Land Use Classification in Construction Areas Based on Volunteered Geographic Information

Chenru Chen[1], Zhenbo Du[1], Dehai Zhu[1*], Chao Zhang[1], and Jianyu Yang[1]
College of Information and Electrical Engineering, China Agricultural University
No. 17 Tsing Hua East Road, HaiDian District, Beijing, China
*Corresponding Author: zhudehai@cau.edu.cn

***Abstract*—In the process of land use management, the relative low construction land use efficiency may lead to more other types of land into construction land, and thus will also affect the amount of cultivated land, while reducing the ecological land as well. According to the current land use classification standard (GB / T 21010-2007), the construction land can be divided into business service land, residential land, public management and public service land, and storage land for industry, etc.. To enhance the utilization efficiency of construction land, this research aims to develop a better way to divide the space distribution of different current land use types, which also provide both database and clues to concentrated land-use planning and monitoring. In this study, a new methodology (hierarchical grading classification method) will be proposed to solve the problems existing in the traditional division methods of construction land, and the experimental results will deliver a series of meaningful interpretations across discipline. To demonstrate this, taking the Fifth Ring of Hai Dian district, Beijing city as the research area, a variety of volunteer geographic information is determined, which including the Open Street Map (OSM), Points of Interest (POI), blogging sign data and Panoramio photos, etc.. Firstly, the Open Street Map road data is used as a block boundary to divide the construction land into the different hierarchy of land parcels. Point of interest, its essence is the abstract expression of geographical entities. Since the Point of Interest and the divided land parcels share the consistent feature attributes, it is possible to use the different grade of POI to assign attribute to the different hierarchy of land parcels, and then the final results of the multi-layers will be combined together to do analysis (overlay, merge) to determine the construction land use type in this region. Finally, the confusion matrix is generated to compare the results among the Google street view, fieldwork and urban planning map. The accuracy rate of commercial and business facilities, industrial and warehouse, residential, administration and public services, street and transportation, and other construction land are 94.7%, 69.2%, 81.4%, 75.0%, 96.7% and 74.7% respectively. Furthermore, the kappa indices of classification is 0.83, showing that in this study, both the adopted data and the newly proposed method used in the process of classification of construction land are feasible, and the new method will have significant impact on the process of division construction land.**

Keywords—volunteered geographic information; classification of construction land use; land use management; big data

I. INTRODUCTION

VGI (Volunteered Geographic Information) is an open sensor that everyone may achieve data collection, which means, in the era of Web 2.0, normal individuals is gradually shifting from the traditional data user to the data producer and disseminator [1, 2]. Whilst, an explosion of crowdsourcing data and crowdsourcing platforms rapidly rushed into the view of scientists [3].

The objects of land cover classification including water bodies, vegetation, and the construction land. With the rapidly development of image technology, the classification method is becoming increasingly mature [4, 5, 6], while the different land use in build-up areas still remains a problem which cannot simply depends on spectral characteristics. At present, VGI database has been widely used in the world. There is a significant tendency shows that many studies carried out to apply VGI in detailed land use classification [7]. Rodrigues et al. applied Points of Interest (POI) to evaluate and analyze land-use situation from the perspective of urban planning [8]. Jamal et al. simply represented Vienna Land Use Map based on the Open Street Map (OSM) database [9]. While in China, Ying et al. achieved functional division according to the types and numbers of the point provided by the public sign of Sina Weibo users [10]. Though the VGI studies have huge improvements over traditional methods for the classification of construction land [11, 12, 13], it's still facing many challenges: 1) the lack of Volunteer Geographic Information data source may cause the low accuracy problem in the city with relative high complex construction land-use; 2) the actual feature covers are represented by different types of points, which may not meet the actual area of the land covers, and in the process of determining the land type, the degree of importance may vary in different situation. Thus, this study will propose a new methodology (hierarchical grading classification method) to solve the problems existing in the traditional division methods of construction land, and the experimental results will deliver a series of meaningful interpretations across discipline. This method mainly use a collection of VGI data, "break" the urban construction land (considered as the whole before), and further to finish the re-fine classification.

II. METHOD

In this research, taking the Fifth Ring of Hai Dian district, Beijing city as the research area, a variety of volunteer geographic information is determined, which including the Open Street Map (OSM), Points of Interest (POI), blogging sign data and Panoramio photos, etc.. The first step is data cleaning. When finished, the construction land will be divided into commercial and business facilities, industrial and warehouse, residential, administration and public services, street and transportation, and other construction land based on the hierarchical grading classification method. Finally, a comparison between the division results, the actual survey results, and the land utilization maps should be developed to deliver the final evaluation. Below the *Fig. 1* illustrates the brief flowchart of the whole technical process.

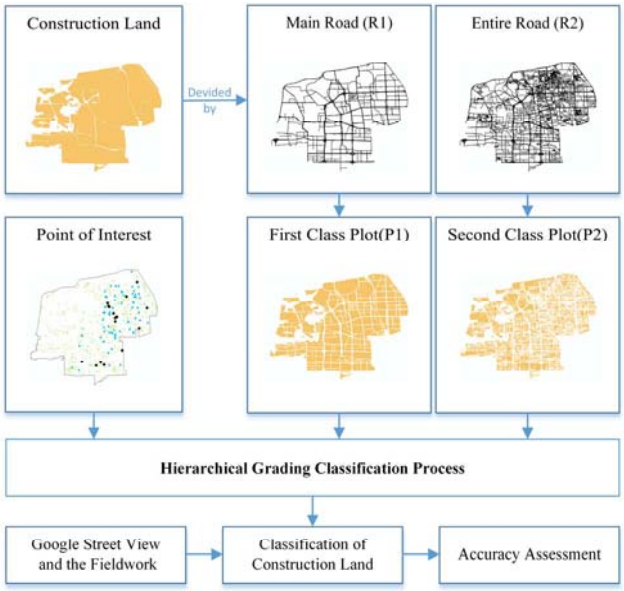


Fig. 1. the brief flowchart of the whole technical process

A. Data Preprocessing

This process include two parts: OSM data management and POI data management. To the OSM road data, firstly, the road length should be extended 20 meters to eliminate the deviation of the OSM dataset that may from the artificial data collection process. Next, the roads where cannot reach to the link roads, are supposed to be modified. When these two procedures finished, the connectivity of the whole road net will be achieved. While for the POI dataset, the most critical thing is to confirm the accuracy of the data property value. In this process, though analyzing the character feature class of every property, a corresponding regular expression should be identified to match the correct text and digital information.

B. Land Block Division and POI Classification

Since in this research road data is all extracted from the Open Street Map, according to the property value of the OSM layer, the road dataset will be divided into main road (R1) and the entire road (R2) [14], and these two classes become as the basis for division of land. For different levels of roads, a different width buffer is calculated to split construction land. Plot origin from the main road is determined as the first class plot (P1), and the second class plot (P2) is extracted from the entire road.

The POI data divide into four levels based on the different data characteristics. The first level (L1) mainly contains the point with integrated multi-buildings like most of the colleges and universities, airports, etc.. Inside these buildings, always exist relative more roads, taking almost the whole block that may account for a large area of the construction land. The second level (L2) POI data combined with series of concentrated buildings. The difference between L1 with L2 is that the L2 points usually have less (even none) roads, and because of this, the construction purpose of L2 is not that complicated when comparing with L1, for example, the residential areas, industrial parks and other categories. As for the level three (L3), the points of interest feature should occupy the whole building. The typical representatives are large companies, star-rating hotels, libraries, etc.. When a plurality of points of interest share one building such as shopping malls and restaurants, or a bungalow with an area not exceeding 30m2, these points will be classified as the forth Level (L4). At this point, the grading work of POI dataset is completed.

C. Hierarchical Grading Classification Process

The entire hierarchical grading classification process can be illustrated in the following steps:

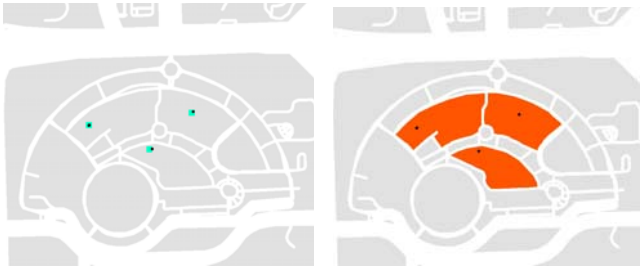
- Assign the type value of the POI in L1 layer to the first land parcel;
- Repeat the first step, assign the type value of the POI in L2 and L3 layer to the second land parcel;
- Taking into account the characteristics of the L4 points of interest, these points should be retrieved in the Atlas indoor-mall database. If the point exist, that means the type value assignment process described in (1) and (2) is completed, then this point should be excluded from L4 layer; by contrast, when this point do not exist, it can be counted as a single feature, and this kind of land parcel will be defined depending on the average size land area.
- Finished the type value assignment process, to get the construction land type, different levels of classification results should be merged together and reunion according to the Situation Land Use Classification Regulation. It can be clearly seen that the construction land includes commercial and business facilities, industrial and warehouse, residential, administration and public services, street and transportation, etc..

It is need to be explicitly notified here that multiple hierarchical grading classification mainly taking into account the problems between L1 POI layer and L4 POI layer during the classification process. As previously stated, the points in L1 layer usually contain more roads, or/and taking a large area, if they are managed as the same way to other categories, it will increase the risk of missing plots（as shown in *fig. 2*）. While

L4 layer represents the points of interest with small area. For these points, the type value assignment process cannot operate directly because it will lead an end that the classification result may be far larger than the actual area (shown in *Fig. 3*).

(a) (b)

Fig. 2. Discription of the Missing Plots Situation (*(a) presented the right classification results, and (b) the red area showed the missing plots*)



(a) (b)

Fig. 3. Discription of the Excess Plots Situation (*(a) presented the right classification results, and (b) the red area showed the excess plots*)

Additionally, there still has another situation that two or more types of POI in the same grade show up in one land parcel unit. The presence of this setting generally occurred in the L2 and L3 POI layers' type value assignment process. Thus, as for POI in the same level, it can be divided based on the proportion of different types of plot points, and then classified.

III. Experiment

A. Determine the Parcel Type

According to the division process referred in part II, the analysis (overlay, merge) results distinctly shown in *Fig. 4* that the construction land can be divided into commercial and business facilities (05), industrial and warehouse (06), residential (07), administration and public services (08), street and transportation (10), and other construction land. This result is a combination of volunteer geographic information, which including the Open Street Map (OSM), Points of Interest (POI), blogging sign data and Panoramio photos, and it will deliver a series of meaningful interpretations across discipline.

B. Result and Accuracy Assessment

The urban land use planning map, Google street view and the fieldwork are treated as the true value, compared to the classification results, then, the confusion matrix is generated to evaluate this outcome. Table 1 helps to exemplify the evaluation result. It can be seen that the accuracy rate of commercial and business facilities, industrial and warehouse, residential, administration and public services, street and transportation, and other construction land are 94.7%, 69.2%, 81.4%, 75.0%, 96.7% and 74.7% respectively. Furthermore, the kappa indices of classification is 0.83, showing that in this study, both the adopted data and the newly proposed method used in the process of classification of construction land are feasible, and the new method will have significant impact on the process of division construction land.

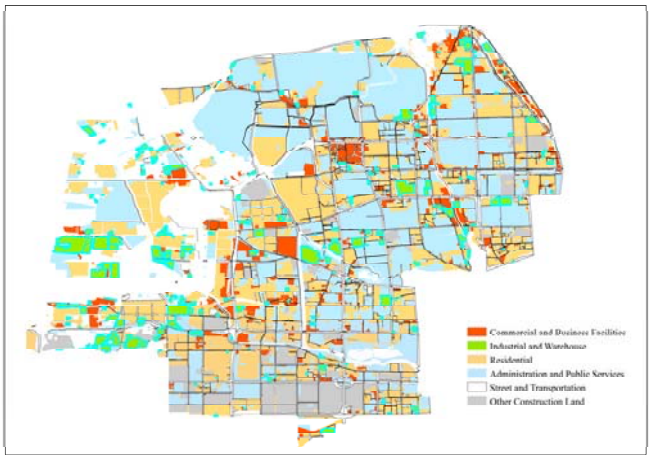


Fig. 4. The Final Classification Results of Construction Land

TABLE I. CONFUSION MATRIX OF CONSTRUCTION LAND USE CATEGORIES ACCURACY ASSESSMENT

		Generated Map						
True Value	***classes***	***05***	***06***	***07***	***08***	***10***	***other***	***total***
	05	881	0	20	24	0	5	930
	06	0	220	31	66	0	1	318
	07	0	0	1508	345	0	0	1853
	08	284	97	7	1165	0	0	1553
	09	33	0	0	6	1167	0	1206
	other	0	0	37	41	0	230	380
	total	1203	323	1610	1655	1177	236	6168

IV. Discussion

This research have developed a more efficient way to divide the space distribution of d0ifferent current land use types, which solved the traditional artificial techniques' defects such as high labor costs, inefficient and time-consuming process, and low automation level, etc.. However, there still leave some problems in the accuracy evaluation process, especially in the acquisition of data source. For example, large flows of people always gather in downtown area, which may lead to a high usage rate of smartphones, and for this reason, the volunteered geographic information in downtown show a

more comprehensive circumstance to get better classification results. Hence, it can be inferred that the farther from the center of city, the larger classification error deviation there will be, which indicate that in the future studies, more data source is required to ensure the accuracy of the division of the construction land.

Overall, with the increasingly growing number of volunteered geographic data and other types of big data, the efficiency of classification is becoming progressively mature. The update rate, accuracy and classification period of data see an apparently substantial growth. Whether from the perspective of country utilization or city planning, or the perspective of construction land classification VGI (Volunteered Geographic Information) will received more attention and be used as the basis of dataset. The data presented herein is also obtained as a clue to the concentration of land planning or monitoring, and in the following days, it will play a major role in land management.

REFERENCES

[1] Goodchild, Michael F. "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0." International Journal of Spatial Data Infrastructures Research 2(2007):24--32.

[2] Goodchild, Michael F., and L. Li. "Assuring the quality of volunteered geographic information." Spatial Statistics 1(2012):110-120.

[3] Goodchild, Michael F. "Commentary: whither VGI?." Geojournal 72.3(2008):239-244.

[4] Foley, Jonathan A., et al. "Global consequences of land use." Science 309.5734(2005):570-4.

[5] Huang, C., L. S. Davis, and J. R. G. Townshend. "An assessment of support vector machines for land cover classification." International Journal of Remote Sensing 23.4(2002):725-749.

[6] Herold, M. "Spatial Metrics and Image Texture for Mapping Urban Land Use." Photogrammetric Engineering & Remote Sensing 69.9(2003):991-1001.

[7] Fritz, S., et al. "Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover." Remote Sensing 1.3(2009):345-354.

[8] Rodrigues, F., et al. "Automatic Classification of Points-of-Interest for Land-use Analysis." GEOProcessing 2012 : The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services 2015:41-49.

[9] Jokar Arsanjani, Jamal, et al. "Toward mapping land-use patterns from volunteered geographic information." International Journal of Geographical Information Science 27.12(2013):2264-2278.

[10] Long, Ying, and X. Liu. "Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest." Journal of International Medical Research 38.38(2015):498-510.

[11] Newsam, S. "Crowdsourcing What Is Where: Community-Contributed Photos as Volunteered Geographic Information." IEEE Multimedia 17.4(2010):36-45.

[12] Hardy, Darren, J. Frew, and M. Goodchild. "Volunteered geographic information production as a spatial process." International Journal of Geographical Information Science 26.7(2012):1191-1212.

[13] Sarah Elwood, Michael F. Goodchild, and Daniel Z. Sui. "Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice." Annals of the Association of American Geographers 102.3(2012):571-590.

[14] Kuang, Wenhui, S. Zhang, and J. Liu. "Methodology for classifying and detecting intra-urban land use change: A case study of Changchun city during the last 100 years." Journal of Remote Sensing 14.2(2010):345-355.