Image Processing and Analysis by Remote Sensing: A Review Paper

[1]Ajay Kumar

Abstract

High-resolution urban land use maps have important applications in urban planning and management, but the availability of these maps is low in countries such as India due to lack of resources. To address this issue, we have developed a protocol to identify urban land use functions over large are as using satellite images and open social data. We first derived parcels from road networks contained in Open Street Map (OSM) and used the parcels as the basic mapping unit. We then used 10 features derived from Points of Interest (POI) data and two indices obtained from Landsat. Similarity measures and threshold methods were used to identify land use types in the classification process. This protocol was tested in Haryana, India. The results showed that the generated land use map had an overall accuracy of 91.04 %. The map revealed significantly more details of the spatial pattern of land uses in Haryana than the land use map released by the government.

Keywords: urban land parcel; remote sensing; social data; land use

Introduction

Urbanization in India is taking place at fast rate. The proportion of urban population increased from 20% in 1982 to more than 50% in 2015. Large-scale urbanization has had a dramatic impact on the environment. Studies that assess this process and its impacts are important for taking remedial actions and designing better urbanization strategies for the future. To achieve these goals, detailed urban land cover/use maps are required.

Currently, land cover information with resolutions ranging from low to high is the primary data source used in studies such as urban growth simulation, evaluation of urban public health, and assessment of urban ecosystem services .However, to study issues such as housing provision, urban transportation, job accessibility and residential relocation ,and land use patterns, detailed information on urban land use is needed due to the difference between the two concepts: land use is a cultural concept that describes human activities and their use of land, whereas land cover is a physical description of land surface. Land cover can be used to infer land use, but the two concepts are not entirely interchangeable.

Nevertheless, high-resolution urban land use maps covering large spatial extents are relatively rare because local knowledge and the techniques necessary for developing these types of maps are often not available, particularly for developing regions. Moreover, urban land use maps are normally produced by interpreting aerial photographs, field survey results, and auxiliary materials, such as appraisal records or statistical data..

Satellite-based remote sensing holds certain advantages in monitoring the dynamics of urban land use because of the large spatial coverage, high time resolution, and wide availability. Pixel-based image classification methods using spectral and/or textural properties are frequently applied to extract urban land use information. Recently, per-field and object-based classification methods have gained popularity in deriving land uses from the satellite images because per-field classification methods can better describe the function of urban areas and serve the needs of urban planning . Although significant progress has been achieved, deriving high-resolution urban land use maps from satellite images is still a difficult task. The medium-resolution satellite images (e.g., Landsat images) allow for mapping urban areas at the large spatial scale, but it is still difficult to extract socioeconomic features of urban areas from these images. Land cover information derived from medium-resolution satellite images cannot provide sufficient separation among urban functional zones.

Satellite images with high spatial and spectral resolution provide more detailed information on urban structures and thus facilitate the assignment of socioeconomic functions to different zones. Nevertheless, these images are prohibitively expensive in general.

Data Collection

The administrative boundary of Haryana falls over two sets of Landsat images (path/row: 123/32 and 123/33). Totally 14 Landsat 8 Operational Land Imager (OLI) images of 2013 were procured as our primary data source from the U.S. Geological Survey (http://earthexplorer.usgs.gov/).These images were selected because of the low cloud proportions (<10%). Multiple available Landsat images with good quality in 2013 were used to remove the impact of cloud contamination, phenology of vegetation and cropland rotation. In addition, a seasonal dynamic of land cover series,

Data on the road networks of Haryana were collected from Open Street Map (OSM) (https://www.openstreetmap.org), a provider of free open geographical data. The data are in vector format and contain different classes of streets organized using street levels and sizes. Street levels, in descending order, correspond to primary highways, primary roads, secondary roads, and small roads (i.e., local, neighborhood and rural streets). Each point contains the functional and locational properties of a site, The initial twenty types of POI were aggregated into 10 general categories, including residential, marketing and recreation, service building, hotel and restaurant, industrial, medical, educational, institutional infrastructure, government and social organization, and transportation land .POIs that did not belong to the aforementioned groups were removed. The quality of the POI data were verified by checking 100 randomly sampled sites manually and the resulting accuracy level was 97%. Although spurious social data may occur, the overall pattern (or distribution) can be accurately reflected by using a huge amount of points.

Method

The overall structure of the protocol is shown in Figure. First, the entire study area was segmented into parcels based on road networks following. Parcels are basic units used in this classification scheme with the assumption that they are homogeneous in terms of urban functions . The parcels were then separated into built-up areas and non-built-up areas based on classified impervious surface areas and defined our classification system based on these two regions. The function of each parcel was inferred using the normalized feature distance (or similarity) to the pre-collected training sample units. The similarity of the built-up parcels was based on 10 socioeconomic features (i.e., residential, marketing and recreation, service building, hotel and restaurant, industrial, medical, educational, institutional infrastructure, government and social organization and transportation land) that were derived from the normalized kernel densities of the different functions of POI data and two physical indices derived from multi-temporal Landsat images.

Processing POIs

Within the spatial extent of a parcel, there may be a variety of POIs of different types, which can be regarded as having compound functions instead of a single function. In addition, the qualities of POIs vary among different categories, i.e., the number of POIs associated with the commercial type is greater than the other types. This results in an unbalanced distribution of the numbers of points among different POI types. To cope with these issues, we normalized the functional intensity of the different POI types using kernel density estimation. Kernel density analysis was implemented using the quadratic kernel function with a search radius of 500 m. The output is a smooth surface indicating the densities, and regions with relatively higher density values indicate that there are more POI points. This processing can mitigate possible errors caused by unbalanced quantity gaps among different POI types.

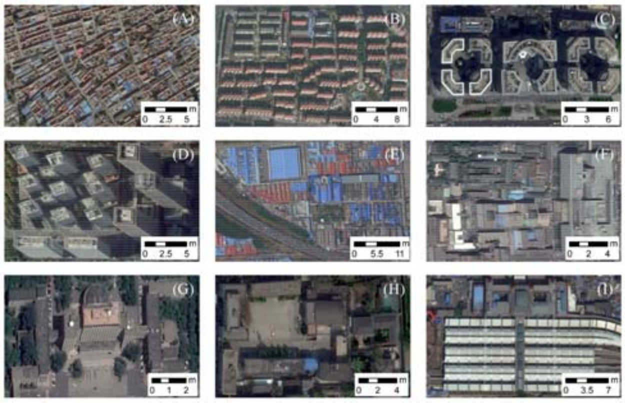

Figure: Training samples for nine subclasses of land use in the built up regions: (a) cottage; (b) community; (c)retail place; (d) service building; (e) industrial lands; (f) medical places; (g) education /research places; (h) administrative office.

Determination of Parcel-Based Land Use

Using the training parcels that were collected based on the definition of the land use classification, a normalized feature distance (similarity) in built-up regions was computed on a parcel-by-parcel basis. The features used for calculating the similarity index include 10 POI density images, one NDVI band and one NDBI band. Two statistical parameters (i.e., the mean value and standard deviation) were previously estimated using the collected training samples. Then, the similarity index of a given parcel was compared. x_i and s_i are the mean and standard deviation of the pre-defined land use type i acquired from the training parcel; m is the total number of land use classes; and x_j is the parcel value (i.e., the mean of all pixels within the parcel) for feature j in either normalized POI density images, NDVI or NDBI. The smaller feature distance means higher similarity to a corresponding land use type. The urban land use type of a parcel was determined by calculating similarity S_i of the parcel to training samples, and the pre-defined land use type of training samples which has the minimum value of S_i was assigned to this parcel.

In addition, the land cover map was adopted to determine the land use of parcels in non-built-up lands. A land parcel commonly has multiple land cover types. The dominant land cover type was identified and the land use function of this land cover type was assigned to the corresponding parcel. Finally, the classified built-up and non-built-up areas were combined to form the detailed land use map for the entire city.

Accuracy Assessment and Uncertainty

To assess the performances of land use classification, a random sampling scheme was adopted to collect a testing sample set over the study area. All testing parcels were surveyed by a field crew with a relatively high level of confidence. The total number of collected testing parcels was 269, among which 180 were located in built-up regions, and 89 were in non-built-up areas. Thereafter, confusion matrixes for Level I and Level II were built.

To assess the uncertainty of the obtained urban land use map, the standard deviation of similarities among all land use types for each parcel was used as an approximate indicator. Low standard deviation values suggested that there was a relatively little similarity difference among the different types of land use, which meant that the identified land use was more uncertain. High standard deviation values indicated considerable variance of similarities among different land use types, which would tend to result in a more credible assigned type of land use to a parcel because the minimum similarity was considerably different. This approach qualitatively captured the overall pattern of the uncertainty distribution.

Discussion

In This study, we utilized both medium resolutions. Satellite remote sensing data and open social data describe the biophysical elements of urban areas. The open social data indicate human activities of a place, especially inside the built-up area. The use of both data sources is beneficial for urban land use type identification in urban areas. In addition, the sensitivity of features adopted in the similarity assessment was tested. If we used only the POI or two biophysical features derived from the Landsat images (i.e., NDVI, NDBI), the obtained accuracies of the urban land parcels (Level I) in the built-up region are 41% and 31%, respectively. However, when combined, the accuracy reaches 75%. Moreover, some function types that were not identified in previous studies have been classified in our result, such as service buildings, medical and public places. Hence, the use of both features at the parcel level produced more detailed and accurate land use maps than studies using single source data.

Conclusions

High-resolution land use maps are needed for academic research and urban management. However, complex and heterogeneous urban landscapes pose challenges to land use mapping. The use of both physical features (i.e., spectral information) derived from remotely sensed data and social attributes (e.g., socio-economic function) derived from open social data can help to delineate the detailed land use patterns in urban areas. We developed an approach to combine the strength of these two types of data to identify land use types quickly over a large area. The use of both biophysical features and socioeconomic features resulted in a land use map with higher accuracy and more detail. The overall accuracy of the land use map for this extremely heterogeneous urban area reached 81.04% and 69.89% for the Level I and Level II categories, respectively. This approach can be applied to derive land use maps of a large area in a relatively short time wherever satellite data and open social data are available, especially for fast-growing urban areas in developing countries.

More efforts can be undertaken for further improvement. First, incorporating both road networks derived from open social data and segmentation derived from biophysical features (e.g., bands in Landsat images) may be helpful to generate more detailed parcels, especially for suburban areas in which road networks are relatively sparse. Second, the shapes and sizes of parcels can be considered when assigning possible land use types (e.g., parcels in industrial areas are relatively large, but parcels in residential areas are small). Third, non-linear approaches, such as neural networks, can be considered when building the relationships between physical and socioeconomic features, if the number of pre-collected training parcels is adequate.

References

[1] Hunter, G., Cox, C., Kremer, J., 2006. Development of a commercial laser scanning mobile mapping system – StreetMapper. Int.

[2] Jason, B. D., Ralph, O., Dubayaha, D.B., Clark, Estimation of tropical forest structural characteristics using largefootprint lidar. Remote Sensing of Environment.

[3] Mahmoud, F.A.,Ahmad,R., Mohd Hasmadi,I., Alias,M.S., 2010. Utilization of the dynamic laser scanning technology.

[4] Mohd Hasmadi,I., Kamaruzaman, J., Pakhriazad, H.Z., Frisco, N,.2007. Geoinformatics for better forest.

[5] Murphy, G.E., 2008. Determining stand value and log product yields using terrestrial lidar and optimal bucking: a case study. Journal of Forestry, 106.

[6] Nohr, H., 2001, Management der Informationsqualität, Nr. 3/2001, Fachhochschule Stuttgart, Stuttgart.

[7] Olson, J.E, 2003. Data quality - The Accuracy Dimension, Morgan Kaufmann Publishers, San Francisco.

[8] Palmer, A.L., 2008. Historical dictionary of architecture: The Scarecrow Press, Inc.

[9] Popescu, S. C., Wynne, R. H., Scrivani,J. A., 2004. Fusion of small footprint lidar and multispectral data to estimate plot-level volume and biomass in deciduous and pine forests in Virginia.

[10] Reutebuch, S.E., Andersen, H.E., McGaughey, R.J., 2005. Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory. Jour. For. Sept., 286-292

[11] Richardson, J., Moskal, L. M., Kim,S., 2009. Modeling approaches to estimate effective leaf area index from aerial discrete-return LIDAR., Agricultural and Forest Meteorology, 149, 1152-1160.

[12] Wehr, A., Lohr, U., 1999. Airborne laser scanning – an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing, 54, 68–82. Sebestyen, G., 1998. Construction: Craft to industry. London: Taylor & Francis, UK.

[13] Young, M.,1986. Optics and lasers: Including fibers and optical waveguide. Berlin, Springer Verlag.

Author's details

[1]M.Tech Scholar, Department of Civil Engineering, Patel Institute of Technology, Bhopal

Email: ajaykumar8mail@gmail.com