URBAN FUNCTION ZONING USING GEOTAGGED PHOTOS AND OPENSTREETMAP

Lu Wang[1], Fang Fang[1,], Xiaohui Yuan[1, 2], Zhongwen Luo[1], Yuanyuan Liu[1], Bo Wan[1], Yishi Zhao[1]*

[1]College of Information Engineering, China University of Geosciences, Wuhan, 430074, China
[2]Department of Computer Science and Engineering, University of North Texas, Denton, 76203, USA

ABSTRACT

Urban function zoning is of great importance for urban structure optimization, urban resource allocation, and urban development planning. Since citizens usually act as a network of motion sensors of the city, their activities could reflect the environment around them. We considered taking advantage of VGI data to classify urban function zones. In this paper, we proposed a framework for automated urban function zoning which is based on VGI geo-tagged photos and OpenStreetMap (OSM) data. Through combining the high-level image features of geo-tagged photos with the road network data, we obtained the functional zoning map of the study area. The experiment result shows the effectiveness of the framework we proposed.

Index Terms— VGI, geo-tagged photos, image scene classification, urban function zoning

1. INTRODUCTION

During the course of urban development, a variety of function zones such as commercial zones, residential zones, industry zones and some mixture function zones, gradually formed to meet peoples' needs. In order to grasp the urban spatial structure, planners make divisions of urban function zones. Besides, since cities are complex and dynamic [1], understanding the structure and function zone of cities can promote the rational distribution of space and provide strategies for a city's sustainable development. What's more, further study how people live, work, leisure and tourism in a complex city system and develop the changes of these spatial activities helps to understand the interaction of human's activities and urban space. Therefore, develop a new method to realize the automation of city function zoning is of strong practical significance.

In recent years, with the development of Web 2.0 applications, a lot of networks such as micro-blog and other new internet model appears. They continue to generate a large amount of data which attracted the attention of more and more scholars to dig geographic information from these easily accessible social media data for urban function zoning[2],[3]. People could obtain geo-tagged photos through the device with GPS and share them on social media platforms, such as Flickr, Facebook, etc. Information from different users, especially information with geo-tags, has become an important source for geographical object study and feature description[4]. With the help of the image information (name, description, scene feature, location, etc.), we can dig out the function of urban zones Inspired by this idea, this paper employs VGI geotagged data [5] to achieve urban function zoning.

2. Data and Method

London is the famous tourist attraction of the world and the visitors provides abundant photos data. It is the capital and the largest city of England, there are more volunteers contribute to open block data, e.g. OSM. Therefore, London is adopted as the study area.

2.1. Data

The photos and their XML files we use were downloaded from Geograph by the API offered by the website. All the images from this website contain artificial discrimination classify information, photo caption, coordinate of photographer and coordinates, address and accuracy of the shooting feature. So the photos from it are very suitable [6] for our research.

We use road networks to divide the study area. Such division could reflect the urban planning more reasonable than grid division. The data from OSM can roughly be divided into 3 types and they are organized by several layers. But only line data layer data were used for study area division. These data are credible.

The ground truth we used was downloaded from Global Monitoring for Environment and Security (GMES), due to this program is in real time and reliability, while the geo-tagged images from Geography are also continuous updated.

2.2. Object Bank (OB)

OB [7] is a high-level image representation task that describes each natural image by the objects present, encodes semantic and spatial information of the objects within an image. It is an effective solution that reflects the way our brains recognize a scene. Several experiments [8] have demonstrated that OB representation carries rich semantic level image information, which is more powerful on scene classification tasks than many others popular methods. The OB representation we have used employs 177 pre-trained 2-view Deformable Part Model (DPM) [9], which is widely used for describing multi-view objects, as the object detector. We conducted experiments to compare the classification accuracy of some images features on the same dataset as shown in Figure 1.

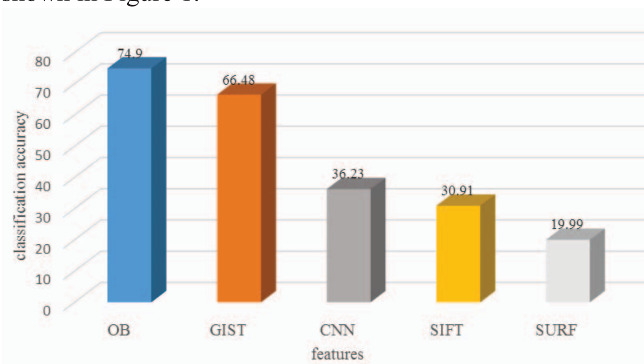

Figure 1 Comparison between different image features on the same dataset

2.3. Framework of the experiment

First, preprocess origin image data from the Geograph to divide them into 6 categories by their tag information. Second, extract the OB feature from the preprocessed images. Then conduct SVM to train and predict these features to divide into 6 classes which are water, commercial, institution, church inertial, open space and residential. In addition, we use road networks to split the research area into regions. Our method maps all the classified points onto the split regions to decide which kind of function zone that every region belongs to. Figure 2 shows the framework of our proposed method.

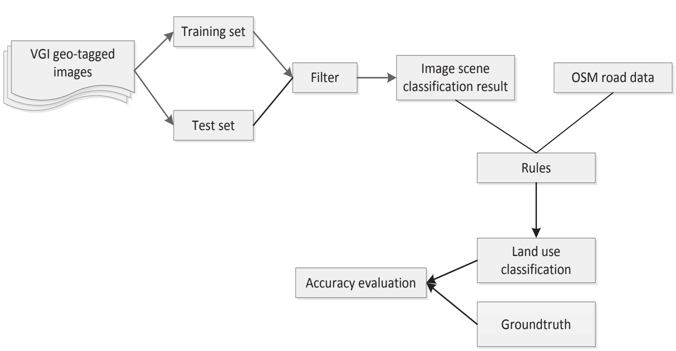

Figure 2 Framework of the experiment

3. EXPERIMENTS AND RESULTS

3.1. Data preprocessing

3.1.1. Photos data preprocess

The photos we have downloaded contain all kinds of features and they belong to many categories, but not all of them are useful. For instance, some categories may have only one or two photos and some are close shooting images. Each photo we have obtained bears a corresponding XML tag and the tag contains many information, like geospatial, temporal and contextual information.

We extract tag information of all these images by parsing their XML files. After obtaining their tags, we group their tags into 6 categories. We abandoned these photos that belong to none of the 6 categories. Then we select 75% of the categories photos as training set and the other as a testing set.

3.1.2. Road networks preprocess

Because the coordinate of the road network is different from the images, we have to unify the coordinate system to make sure images are at the correct location of the study area. Since the road network we obtained is line data, we need to convert them into polygon data. For this, our solution approach is to make a buffer with different widths for each road according to their road grade attribute. To divide urban function areas in multiscale, we get several different levels of the road network. In our experiments, the road network has been divided into three levels. The first level is the largest segmentation scale, which only contains the main road of London. The second level road network contains all roads except 'residential' and the roads whose priority is lower than 'residential'. The remaining level contains whole roads of the study area and it is of the minimum segmentation scale.

3.2. Image feature extraction and classification

During the feature extraction step, we filter an input image in 6 scales, then we got *NumObjec * NumScales * 2* (each view of DPM has one response map) response maps. Our method encodes the response map with a 3-level (L0, L1, L2) pyramid pooling strategy. Hence, we have a *NumObject * NumScales * 2 * (L0+L1+L2)* dimensions in an OB presentation for each image.

Before training with the image features, we have to format the input features according to the input data requirements at first. Our method scales both training and testing data to a certain range. This operation is of great importance in feature classification. Then, implement a 10-folder cross-validation to train a 6-class classifier. The classification result is shown in Table 1.

TABLE 1 Photos classification result

	Church inertial	Commercial	Institution	Open Space	Residential	Water	Producer Accuracy
Church inertial	**150**	8	7	0	0	3	89.3%
Commercial	7	**374**	45	3	13	4	83.7%
Institution	14	38	**181**	3	8	2	73.6%
Open space	0	8	12	**47**	7	9	56.6%
Residential	0	56	20	2	**93**	2	53.8%
Water	0	12	4	17	5	**72**	65.5%
User Accuracy	87.7%	75.4%	67.3%	65.3%	73.8%	78.3%	**74.9%**

3.3. Determination of the urban function zones

Each image can be regard as a single geo-tagged point. When the images are classified, they are mapped to urban function zones.

We locate classified geo-tagged images on the segmented study area and remove the photos located in road buffer. Then, a statistic the number of images and categories of them in each function zone on each scale. For each zone, the type of its function is determined as follows:

$$\mathrm{P} = \frac{\max(c_i)}{2 * \sum_{i=1}^{n} c_i} * n \qquad (1)$$

where n is the count of the type of photos located in a function zone and c_i the number of one kind of photo located in this function zone. When P >= 1, the type of this zone is the same with *max(c_i)*, otherwise, it is a mixture function zone. If there is a function zone in which no photo located, we name it “Unclassified”.

To visualize the result, we made a superposition of experiment results in a different layer. The leveled results are showed in Figure 3. In the superposition process, the function of a region depends on the function of the region in tiniest scale.

We evaluate the classification accuracy by comparing our result with the ground truth. When the function of an area is the same with ground truth, we think this area is correctly classified; otherwise, it is wrong. The collation map is shown in Figure 4. And the average classification accuracy is 56.97%.

3.4. Error analysis

3.4.1. Mixture urban function zone

There are many kinds of mixture function zone types in the urban area, such as the mixture of commercial and residential, commercial and entertainment. We usually are unable to assign a distinct function type to such an urban zone. So mixture urban function zone must be taken into consideration. Since citizens are sensors of their surrounding environment, the photos they have taken could reflect the situation around them more realistically than most of the remote sensing image. In views of this, our experiment result could well reflect urban function zone type. But the ground-truth remote sensing image we have got contains no mixture urban function zone, therefore, the classification accuracy is affected.

3.4.2. Photo collection preference

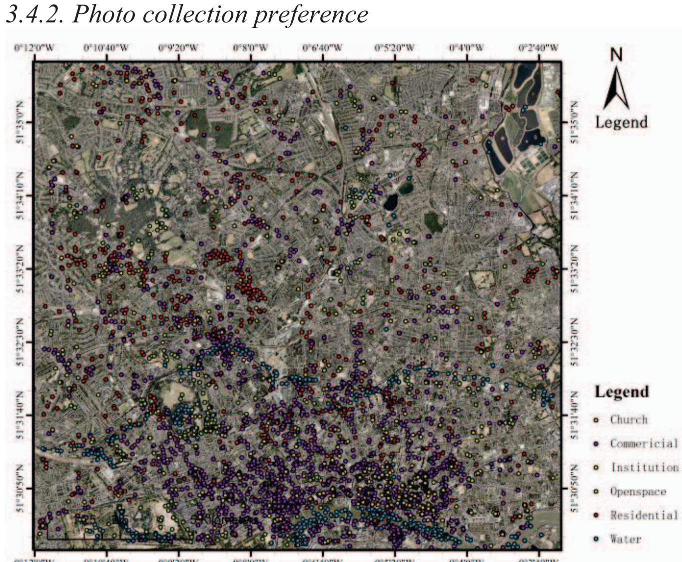

Figure 5 Spatial distributions of VGI photos

The photos and their extra information people sharing on websites can tell the behavior and location of the photographer.

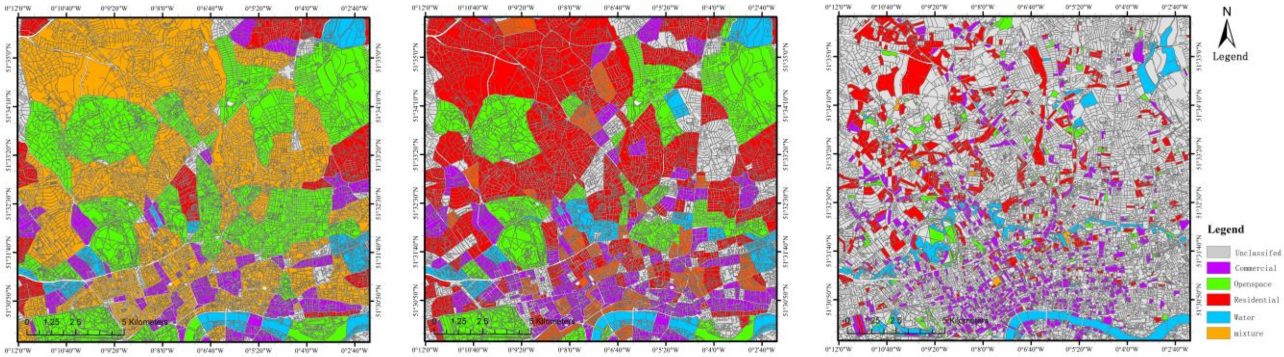

Figure 3 Results of urban function zone

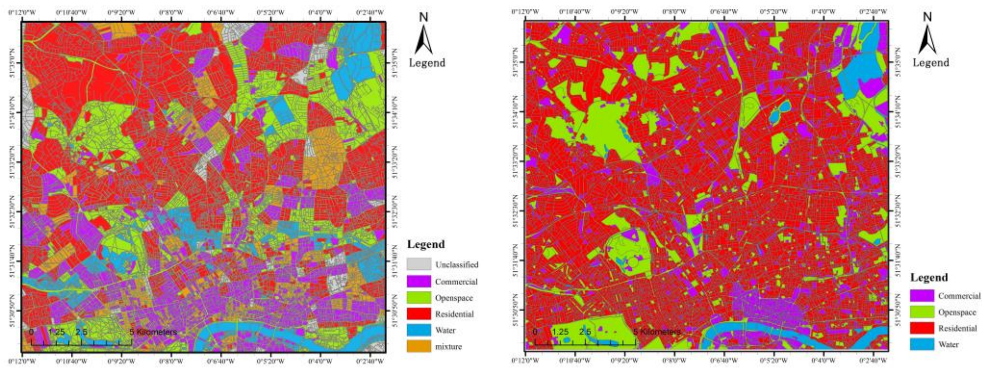

Figure 4 Comparison between experiment result (left) and ground truth (right)

But although these human activities can reflect the function of urban zones to a certain extent, they couldn't cover every corner of a city. People usually like sharing memorable photos and many of their shared photos are food, people or characteristic buildings. Besides, most people like move around bustling places, so the photos people sharing contain more interesting places and fewer residential space or some relatively remote place. Figure 5 illustrates the spatial distribution of the photos.

4. CONCLUSIONS

The main contribution of this paper is a framework for automated urban function zoning. This framework takes advantage of social media geo-tagged photos uploaded by users and image annotations. Using these VGI data could greatly reduce the sampling cost and human efforts of classifying remote sensing images, and can reduce the subjectivity and shorten the time. Besides, the VGI data reflect the actual situation more objective.

ACKNOWLEDGMENTS

Thanks to OSM, Geograph and GMES for providing data. This work is supported by the National Key Research and Development Program of China (No.2016YFB0502304).

REFERENCES

[1] Suteanu, C., *Complexity, Cognition and the City by Juval Portugali.* Journal of Regional Science, 2012. 52(3): p. 506–507.

[2] Long, Y. and X. Liu, *Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest.* Environment and Planning B: Planning and Design, 2016. **43**(2): p. 498-510.

[3] Long, Y. and Z. Shen, *Discovering Functional Zones Using Bus Smart Card Data and Points of Interest in Beijing*. 2015: Springer International Publishing.

[4] Mazimpaka, J.D. and S. Timpf, *Exploring the Potential of Combining Taxi GPS and Flickr Data for Discovering Functional Regions.* 2015: p. 3-18.

[5] Fonte, C.C., et al., *Usability of VGI for validation of land cover maps.* International Journal of Geographical Information Science, 2015. **29**(7): p. 1269-1291.

[6] Antoniou, V., et al., *Investigating the Feasibility of Geo-Tagged Photographs as Sources of Land Cover Input Data.* Isprs International Journal of Geo-Information, 2016. **5**(5).

[7] Zhang, L., S.Z. Xie, and X.T. Zhen, *Towards Optimal Object Bank for Scene Classification.* 2013 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp), 2013: p. 1967-1970.

[8] Li, L.J., et al., *Object Bank: An Object-Level Image Representation for High-Level Visual Recognition.* International Journal of Computer Vision, 2014. **107**(1): p. 20-39.

[9] Felzenszwalb, P.F., et al., *Object Detection with Discriminatively Trained Part-Based Models.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 2010. **32**(9): p. 1627-1645.