmentation technique to parse street-level images and obtain 64 scene elements (building, sky, grass, etc.) that constitute a typical street scene. Base on these 64 categories of scene elements, we propose a framework for locale representation. In particular, first, we propose street scene ontology, which is a hierarchical semantic tree that illustrates the conceptual relationships among the 64 scene elements, providing insight regarding the qualitative understanding of a locale. Second, driven by the scene ontology, we propose a scene visual descriptor - a 64-dimensional computational vector that can be used to represent and measure a general street scene quantitatively. Each dimension of the vector corresponds to the presence ratio of each scene element in the street view image. The scene visual descriptor enables the carrying out of measurements among places and contributes to the calculations between place and other spatial, demographic, and socioeconomic factors.

The remainder of this paper is organized as follows. In Section 2, the literature on place formalization and those studies using street-level images are briefly reviewed. Section 3 introduces the proposed locale representation framework, including the street scene ontology and street visual descriptor. Section 4 reports a case study in Hong Kong, with a series of analyses using the street visual descriptor being conducted. The case study shows that the visual appearance of streets is highly correlated with the spatial structure of street network. In Section 5, we draw some conclusions and discuss future work.

2. Literature review

2.1. Big data advances place formalization

In an earlier stage, most of the studies conducted to understand place were based on traditional interviews and questionnaires. For example, residents in Charleston, South Carolina, following Hurricane Hugo in 1989 were interviewed about their view on the place feature they had lost in the storm to find the major place-identity symbols (Hull, Lam, & Vigo, 1994). Also, earlier in 2001, Hidalgo and Hernandez interviewed 177 people, and found that attachment to place developed to different degrees with different spatial ranges and dimensions, obtaining more quantified results (Hidalgo & Hernandez, 2001). With the development of technologies, Quercia, O'hare and Cramer collected the voting results of 3.3 k individuals to quantify the aesthetic capital of each neighborhood and tried to determine the visual cues that may cause a street to be perceived as being beautiful, quiet or happy (Quercia, O'Hare, & Cramer, 2014). However, these traditional studies relying on case studies and interviews, which are costly and time-consuming, were limited by the lack of accessibility of data and the approaches to handle it, and faced difficulties resulting from a lack of resolution, scale and throughput in representing place (Goodchild, 2011; Salesses, Schechtner, & Hidalgo, 2013).

Enabled by the proliferation of wireless communication and mobile networking, the emergence of big data era in recent years has resulted in massive number of geo-tagged and semantic-tagged datasets, potentially providing new opportunities to enrich and enhance the representation of place semantics (Glaeser, Kominers, Luca, & Naik, 2016; Li & Goodchild, 2012; Liu et al., 2015). By analyzing the spatial-temporal distribution of footprints, Li et al. investigated the relationships between uneven data spatial patterns and the socioeconomic characteristics of geographic data creators in the United States (Li, Goodchild, & Xu, 2013). Sigurbjörnsson (Sigurbjörnsson & Van Zwol, 2008) explored the connections between a photo's geo-location and the semantic information that was contained in the photo tags, which somehow implied that urban place can be typically represented by semantic tags from crowd-sourced photo data. With the aid of place semantics mining methods (Rattenbury, Good, & Naaman, 2007; Rattenbury & Naaman, 2009), later works have demonstrated it (Hollenstein & Purves, 2010; Li & Goodchild, 2012)).

Semantic tags could provide a subjective description of a place according to users; on the other hand, street-level images depict the objective locale of a place. By integrating image content analyses with computer vision and deep learning, researchers can not only understand city's physical and dynamic characteristics - by detecting landmark (Hays & Efros, 2015), recognizing urban identities (Liu, Zhou, Zhao, & Ryan, 2016), evaluating the inequality of living environment (Li, Ratti, & Seiferling, 2018; Salesses, Schechtner, & Hidalgo, 2013), and modeling human activities (Arase, Xie, Hara, & Nishio, 2010) and popular places (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009); but also examine the physical and social structures of dynamic urban environments (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009; Less et al., 2015). Overall, street-level images potentially offer about new dimensions and perspectives for place formalization.

2.2. Ontology model for representing a natural scene

Ontology is the study of the natural aspects of being, reality, and relations from a common perspective and represents them in a formal language (Guarino, Oberle, & Staab, 2009). Most studies on scene ontology have been conducted in the field of information science, especially in the areas of image/video segmentation, detection and tracking (Francois, Nevatia, Hobbs, Bolles, & Smith, 2005; Gómez-Romero, Patricio, García, & Molina, 2011; Nevatia, Hobbs, & Bolles, 2004; Oliva & Torralba, 2001; Town, 2004). Others have used scene ontology for scene natural language description (Nwogu, Zhou, & Brown, 2011). Most of the previous studies were focused on a specific domain, such as instance tracking and event reasoning. However, geographers and urban researchers are concerned with the reciprocal interaction between a place and the individuals that live in it. From the perspective of urban studies, practical implementations of the formal representation of a scene are still scarce. Consequently, drawing upon the previous discussions on the ontology of geographic information science (Montenegro & Duarte, 2008; Smith & Mark, 2001; Smith & Mark, 2003), we introduce the street scene knowledge ontology - a semantic taxonomic tree - in this study to support the qualitative understanding of a general street scene.

2.3. Street-level images for representing the locale of a place

A street, as a major place for human mobility and activity, should be regarded as a basic element in urban studies. Of the 5 elements of the image of a city, i.e. landmarks, paths, nodes, districts and edges, (Lynch, 1960)) stated that the paths are the most predominant elements and thus deserve to be studied, the investigation of which would provide a foundation for the clustering and organization of the meanings and associations of the other four elements and the city as a whole. The literature advanced by other urban planners and geographers has also demonstrated the importance of streets and the necessity of taking them as the elementary unit from various perspectives (Jacobs, 1992; Jacobs, 1993; Jiang & Claramunt, 2004; Long, 2016; Zhu, Wang, Wu, & Liu, 2017).

Among various multi-source data, street-level imagery contains a large amount of visual information regarding an urban space and serves as an ideal proxy for representing the physical built environment. These numerous street-level images not only scale up the size and scope of the related research but also provide new perspectives and dimensions for formalizing the concept of place (Gebru et al., 2017; Less et al., 2015; Rundle, Bader, Richards, Neckerman, & Teitler, 2011). Researchers have employed street view images to reconstruct 3-dimensional urban models (Kim, Hyungki, Yuna, & Soonhung, 2014; Micusik & Kosecka, 2009; Sturm, 2000), to explore urban morphologies by mapping the distribution of image locations (Crandall, Backstrom, Huttenlocher, & Kleinberg, 2009), and to analyze the visual elements of an urban space in terms of human perception (Zhang, Hu, Che, Lin, & Fang, 2018),