Scale Effect on Fusing Remote Sensing and Human Sensing to Portray Urban Functions

Wei Tu, *Member, IEEE*, Yatao Zhang, Qingquan Li, Ke Mai, and Jinzhou Cao

***Abstract*— The development of information and communication technologies has produced massive human sensing data sets, such as point of interest, mobile phone data, and social media data sets. These data sets provide alternative human perceptions of urban spaces; therefore, they have become effective supplements for remote sensing tasks. This letter presents an exploratory framework to examine the scale effect of fusing remote sensing and human sensing. The physical and social semantics are extracted from raw remote sensing images and human sensing data, respectively. A dynamic weighting strategy is developed to explore the fusion of remote sensing and human sensing. Taking urban function inference as an example, the scale effect is evaluated by weighting remote sensing and human sensing. The experiment demonstrates that fusing remote sensing and human sensing enables us to recognize multiple types of urban functions. Meanwhile, the results are significantly affected by the scale.**

***Index Terms*— Data fusion, human sensing, scale, urban functions.**

I. Introduction

REMOTE sensing has been widely used to monitor urban environments from satellite views. Remote sensing images provide rich physical information, enabling the extraction of land use and land cover (LULC), impervious surfaces, and heat islands in cities [1]–[5]. Therefore, such images can foster the sustainable development goals of the United Nations [6].

With the evolution of 3G/4G/5G technologies, smartphones, and the Internet of Things (IoT), ubiquitous sensors are deployed in cities and capture urban data along with human movements or activities [7]. Humans are accepted as sensors to perceive the surrounding environment [8], [9]. Many alternative human sensing data sets, e.g., point of interest (POI), mobile phone records, street-view images, social media check-in, and geotagged photographs [4], [10], have been produced. Compared to remote sensing, human sensing is characterized by: 1) rich information of trajectories, human activities, and even human emotions [11]; 2) strong dynamics, as they are produced during human movements [12]; and (3) human preferences for urban spaces. Consequently, human sensing has been recognized as an effective supplement to remote sensing [13].

Fusing remote sensing images and human sensing data is a promising approach to understanding urban dynamics [14], [15], especially for portraying urban functions in high-density cities, e.g., Shenzhen, Tokyo, London, and New York. Recently, Zhang *et al.* [16] developed a new data fusion method by integrating remote sensing and human sensing to reveal functional urban land use. Tu *et al.* [11] integrated MODIS images and mobile phone records to detect the gradient of urban functions. These studies pushed the envelope of the convergence of remote sensing and human sensing.

Scale is an issue that must be explored in the fusion of multisource geospatial big data, as geographical phenomena, distributions, and processes are generally scale-dependent [17]. First, geographical objects might exhibit different attributes at different spatial scales, e.g., a building may be recognized as impervious land at a low scale but regarded as an industrial park at a high scale. Second, the scale-dependent phenomena can be reflected by different characteristics of various geospatial data. Intuitively, impervious land has a strong physical characteristic that can be expressed by remote sensing images, whereas an industrial park is more inclined to social cognition that can be expressed by human sensing data, e.g., POI. Furthermore, urban spaces also possess a hierarchical structure [18]. For example, a neighborhood can be deeply divided into several patches with different functions. A residential community usually includes buildings, green land, and facilities. Hence, these facts highlight the need to explore the scale effect on data fusion.

The goal of this letter is to explore the scale effect when fusing remote sensing and human sensing. Taking urban function

Manuscript received August 13, 2019; revised December 1, 2019; accepted January 4, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 71961137003, in part by the Natural Science Foundation of Guangdong Province under Grant 2019A1515011049, and in part by the Basic Research Program of Shenzhen Science and Technology Innovation Committee under Grant JCJY20180305312511388 3 and Grant JCYJ20170412105839839. *(Corresponding authors: Yatao Zhang; Qingquan Li.)*

Wei Tu, Ke Mai, and Jinzhou Cao are with the Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen 518060, China, also with the Guangdong Laboratory of Artificial Intelligence and Digital Economy (Shenzhen), Shenzhen University, Shenzhen 518060, China, and also with the Department of Urban Informatics, School of Architecture and Urban Planning, Research Institute of Smart Cities, Shenzhen University, Shenzhen 518060, China (e-mail: tuwei@szu.edu.cn; maikspring@foxmail.com; cjzszu@gmail.com).

Yatao Zhang is with the State Key Laboratory of Information Engineering of Survey, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: yatau@foxmail.com).

Qingquan Li is with the Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen 518060, China, also with the Guangdong Laboratory of Artificial Intelligence and Digital Economy (Shenzhen), Shenzhen University, Shenzhen 518060, China, also with the Department of Urban Informatics, School of Architecture and Urban Planning, Research Institute of Smart Cities, Shenzhen University, Shenzhen 518060, China, and also with the State Key Laboratory of Information Engineering of Survey, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: liqq@szu.edu.cn).

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.2020.2965247

recognition as an example, physical and social semantics are extracted from raw remote sensing images and human sensing data, respectively. A dynamic weighting framework is presented to link the physical and social semantics, and the scale effect is evaluated by means of clustering and classification while taking urban function inference as an example. The experiment demonstrates that fusing remote sensing and human sensing enables us to recognize multiple-type and multiple-scale urban functions.

The rest of this letter is organized as follows. Section II introduces the study area and data sets. Section III describes the proposed method. Section IV reports the experiments and results, and Section V concludes this letter.

II. Study Area and Data Sets

This study was conducted in Futian, Shenzhen, one of the most developed districts in China. Futian has a high population density, with 19 847 persons/km$_2$ in 2017. The utilized data sets include one high-spatial-resolution (HSR) remote sensing image and one POI data set. The HSR image with a spatial resolution of 2.5 m was obtained through the pan-sharpening fusion method applied to SPOT-5 images. The POIs were crawled from an online map service and held long-term human interactions with urban spaces [19]. The POIs were reclassified into 16 types, such as residential communities, entertainment facilities, and education sites. The land use maps were provided by the Commission of Natural Resource. Multiscale urban functions were further manually labeled as ground truth.

III. Methodology

This letter proposes an exploratory framework for evaluating the scale effect on fusing remote sensing and human sensing by taking urban function inference as an example. Fig. 1 shows the workflow, which contains four modules: hierarchical segmentation, semantic extraction, feature fusion, and scale effect exploration. First, land patches at multiple scales are produced by hierarchical segmentation. Second, the physical and social semantics are extracted from the HSR image and the POIs, respectively, using the probabilistic topic model. Third, a dynamic weight-based fusion model is presented to integrate two sourced semantics. Finally, clustering and classification are used to explore the scale effect.

A. Hierarchical Segmentation

This module produces the hierarchical partitions of the HSR image that contain a series of nonoverlapped image blocks. Given an image I, the segmented blocks are expressed as $\Lambda = \{R_1, R_2, \ldots, R_m\}$, where R_i is one block and $r_i(x, y)$ is the value of a pixel in R_i. If a model $f : r_i(x, y) \rightarrow s_i(x, y)$ satisfying $s_i(x, y) \approx r_i(x, y)$ is employed in I, then I can be approximately depicted as $S = \{S_1, S_2, \ldots, S_m\}$. The approximation introduces the errors between the original images I and the simplified images S, defined as $e(I)$

$$e(I) = \sum_{i=1}^{m} \sum_{(x,y)\in R_i} d(r_i(x, y) - s_i(x, y)). \quad (1)$$

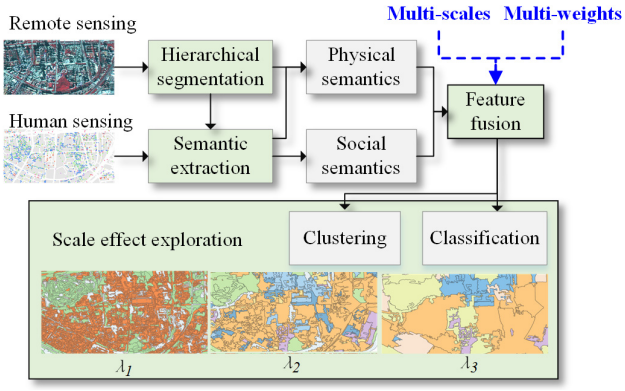

Fig. 1. Workflow for exploring the scale effect.

$d(\cdot)$ is a distance function, which refers to the differences in the spectrum, texture, or other information. When the number of segmented blocks increases, the block size decreases, and the simplification of the original image decreases.

It is implemented by the hierarchical step-wise region merging process that merges the partitioned image blocks from bottom to top [20]. Graph-based image segmentation is adopted in the HSR image to generate the initial image blocks. The layer of these image blocks is indeed the lowest hierarchy. The two adjacent image blocks with the minimum merging cost are merged to produce a new layer of image blocks with another scale (recorded as λ_k). Here, the merging cost between two adjacent image blocks is determined by the Baatz–Schape criterion and the color–texture dissimilarity. The Baatz–Schape criterion refers to the spectral homogeneity and spatial heterogeneity among image blocks, while the color–texture dissimilarity considers the spectral histogram and texture features among the image blocks. The merging is repeated iteratively until all the image blocks are merged. For details, please refer to [20].

The segmented results can be represented by the scale set $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Here, we denote the mean diameter of the image blocks of a hierarchy to indicate the spatial scale (λ_k). Fig. 2 shows the segmentation results from the small scale λ_1 (76.5 m) to the large scale λ_5 (671 m). The larger the scale, the greater the block size.

B. Semantic Extraction

Semantic extraction generates the high-level semantic features from the HSR image and the POIs. The bag-of-words (BOW) model [3] is adopted for feature selection. The BOW model quantizes the input features into visual words and then represents the raw data with the histogram of the visual words [21]. For the HSR images, three widely used features, including the spectral, texture, and scale-invariant-feature-transform (SIFT) features, are extracted by a moving window with a size of 25 pixels [16]. These extracted features are clustered into the corresponding visual words by the k-means method, and the optimal number of clusters is identified by the clustering metric [16], i.e., the number of one kind visual words (5 for the spectral features, 5 for the texture features, and 100 for the SIFT features). For the POIs, the visual words are directly defined as the types [22]. We count the frequencies of these two types of visual words to produce their BOW histograms.

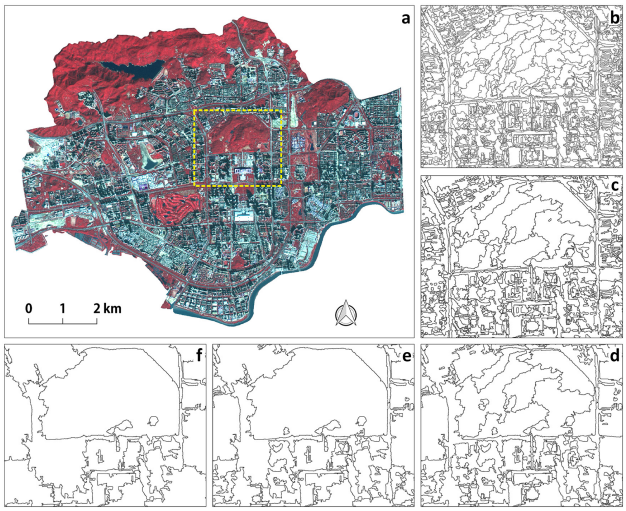

Fig. 2. Hierarchical segmentation results. (a) HSR image. (b) $\lambda_1 = 76.5$ m. (c) $\lambda_2 = 152$ m. (d) $\lambda_3 = 294$ m. (e) $\lambda_4 = 462$ m. (f) $\lambda_5 = 671$ m.

Latent Dirichlet allocation (LDA), one of the most widely used probabilistic topic models [3], is then utilized to produce the semantic information. The LDA model depicts each land patch with a mixture of underlying topics that are characterized as a mixture over visual words [3]. However, the LDA depends on the topic number [23]. Here, the hierarchical Dirichlet process (HDP) is introduced to avoid identifying the topic number. HDP is a Bayesian nonparametric model that can be used to infer the number of topics from the data automatically [23]. For the HSR image, the BOW histograms of the spectral, texture, and SIFT features are input into the HDP to extract their underlying mixed topics, i.e., the physical semantics. For the POI data, the BOW histogram of the POI features is input into the HDP to generate the social semantics.

C. Feature Fusion

Previous studies simply concatenated the physical semantics (s_{phy}) and the social semantics (s_{soc}) in a sequence to fuse them for future tasks [22]. These methods ignore the characteristic differences of remote sensing and human sensing and the variations at multiple spatial scales. Thus, we develop a dynamic weight-based fusion approach as follows:

$$\varphi_l = \{w \cdot s_{\text{phy}}, (1-w) \cdot s_{\text{soc}}\} \tag{2}$$

where φ_l is the fused semantics, s_{phy} and s_{soc} are the physical and social semantic features, respectively, and w and $(1-w)$ are the corresponding weights. The weight w represents the interaction between the physical and social semantic features. A lower w value indicates a higher importance of the social semantics, while a higher w value indicates a higher importance of the physical semantics.

Given land patches at a scale of λ_k, all possible fused representations are collected in $\phi_k = \{\varphi_1, \varphi_2, \ldots, \varphi_l, \ldots\}$. By varying the weight w from 0 to 1, the fused representations can cover all combinations of the physical and social semantics. Here, the weight step is set as 0.1. This simple and effective method enables us to evaluate the interaction of the HSR image and POI data over multiple scales.

D. Scale Effect Exploration

Clustering and classification are used to explore the scale effect. For clustering, given a scale λ_k, we utilize k-means clustering to group its land patches. The Davies–Bouldin index (DBI) is used to measure the appropriateness of the clustering results with different fused feature representations [24]. The index is defined by (3), where i and j are two clusters, S_i (S_j) is the intraclass dispersion, and M_{ij} is the interclass distance. R_{ij} is the ratio of the intraclass dispersion to the interclass distance. A lower DBI score indicates a better clustering result. For one of the fused feature representations φ_l, its DBIs with different partitions are recorded as db_l. $\text{DB}_k = \{db_1, db_2, \ldots, db_L\}$ refers to the DBIs with all fused feature representations at the scale of λ_k. Consequently, the DBI results can be obtained on the scale set Λ

$$db = \frac{1}{N}\sum_{i=1}^{N} R_i, \quad R_i = \max_{j \neq i} R_{ij}, \quad R_{ij} = \frac{S_i + S_j}{M_{ij}}. \tag{3}$$

Supervised classification is adopted to further evaluate the fusion for urban function inference at a given scale λ_k. The function of the land patches is first manually labeled with the help of street-view images, the land cover map, and the urban planning map. Then, the labeled samples are randomly divided into two sets, with 75% randomly selected as the training set and the remaining 25% used as the validation set. For the classification task, we utilize the support vector machine (SVM) classifier, which has achieved good performance in many studies [25]. The grid search method is used to identify the optimal parameters of the SVM with a step by times 10 in the search range. The range of the parameter Gamma is 10^{-8}–10^{8}, while the range of the parameter C is 10^{-6}–10^{6}. At a scale λ_k, several fusions with different weight configurations are conducted. The classification of fusions is run 50 times with randomly selected training and validation sets. The average overall accuracy (OA) and Kappa coefficient are used to evaluate the classification performance. The OA coefficient calculates the ratio of correctly classified items to the total number of items, while the Kappa coefficient measures the reliability of classification using a confusion matrix. Finally, the classification results are compared to assess the performance of fusing remote sensing and human sensing at multiple scales.

IV. Experiment and Results Analysis

This section reports the results of fusing HSR images and POIs via clustering and classification. The scale effect on data fusion is analyzed.

A. Exploring the Scale Effect on Data Fusion With Clustering

Clustering was implemented to group land patches with the fusion of remote sensing and human sensing at multiple scales. We varied the number of clusters from 5 to 50 and examined the change in DBI. Fig. 3 shows the distribution of DBIs with different weights for remote sensing and human sensing across multiple spatial scales. The results demonstrate that

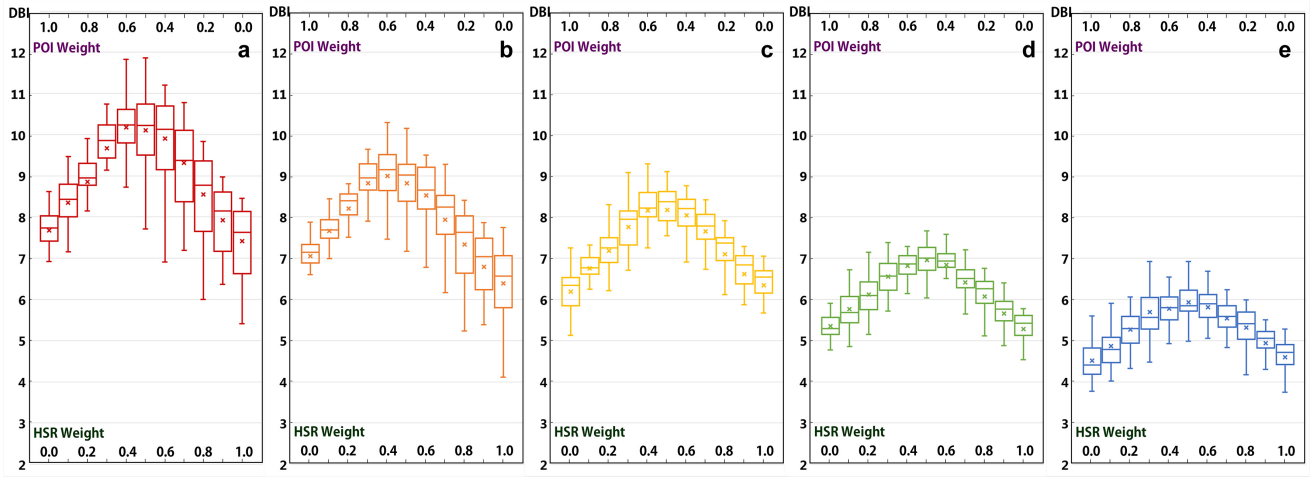

Fig. 3. DBIs for the multiscale fusion of remote sensing images and POIs. (a) $\lambda_1 = 76.5$ m. (b) $\lambda_2 = 152$ m. (c) $\lambda_3 = 294$ m. (d) $\lambda_4 = 462$ m. (e) $\lambda_5 = 671$ m.

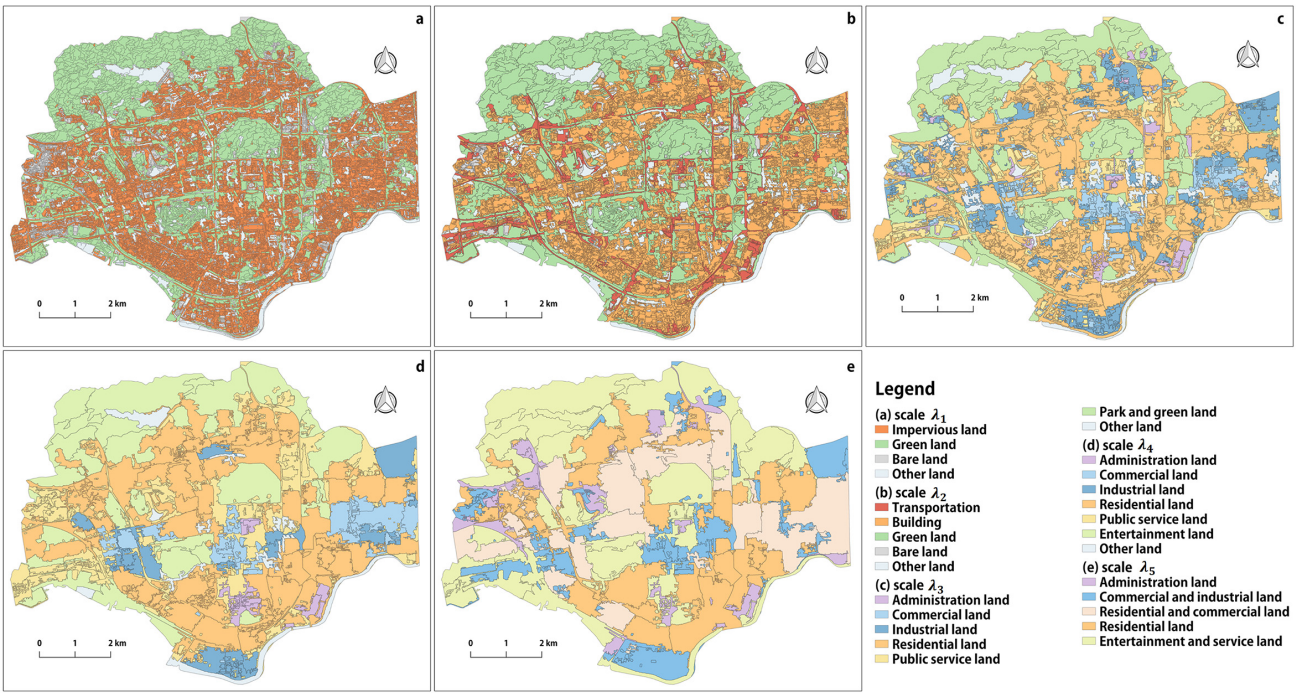

Fig. 4. Portraying urban functions using the best performing fused representations at five scales. (a) $\lambda_1 = 76.5$ m. (b) $\lambda_2 = 152$ m. (c) $\lambda_3 = 294$ m. (d) $\lambda_4 = 462$ m. (e) $\lambda_5 = 671$ m.

lower DBIs, suggesting good clustering performance, appear at the sides with the polarized weights of the HSR image and POIs. The highest DBIs appear at the center with the weight in the range [0.4, 0.6]. Therefore, fusing remote sensing and human sensing increases the diversity of features in the land patches, as more geospatial data will introduce more features.

These clustering results further demonstrate that the distribution of the clustering quality changes from the small scale of $\lambda_1 = 76.5$ m to the large scale of $\lambda_5 = 671$ meters. Fig. 3(a) and (b) shows that fusion with a higher weight for HSR images generates a lower DBI at the small scale, which suggests better clustering performance. Fig. 3(c)–(f) shows that higher POI weights produce better clustering results at the scales λ_3, λ_4, and λ_5. These results reveal that remote sensing image and POIs have stronger effects at low and high scales, respectively, because the HSR image contains higher physical homogeneity at low scales. As the scale increases, the adjacent patches are merged to generate new larger patches. Consequently, the physical homogeneity decreases. Meanwhile, the social properties of land patches emerge as the scale increases [18].

TABLE I

SVM CLASSIFICATION ACCURACY ASSESSMENT WITH DIFFERENT FUSED FEATURE REPRESENTATIONS AT FIVE SCALES

Scale	Ave.diam -eter (m)	$\langle 0.2, 0.8 \rangle$		$\langle 0.5, 0.5 \rangle$		$\langle 0.8, 0.2 \rangle$	
		OA	Kappa	OA	Kappa	OA	Kappa
λ_1	76.5	0.949	0.923	0.950	0.925	**0.955**	**0.932**
λ_2	152	0.890	0.851	0.931	0.906	**0.933**	**0.908**
λ_3	294	0.861	0.829	**0.866**	**0.834**	0.789	0.738
λ_4	462	**0.863**	**0.830**	0.850	0.812	0.782	0.726
λ_5	671	**0.817**	**0.751**	0.810	0.745	0.749	0.661

B. Exploring Scale Effect on Data Fusion With Classification

Taking the classification of multiscale urban functions as an example, the scale-dependent fusion of remote sensing

images and POIs was implemented by the SVM classifier with different weight settings. For each scale, the categories of urban functions are manually labeled according to the land use map. The numbers of labels from λ_1 to λ_2 are 215, 255, 137, 80, and 76, respectively. Considering the clustering results, we set three group weights: higher remote sensing weight $\langle 0.8, 0.2\rangle$, equal weights $\langle 0.5, 0.5\rangle$, and higher human sensing weight $\langle 0.2, 0.8\rangle$. Then, the classification accuracy was evaluated. The OA and kappa coefficients are reported in Table I. At multiple spatial scales, the fusion of remote sensing images and POIs achieves the highest accuracy with different weight settings. Similar to the clustering results, a higher remote sensing weight and a higher human sensing weight produce better urban function results at the small scale (λ_1 and λ_2) and the large scale (λ_4 and λ_5), respectively. Furthermore, fusion at the medium scale λ_3 produces better results with equal weights.

Fig. 4 shows the inferred urban functions across multiple scales. This figure demonstrates that scale-dependent fusion produces different land function maps. At the small scale λ_1 and λ_2, the interpreted land types are dominated by physical characteristics, i.e., buildings, transportation, and green land. When the scale increases, social urban functions appear. For example, at the scales λ_3 and λ_4, the interpreted land functions combine physical and social characteristics, such as public service land, commercial land, industrial land, and residential land. However, at the largest scale λ_5, the diversity of the inferred urban functions decreases, as the patches are larger than a street block, and the urban functions are mixed.

V. Conclusion

The fusion of remote sensing images and human sensing data is associated with the spatial scale. Taking inferring urban functions as an example, this letter presented a hierarchical scale framework to explore the scale effect on data fusion. The contributions of this letter are twofold: a data-driven exploration framework investigating the fusion of remote sensing and human sensing and a practical application illustrating the scale effect by taking the recognition of urban function as an example. The experiment in Shenzhen demonstrates that the performance of fusing remote sensing and human sensing depends on the targeted spatial scale. The weights of remote sensing and human sensing emphasize the lower and higher spatial scales, respectively. This further reveals that urban land properties can be classified into different categories on different spatial scales, such as urban land cover, land use, and land function. In the future, we will include more human sensing data sets in the presented framework, e.g., mobile phone data and social media data.

Acknowledgment

The authors would like to thank Prof. Z. Hu for the hierarchical remote sensing image segmentation software. They would also like to thank anonymous reviewers for their constructive suggestions and comments.

References

[1] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[2] P. Zhou, J. Han, G. Cheng, and B. Zhang, "Learning compact and discriminative stacked autoencoder for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4823–4833, Jul. 2019.

[3] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[4] R. Cao *et al.*, "Integrating aerial and street view images for urban land use classification," *Remote Sens.*, vol. 10, no. 10, p. 1553, Sep. 2018.

[5] X. Zhang, X. Wang, X. Tang, H. Zhou, and C. Li, "Description generation for remote sensing images using attribute attention mechanism," *Remote Sens.*, vol. 11, no. 6, p. 612, Mar. 2019.

[6] J. Trinder, S. Zlatanova, and J. Jiang, "Editorial to theme section on UN sustainable development goals (SDG)," *ISPRS J. Photogram. Remote Sens.*, vol. 142, pp. 342–343, Aug. 2018.

[7] W. Tu *et al.*, "Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 12, pp. 2331–2358, Dec. 2017.

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

[9] T. Blaschke, G. J. Hay, Q. Weng, and B. Resch, "Collective sensing: Integrating geospatial technologies to understand urban systems—An overview," *Remote Sens.*, vol. 3, no. 8, pp. 1743–1776, Aug. 2011.

[10] Y. Liu *et al.*, "Social sensing: A new approach to understanding our socioeconomic environments," *Ann. Assoc. Amer. Geographers*, vol. 105, no. 3, pp. 512–530, May 2015.

[11] W. Tu *et al.*, "Portraying urban functional zones by coupling remote sensing imagery and human sensing data," *Remote Sens.*, vol. 10, no. 1, p. 141, Jan. 2018.

[12] J. Li *et al.*, "Social media: New perspectives to improve remote sensing for emergency response," *Proc. IEEE*, vol. 105, no. 10, pp. 1900–1912, Oct. 2017.

[13] Y. Zhou *et al.*, "A global map of urban extent from nightlights," *Environ. Res. Lett.*, vol. 10, no. 5, May 2015, Art. no. 054011.

[14] N. Zhao, W. Zhang, Y. Liu, E. L. Samson, Y. Chen, and G. Cao, "Improving nighttime light imagery with location-based social media data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2161–2172, Apr. 2019.

[15] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, p. 38, 2014.

[16] Y. Zhang, Q. Li, W. Tu, K. Mai, Y. Yao, and Y. Chen, "Functional urban land use recognition integrating multi-source geospatial data and cross-correlations," *Comput., Environ. Urban Syst.*, vol. 78, Nov. 2019, Art. no. 101374.

[17] S.-C. Li and Y.-L. Cai, "Some scaling issues of geography," *Geograph. Res.*, vol. 24, no. 1, pp. 11–18, 2005.

[18] X. Zhang, S. Du, and Q. Wang, "Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data," *ISPRS J. Photogram. Remote Sens.*, vol. 132, pp. 170–184, Oct. 2017.

[19] G. Mckenzie, K. Janowicz, S. Gao, and L. Gong, "How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest," *Comput., Environ. Urban Syst.*, vol. 54, pp. 336–346, Nov. 2015.

[20] Z. Hu, Q. Zhang, Q. Zou, Q. Li, and G. Wu, "Stepwise evolution analysis of the region-merging segmentation for scale parameterization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2461–2472, Jul. 2018.

[21] Y. Zhang, R. Jin, and Z.-H. Zhou, "Understanding bag-of-words model: A statistical framework," *Int. J. Mach. Learn. Cyber.*, vol. 1, nos. 1–4, pp. 43–52, Dec. 2010.

[22] X. Liu *et al.*, "Classifying urban land use by integrating remote sensing and social media data," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, Aug. 2017.

[23] C. Wang, J. Paisley, and D. Blei, "Online variational inference for the hierarchical Dirichlet process," in *Proc. 4th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 752–760.

[24] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.

[25] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?" *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133–3181, 2014.