Urban Land-Use Classification From Photographs

Fang Fang, Xiaohui Yuan, *Senior Member, IEEE*, Lu Wang, Yuanyuan Liu, and Zhongwen Luo

***Abstract*— Land-use (LU) classification of urban areas is conventionally achieved via field survey or remote sensing technologies, which is labor-intensive and time-consuming. With the wide development of social networks such as microblog and ubiquitous network access, images are captured by residents and tourists. In this letter, we propose a method for an automatic urban LU classification using geotagged images from public venues. Our method identifies the LU type depicted in those images that are extrapolated to the local regions bounded by street blocks. Experiments were conducted with geotagged photographs and Open Street Map of an urban area in London, U.K. It was demonstrated that the proposed method achieved overall 76.5% accuracy across five LU types. More importantly, our method demonstrated a greater performance in dealing with a mixture of LU types.**

***Index Terms*— Classification, land use (LU), urban, volunteered geographic information (VGI).**

I. Introduction

URBAN land use (LU) provides important derived data for applications such as environmental research, spatial planning, and urban management [1]–[4]. Knowledge of the distribution of vegetation, water body, commercial and residential land usage, as well as information on their changing proportions, is needed by the state and local governments and legislators to allocate resources, e.g., transportation infrastructure and water supply, and to identify regional development.

Urban LU has been conventionally achieved via field survey or remote sensing technologies, which is labor-intensive and time-consuming. The high cost and infrequent updates hardly meet the needs of rapidly changing modern cities. With the development of social networks such as photograph sharing, a large number of images have been generated by tourists and residents. The emergence of volunteered geographic information (VGI) data motivates the LU analysis from a new perspective [5]. Geotagged images are acquired and shared on Flickr, Facebook, and other social networks, which have become an important image source [6]. The geotagged images demonstrate a great promise for generating LU maps [7]. Yet, the unevenly distributed image in the urban space and number of geotagged images for different LU types and the great diversity of image content make the LU classification a great challenge.

In this letter, we propose a method for automatic urban LU classification using unevenly distributed geotagged images. We leverage geotagged photographs from social networks and Open Street Map (OSM) and identify the LU type depicted in those images that are extrapolated to regions bounded by street blocks. The contribution is twofold: 1) a hierarchical street-based land division to circumvent uneven spatial distribution of photographs and 2) concise, specialized object bank (OB) for scenic street photographs classification.

The rest of this letter is organized as follows. Section II reviews recent advances of the LU classification. Section III gives details of the study area and data used in this letter. Section IV describes our proposed method. Section V presents our experimental results. Section VI concludes this letter with a summary.

II. Related Work

Considering the low-cost characteristic of the VGI data, it has shown a great potential for geographers to recognize different LU types [5]. The easily accessible, low-cost VGI data have been used in LU and land-cover (LC) classification [8], [9], and the research has made great achievements. Yuan *et al.* [10] adopted a method of natural language processing to combine GPS data and point of interest (POI) to realize urban function zoning of Beijing through establishing a semantic model. Based on bus smart card data and POI data, Long and Shen [11] demonstrated the spatial distribution of urban function zones in Beijing through the construction of discovering function zones via the cluster analysis. Pei *et al.* [12] used mobile phone data to determine the social functions of the land using mobile phone data to characterize LU types and used a semisupervised fuzzy c-means clustering approach to infer the LU types. Rodrigues *et al.* [13] used multisource POIs to make the classification.

In addition to the POI data, photography has been exploited for inferring LU. Estima and Painho [14] evaluated photographs from Flickr for the LU classification for the city of Coimbra and conducted a comparison with Corine LC classes (levels 1 and 2) for continental Portugal. It was found that the uneven spatial distribution of the photographs induces errors and improvement could be achieved by integrating other VGI sources. Leung and Newsam [7] used geotagged photographs from Flickr and Geograph. The method was evaluated on LU maps from university campuses and the overall results suggested the potential of generalizing this approach to areas, where LU maps are unavailable or are outdated.

In contrast to numerical data, e.g., GPS and phone records, using VGI photographs for the LU generation has not been fully studied. This is due to much greater difficulties in processing images to extract the semantic information as well as the sparseness and unevenness of the available photographs.

Manuscript received December 20, 2017; revised March 24, 2018 and May 17, 2018; accepted August 5, 2018. This work was supported by the Key Laboratory of Agricultural Electronic Commerce through the Ministry of Agriculture of China under Grant AEC2018003. *(Corresponding author: Xiaohui Yuan.)*

F. Fang, L. Wang, Y. Liu, and Z. Luo are with the College of Information Engineering, China University of Geosciences, Wuhan 430074, China.

X. Yuan is with the College of Information Engineering, China University of Geosciences, Wuhan 430074, China, and also with the Department of Computer Science and Engineering, University of North Texas, Denton, TX 76203 USA (e-mail: xiaohui.yuan@unt.edu).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2018.2864282

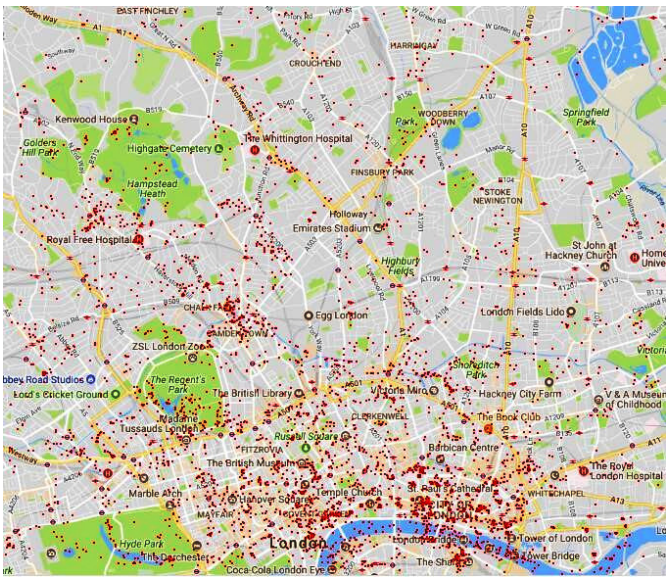


Fig. 1. Spatial distribution of photographs. Each dot marks the location of a photograph and the distribution is extremely uneven. (The background image is obtained from Google Map on March 20, 2018.)

Our hypothesis is that VGI photographs of an urban area allow automatic generation of LU via the data density-driven land division and robust photograph classification algorithms.

III. Study Area and Data Sets

The study area is at the center of the metropolitan London, U.K., with an area of 110 km^2, which is composed of various LU types including residential area, commercial area, water body, open space, and so on. A large volume of the VGI data of London such as photographs is shared via social networks. In this letter, we adopted the geotagged images from the Geograph Britain and Ireland project as our source of geotagged images, which consists of 24 835 images with GPS coordinates. The spatial distribution of these photographs is shown in Fig. 1.

In addition to photographs, the road networks of London are obtained from OSM [15]. The data set contains streets of different classes that are organized according to the street level and width. Street levels include primary highways, primary roads, secondary roads, and small roads (i.e., local and neighborhood streets). The OSM is updated constantly, but for a matured city such as London, changes of road network are minimal.

IV. Method

Our method consists of space division and photograph classification. In the space division, the study area is divided into small regions. The hierarchical street network is used to facilitate the division. In the photograph classification phase, selected features are extracted from each photograph, and a classifier is used to predict the depicted LU type. For a street block with more than one photograph, majority voting is used to make the final decision and extrapolate to the entire block. If no photograph exists in a street block, a large-scale block is used, which is constructed with streets in an upper level in the street hierarchy.

A. Urban Space Division Using Hierarchical Street Networks

Let I denote an urban space and $I(u, v)$ is a geolocation at coordinates u and v. For a photograph p, a pair of coordinates (u, v) is assigned through geotagging, i.e., $p(u, v)$. To decide the LU from a set of photographs, we divide the urban space into parcels, denoted with $\tilde{I}(U, V)$, where U and V are the sets of continuous coordinates. Ideally, each parcel represents a homogeneous LU type, that is,

$$f(\tilde{I}(U, V)) = f(p_i), \quad i \in [1, N] \tag{1}$$

where $f(\cdot)$ is the mapping function that classifies the LU type of a photograph p_i. When inconsistent LU types are identified for the photographs, the dominate LU type is used as the LU type of the parcel $\tilde{I}(U, V)$.

A typical issue of the VGI data is the uneven geospatial density. The spatial distribution of the geotagged photographs is usually discontinuous and varies across the urban space as shown in Fig. 1. In a region with many photographs, it is likely that there is more than one photograph in a parcel. In addition, in a region that is not a tourist attraction or a business, there exist street blocks that contain no photographs. Hence, the division of an urban space needs to be adaptive to the density of the spatial distribution of photographs.

To address this issue, we put streets into three categories: primary, secondary, and tertiary streets. When dividing an urban space, the small-scale parcels, denoted with $\tilde{I}^{(3)}$, are constructed using streets from all categories. If there is no photograph in a small-scale block, a medium-scale parcel, denoted with $\tilde{I}^{(2)}$, is constructed using the secondary and primary streets that enclose the small-scale parcel in question. If again there exists no photograph in the medium-scale block, a large-scale parcel, denoted with $\tilde{I}^{(1)}$, is constructed using the primary streets that enclose the medium-scale parcel. Following this hierarchical urban space division scheme, parcels adapted to the photograph density are generated.

For a parcel $\tilde{I}^{(m)}$ that contains no photograph, its LU type is the same as that of a coarser scale, denoted with $\tilde{I}^{(n)}$ and $n < m$, where $\tilde{I}^{(m)} \subset \tilde{I}^{(n)}$, that is,

$$f(\tilde{I}^{(m)}) = f(\tilde{I}^{(n)}), \quad n < m. \tag{2}$$

When more than one LU type exists in $\tilde{I}^{(n)}$, the LU type for $\tilde{I}^{(m)}$ is decided by the weighted majority from the finer scale parcels in $\tilde{I}^{(n)}$

$$f(\tilde{I}^{(m)}) = \arg\max_k \left\{ \sum_{w_i} w_i f\left(\tilde{I}_i^{(m)}(U, V; k)\right) \right\} \tag{3}$$

where $\tilde{I}_i^{(m)}(U, V;\ k)$ denotes a parcel with the LU type k, and the weight w_i is inverse proportional to the distance to $\tilde{I}^{(m)}$.

B. Photograph Classification Using Object Bank

To correctly classify photographs that contain a variety of mixed ground objects in the complex spatial context, representative image features are crucial. Yet, the variety of photographs posted to the social networks is extraordinary, for instance, a shopping center could have a number of views that differ greatly in scale, perspective, spatial relation to nearby objects, and so on. Hence, low-level image features suffer from consistency and stability. To address the class feature diversity,

TABLE I
LU CATEGORIZATION AND MATCHING SCHEME

Land use	Description	Land-use types in GMESUA data	# of photo
Residential	Multiple unit or multi-family housing (e.g., condos and apartment)	Continuous Urban fabric, Isolated Structures, Discontinuous Dense/Medium Density/Low Density Urban Fabric, Discontinuous very low density urban fabric	698
Commercial	Retail places (e.g., grocery market); Mixed function (e.g., shops and apartments); Financial services (e.g., banks); Service building (e.g., offices)	Industrial, commercial, public, military and private units	1790
Institution	Medical units (e.g., hospitals) Administrative offices (e.g., government facilities) Public service (e.g., transportation terminals)	Industrial, commercial, public, military and private units	1661
Open space	Open land (e.g., parks and sport fields)	Mineral extraction and dump site, green urban area, land without current use, construction sites, forest, sports and leisure facilities, semi-natural and wetland areas, agricultural	339
Water	Rivers, lakes, and ponds	Water	446

we extend the OB [16] by specializing it to the urban scenery through feature selection. A generic OB is obtained by ranking the object according to their frequency of occurrence in data sets including extrasensory perception, LabelMe, ImageNet, and Flickr photograph set.

The key idea of OB presents an image as a collection of semantic-level objects. A generic OB feature collection is constructed by training a number of models for various objects. The initial object responses are obtained by applying filters to an image at various locations and scales. A hierarchical block structure with different granularities is formed according to the spatial pyramid model. In our method, feature selection for urban scenery-specialized OB is performed by evaluating the impact of each feature on the accuracy by removing it from the collection. Many of the features are irrelevant to the classification of LU, e.g., snail and key; hence, the most influential ones are retained for our classification.

C. Land Use of Parcels

For each parcel, we constructed two indexes to identify its LU type: frequency density and category ratio. The frequency density of the type k photographs with respect to the total number of photographs of the type is denoted with F_k, and the ratio of the frequency density of the type k photographs to the density of all the types in a parcel is denoted with C_k and computed as follows:

$$C_k = \frac{F_k}{\sum_{k=1}^{n_k} F_k}, \quad \text{and } F_k = \frac{n_k}{N_k} \tag{4}$$

where n_k denotes the number of photographs of type k in a parcel, and N_k denotes the total number of the type k photographs.

Following (4), we calculate the frequency density and category ratio of each parcel. When the proportion of a certain type of photograph in the unit accounts for 50% or more, it is determined as a parcel of single-used land and the type is the same as that of the photographs. When all types of photograph points in a unit are less than 50%, it is considered as a parcel of multiuse land. The type of mixture depends on the top two dominant types of the photographs in the parcel. When the parcel contains no photographs (i.e., the category ratio is zero), it is considered as an unclassified area.

TABLE II
AVERAGE ACCURACY USING DIFFERENT NUMBER OF FEATURES

Num. of Features	208	150	99	80	60	40
Accuracy	86.1	86.1	83.3	80.6	77.8	75

V. EXPERIMENTAL RESULTS

A. Experimental Data and Settings

To evaluate the proposed approach, the results are compared with the Global Monitoring for Environment and Security Urban Atlas (GMESUA) data, which are derived from the Earth observation data supported by topographic maps. The LU types in GMESUA data are very detailed and many differ only in degrees, e.g., discontinuous dense/medium-density/low-density urban fabric. To be consistent, we put photographs into five categories that match the GMESUA LU types as listed in Table I.

The street networks from OSM data were categorized into three levels according to their width: primary streets that are greater than and equal to 40 m, secondary streets that are lesser than 40 m and greater than and equal to 12 m, and tertiary streets that are less than 12 m. In our experiments, all photographs were resized to 128×128. Support vector machine was used for the classification with a radial basis function kernel. A fivefold cross validation was performed in our studies.

To derive a specialized OB for urban photographs, we evaluated the impact of objects to the classification accuracy. Table II lists the classification accuracy with respect to the number of features. In each case, the less significant features are removed. With all features, the average accuracy is 86.1%.[1] As the number of features decreases, the accuracy decreases. In the rest of our experiments, 150 objects are used in photograph classification.

B. Photograph Classification Analysis

The reliability of photographs inevitably affects the performance of LU classification. Erroneous LU labels arise from the unaligned dominant foreground objects depicted in the

[1]Note that this is different from LU accuracy. It deals only with regions that have photographs, and extrapolation to estimate the regions without photographs is not practiced.

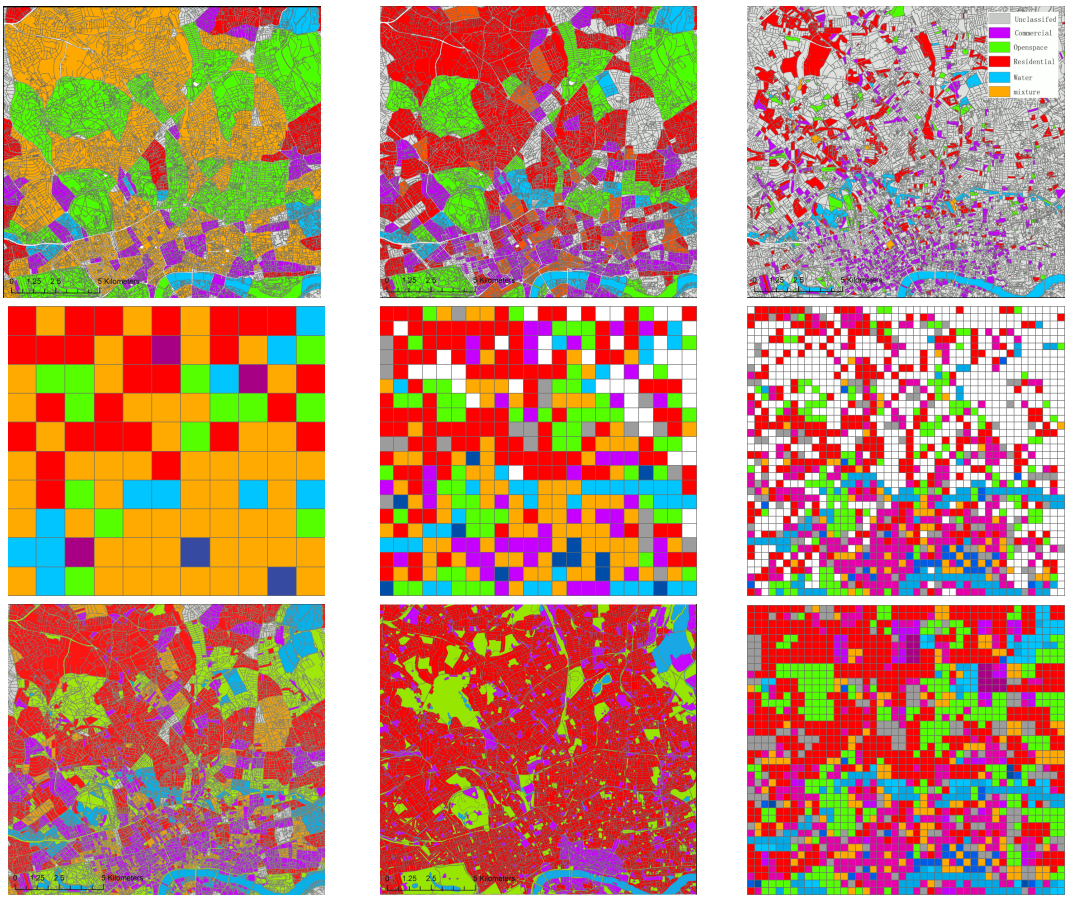


Fig. 2. Comparative view of the LU results. (Top row) Results of the proposed method on different parcel scales (coarse to fine from left to right). (Middle row) Results from grid-based method [7] based on three gridding scales (coarse to fine from left to right). (Bottom row from left to right) Integrated result of our method, GMESUA data, and integrated result of method in [7].

images. For example, a fairly large portion of the foreground of a photograph shows part of a residential building in front of a park, and because of the presence of a building, the classification, which typically underweighs objects in the background, decides the scenery that represents a residential area. To better understand the reliability of photograph classification, we conduct experiments and generate a confusion matrix as given in Table III. We obtained an overall accuracy of 76.5% for the five classes: residential, commercial, institution, open space, and water body. Institution type had the highest producer's accuracy (85%) in built-up areas, whereas an open space had a relatively lower rate of 56.6%. The user's accuracy essentially tells the reliability of the classification to the actual LU types. In our experiments, the user's accuracy of all except open space is above 70% and the institution type yields the best reliability.

The three-level LU mapping results are shown in Fig. 2 (Top row). The mixture type (in orange color) represents the structures for both commercial and residential usages. From the resulted LU map, over 60% areas in London are residential. The residential areas are in the northern part of London, while the commercial areas are mostly along both sides of the Thames river. Fig. 2 (Middle row) depicts the results from the grid-based method [7]. The area is gridded into three scales and the integrated result is shown on the right of Fig. 2 (Bottom row). The result of our method and the LU map of GMESUA data are shown in Fig. 2 (Bottom row). In comparison with the GMESUA data, the results of our method are mostly consistent with the LU map of GMESUA. The overall accuracy of the grid-based method is 49.14%.

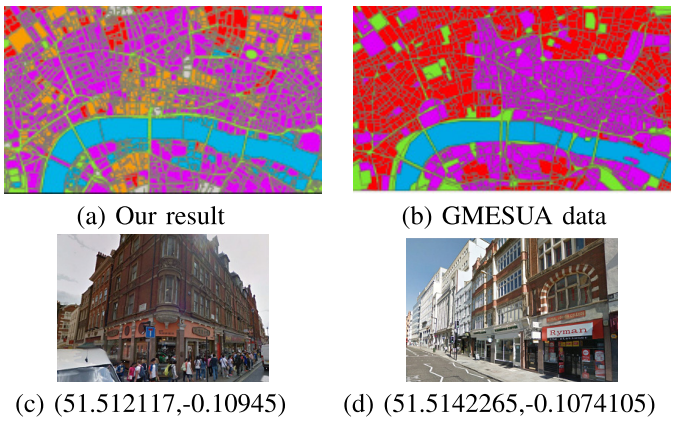
TABLE III
ACCURACY OF PHOTOGRAPH CLASSIFICATION. PA: PRODUCER'S ACCURACY AND UA: USER'S ACCURACY

	Inst.	Comm.	Resi.	Open Sp.	Water	PA
Inst.	352	46	8	3	5	85.0%
Comm.	52	374	13	3	4	83.9%
Resi.	20	56	93	2	2	53.8%
Open Sp.	12	8	7	47	9	56.6%
Water	4	12	5	17	72	65.5%
UA	80.0%	75.4%	73.8%	65.3%	78.3%	76.5%

C. Comparison Study

We selected two representative regions to compare with the GMESUA data and validated by street views: the Thames river bank and the northwest corner of the study area.

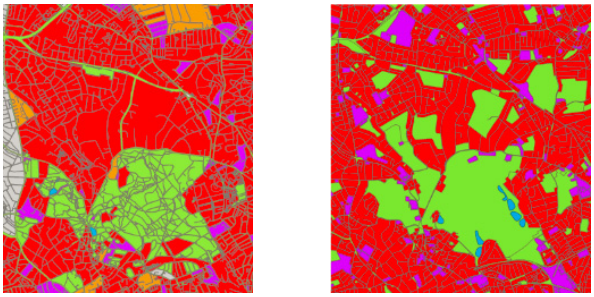
1) Thames River Bank: As shown in Fig. 3(a), a majority of this area is for business functions and a smaller portion is for multifunctional structures as well as residences. However, Fig. 3(b) illustrates two LU types, commercial and residential usages, and they distribute along the river. There exist disagreements between our results and the



(a) Our result (b) GMESUA data

(c) (51.512117,-0.10945) (d) (51.5142265,-0.1074105)

Fig. 3. (a) and (b) LU result and GMESUA LU map along the Thames river bank. (c) and (d) Google street view photographs. The numbers are the GPS coordinates of these photographs.



(a) (b)

Fig. 4. Comparison of the LU map of the northwest corner of the study area. (a) Our result. (b) GMESUA data.

LU map from GMESUA. To understand the disagreements, we pulled the photographs from Google street view as shown in Fig. 3(c) and (d). It is clear that there are many business shops as well as mixtures of commercial and residential usages. Such multifunctional structures are mostly of the same form: the lower floors are for a commercial purpose and the upper floors are for residence. Hence, it is evident that our method provides a more accurate description of the true LU in contrast to the LU presented by GMESUA.

2) Hampstead Heath: Fig. 4 shows the LU map produced by our method and the GMESUA data. The LU map in the GMESUA data in Fig. 4(b) shows large but disconnected areas of an open space that is colored with green; whereas our results as shown in Fig. 4(a) shows relatively limited open space region. A noticeable sized region in our results is marked with gray color, which indicates unclassified LU. In addition, there exist false identified LU types, such as water body depicted in blue in Fig. 4(b). In comparison with Fig. 1, it is easy to see that there are much fewer photographs for the top half of the study area. A very few numbers of photographs provide weak support to the LU classification and make spatial extrapolation error prone. Our proposed method is able to identify LU types by leveraging the urban land unit generated from the hierarchical road networks.

D. Efficiency

Our method is implemented with MATLAB R2016b in a PC with Intel Xeon E5-2650, 16-GB memory, and Ubuntu system. In our evaluation of time, 3000 images are randomly selected and used as the training images, and 1000 images are used as the testing images. The average training time per image is 0.72 s and the average time to classify an image (i.e., testing) is 0.53 s. For a typical urban area such as London, which accumulates about 5000 images (at the time of this letter), our system takes less than an hour to complete the LU classification.

VI. Conclusion

This letter presents a new approach to extract the image semantic information for LU studies. The proposed method employs the VGI data, which make this approach economically advantageous and efficient without the need for fieldwork. The proposed method eases the process of updating LU maps. For future work, integrating VGI data of different modalities and sources such as Wikimedia, POIs, or remote sensing imagery could fill the gap of missing data and lack of examples.

References

[1] J. P. Evans and R. Geerken, "Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure," *Remote Sens. Environ.*, vol. 105, no. 1, pp. 1–8, 2006.

[2] R. L. Cifaldi, J. D. Allan, J. D. Duh, and D. G. Brown, "Spatial patterns in land cover of exurbanizing watersheds in southeastern Michigan," *Landscape Urban Planning*, vol. 66, no. 2, pp. 107–123, 2004.

[3] Z. Masoomi, M. S. Mesgari, and M. Hamrah, "Allocation of urban land uses by multi-objective particle swarm optimization algorithm," *Int. J. Geograph. Inf. Sci.*, vol. 27, no. 3, pp. 542–566, 2013.

[4] X. Yuan and V. Sarma, "Automatic urban water-body detection and segmentation from sparse ALSM data via spatially constrained model-driven clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 73–77, Jan. 2011.

[5] M. F. Goodchild, "Citizens as sensors: The world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, 2007.

[6] T. Rattenbury and M. Naaman, "Methods for extracting place semantics from Flickr tags," *ACM Trans. Web*, vol. 3, no. 1, p. 1, 2009.

[7] D. Leung and S. Newsam, "Land cover classification using geo-referenced photos," *Multimedia Tools Appl.*, vol. 74, no. 24, pp. 11741–11761, Dec. 2015.

[8] V. Antoniou *et al.*, "Investigating the feasibility of geo-tagged photographs as sources of land cover input data," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 5, p. 64, 2016.

[9] C. C. Fonte, L. Bastin, L. See, G. Foody, and F. Lupia, "Usability of VGI for validation of land cover maps," *Int. J. Geograph. Inf. Sci.*, vol. 29, no. 7, pp. 1269–1291, 2015.

[10] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 186–194.

[11] Y. Long and Z. Shen, "Discovering functional zones using bus smart card data and points of interest in Beijing," in *Geospatial Analysis to Support Urban Planning in Beijing*. Cham, Switzerland: Springer, 2015, pp. 193–217.

[12] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou, "A new insight into land use classification based on aggregated mobile phone data," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 9, pp. 1988–2007, 2014.

[13] F. Rodrigues, F. C. Pereira, A. Alves, S. Jiang, and J. Ferreira, "Automatic classification of points-of-interest for land-use analysis," in *Proc. 4th Int. Conf. Adv. Geograph. Inf. Syst., Appl., Services*, 2012, pp. 41–49.

[14] J. Estima and M. Painho, "Photo based volunteered geographic information initiatives: A comparative study of their suitability for helping quality control of corine land cover," *Int. J. Agricult. Environ. Inf. Syst.*, vol. 5, no. 3, pp. 73–89, 2014.

[15] J. Arsanjani, A. Zipf, P. Mooney, and M. Helbich, "An introduction to OpenStreetMap in geographic information science: Experiences, research, and applications," in *OpenStreetMap in GIScience*. Cham, Switzerland: Springer, 2015.

[16] L.-J. Li, H. Su, L. Fei-Fei, and E. P. Xing, "Object bank: A high-level image representation for scene classification & semantic feature sparsification," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2010, pp. 1378–1386.