Research on Rapid Extraction Method of Urban Built-up Area with Multiple Remote Sensing Indexes

Yuntao Ma[1], Zhiwei Xie[1(✉)], Yingchun You[1], and Xin Xuan[2]

[1] School of Transportation Engineering, Shenyang Jianzhu University, Shenyang 100168, China
zwxrs@sina.com
[2] College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

Abstract. The urban built-up area is a key monitoring and statistics data of national and provincial statistical departments, and has become an important scale to reflect the level of urban development and predict the potential of urban development. Based on the Landsat 8 images, by using the way of spectral signature analysis, the study selected four indices, Modified Normalized Difference Built-up Index (MNDBI), Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Water Index (MNDWI) and a set of mineral index (Clay, Iron, Ferrous), and combined with the band operation and K-means unsupervised classification algorithm to extract the range of urban built-up area. Taking Shenyang as an example, the method proposed in this paper was used to extract the built-up area of Shenyang and the Kappa coefficient of the extracted built-up area range was 0.9091 after the post-classification treatment and refinement treatment of the built-up area.

Keywords: Landsat · Built-up area · MNDBI · NDVI · MNDWI

1 Introduction

The urban built-up area refers to the area within the urban administrative region that has actually been developed and constructed, been provided with the municipal public utilities and public facilities basically. Accurate acquisition of urban built-up area has important guiding significance for urban construction, management and research, and is also one of the important indicators reflecting the comprehensive economic strength and urbanization level of a city [1, 2]. However, urban built-up area is a complex land use type, which cannot be accurately extracted from remote sensing images by simple methods [3]. DMSP/OLS nighttime light data are widely used at present [4], but it is susceptible to the problems of saturation, diffusion and low resolution of nighttime light data, so it is still a great challenge to extract urban built-up area by DMSP/OLS data alone [5]. Object-oriented image segmentation method requires multiple experiments to select appropriate threshold, segmentation method and rule set, which takes a lot of time [6, 7]. And the index-based information extraction of built-up area has been increasingly applied [3, 8] due to its simplicity and ease of use. The paper based on

Y. Xie et al. (Eds.): GSES 2019/GeoAI 2019, CCIS 1228, pp. 199–209, 2020.
https://doi.org/10.1007/978-981-15-6106-1_15

Modified Normalized Difference Built-up Index (MNDBI) [9], Modified Normalized Difference Water Index (MNDWI) [10], and Normalized Difference Vegetation Index (NDVI) [11], combined with a group of mineral indexes and unsupervised classification methods [9, 12], proposed a rapid extraction method of urban built-up area based on multiple indexes. The overall accuracy of extraction results after post classification and refinement of built-up area can reach 97.14%, and the Kappa coefficient is 0.9091.

2 Extraction and Post Classification of Built-up Area

2.1 Extraction of Built-up Area

Ground objects in the urban built-up area of Shenyang can be divided into five types: built-up area, forest land, arable land, bare soil and water [9, 10]. This paper randomly selected 2,500 pixel points as samples, 500 pixel points for each type, and the reflection spectrum characteristic curve of ground objects is shown in Fig. 1. The spectral signatures of five types of ground objects was analyzed through the reflection spectrum characteristic curves, which shows the differences between bare soil and built-up area in Swir1 was very small. So if the Normalized Difference Built-up Index (NDBI) [8] (Formula (1)) was selected, the bare soil would be strengthened at the same time as the built area, which is not conducive to the accurate extraction of the built area. This study selected Modified Normalized Difference Built-up Index (MNDBI) proposed by

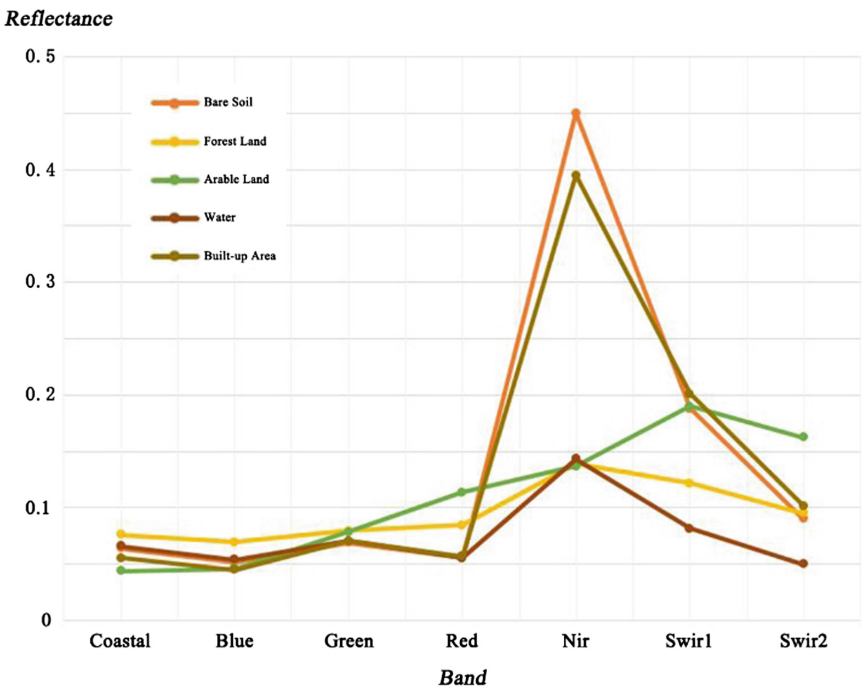


Fig. 1. The reflection spectrum characteristic curves of ground objects

Yueyao Hu to preliminary extract the built-up area of Shenyang. MNDBI used Swir2 instead of Swir1 band of NDBI, enhanced the difference between the built-up area and the bare soil, and effectively reduced the problem of miscellany between bare soil and built-up area [9]. The MNDBI formula is shown in Formula (2).

$$NDBI = \frac{Swir1 - Nir}{Swir1 + Nir} \tag{1}$$

Where $Swir1$ is the reflectance of band 6 Shortwave Infrared 1 (1.57–1.65 μm), Nir is the reflectance of band 5 Near-Infrared (0.85–0.88 μm).

$$MNDBI = \frac{Swir2 - Nir}{Swir2 + Nir} \tag{2}$$

Where $Swir2$ is the reflectance of band 7 Shortwave Infrared 2 (2.11–2.29 μm).

Using MNDBI alone to extract the urban built-up area cannot avoid the interference of water and vegetation in and around the urban built-up area, so vegetation index and water index were introduced to reduce the influence caused by the wrong classification of vegetation and water into built-up area. The Normalized Difference Vegetation Index (NDVI) proposed by Rouse was selected, as shown in Formula (3). The Modified Normalized Difference Water Index (MNDWI) was selected because it still has good extraction effect in the case that the background of the buildings is more [10]. This paper used Swir2 band instead of Mir [13] because the Landsat 8 OLI sensor does not contain Mir band. The formula is shown in (4).

$$NDVI = \frac{Nir - R}{Nir + R} \tag{3}$$

Where R is the reflectance of band 4 Red (0.64–0.67 μm).

$$MNDWI = \frac{G - Mir}{G + Mir} \tag{4}$$

Where G is the reflectance of band 3 Green (0.53–0.59 μm), Mir is the reflectance of band 7 Mid-Infrared (2.08–2.35 μm) of TM.

After index calculation, the threshold models of built-up area, water and vegetation were constructed according to the statistical results, and binary classification was carried out to eliminate the mixed water and vegetation in the built-up area, so as to improve the classification accuracy.

Due to urban development and irregular distribution of bare soil, there were still a lot of bare soil in the built-up area and surrounding towns after removing the disturbance of misclassified water and vegetation, which would affect the extraction accuracy of the built-up area. In this study, the method of combining a group of mineral indexes [12] with k-means unsupervised classification proposed by Ward in 2000 was used to eliminate the bare soil mixed in urban built-up area. This method can effectively reduce the wrong sample points which were misclassified into the built-up area, reduce the discrete points outside the built-up area significantly, reduce the false alarm rate, and improve the overall accuracy and Kappa coefficient [9]. This group of mineral indexes includes:

Iron Oxide Index:

$$Iron = \frac{Red}{Blue} \quad (5)$$

Where *Blue* is the reflectance of band 2 Blue (0.45–0.51 μm).

Clay Mineral Index:

$$Clay = \frac{Swir1}{Swir2} \quad (6)$$

Ferrous Metal Index:

$$Ferrous = \frac{Swir1}{Nir} \quad (7)$$

K-means clustering algorithm is an unsupervised real-time clustering algorithm advanced by Mac Queen [14]. On the basis of minimizing the error function, the data is divided into a predetermined number of clusters. The principle of the algorithm is simple and it is easy to process a large amount of data [15].

2.2 Data Post-classification and Construction Area Refinement

Because the spectral characteristics of buildings in urban villages and urban built-up areas are similar, many small patches were formed around the built-up area, which affected the extraction accuracy. Within the built-up area, due to the existence of large lakes, rivers and parks, the spectral characteristics of them are quite different from those of the most widely distributed buildings in the built-up area, which often be missed and form voids. In order to reduce the impact of small patches and voids and further meet the needs of classification, the necessary post-classification and the refinement steps of build-up area were needed. Majority Analysis and Clump were used to post process the classification results.

Majority Analysis is a form of convolution filtering for smoothing and sharpening purposes actually. It replaces the central pixel category with the pixel category which occupies the most important position (the largest number of pixels) in the transform kernel.

For the voids within the built-up area, although the low-pass filtering can smooth the image well, it is affected by the coding of surrounding objects easily. Clump processing solved this problem well. Clump processing is an algorithm that uses mathematical morphological operators (erosion and dilation) to cluster and merge adjacent similar classification regions.

The extraction results of built-up area after post-classification still had some voids that cannot be filled and small patches that cannot be removed. And it did not conform to the definition of built-up area and was not conducive to data statistics, so built-up area needed to be refined. The range of built-up area was transformed into vector and imported into ArcMap 10.2. The fragmented map patches outside the built-up area were deleted and the voids in the built-up area were filled, and the built-up area became a whole map patch.

3 Experiment and Analysis

In this paper, Landsat 8 OLI image data in Shenyang on August 31, 2017 were adopted, including 8 multispectral bands with resolution of 30 meters and a panchromatic band with resolution of 15 meters. In order to obtain surface albedo, radiometric calibration and atmospheric correction were carried out. In this paper, the processing of remote sensing image data also includes image fusion and clipping. The technical route of this experiment is as Fig. 2.

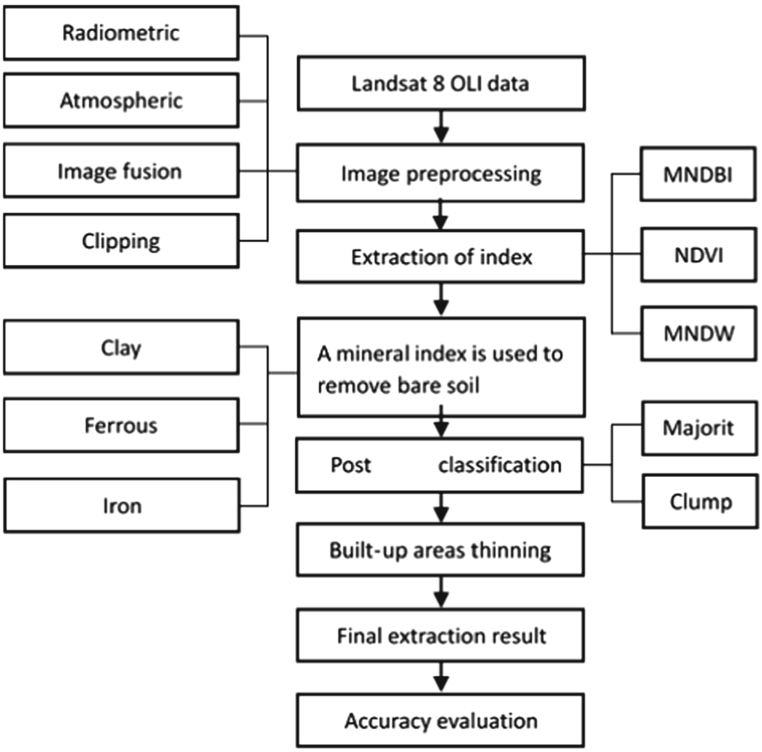


Fig. 2. The flow chart of the built-up area extraction

3.1 Extraction of Index and Establishment of Threshold Model

Figure 2 is the flow chart of the built-up area extraction. After many experiments and statistical analysis, it is found that the brightness of the ground object corresponding to each index is enhanced after calculation, while other objects are suppressed. Therefore, it is easy to achieve the goal of binary classification by establishing a threshold model, so as to extract the required ground objects for the next built-up area extraction and vegetation and water mixed removal. The calculated value range of each classification is shown in Table 1.

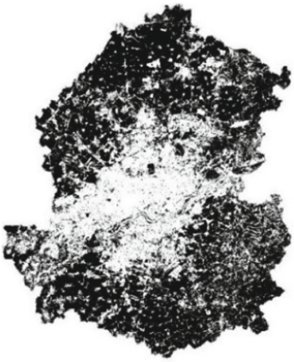
Table 1. The classification value range

Index	Features	The minimum	The maximum
MNDBI	Bare soil	−0.3576	1.300
	Forest land	−0.6664	−0.3820
	Arable land	−0.7255	−0.4907
	Water	−0.6695	−0.3004
	Built-up area	−0.5032	0.5248
NDVI	Bare soil	−1.2022	0.5553
	Forest land	0.5698	0.8286
	Arable land	0.6659	0.8137
	Water	0.1328	0.8106
	Built-up area	−0.4083	0.5391
MNDWI	Bare soil	−0.5151	−0.2320
	Forest land	−0.2709	−0.0692
	Arable land	−0.2020	−0.0546
	Water	−0.1240	0.4789
	Built-up area	−0.5122	0.1162

According to the value range of five types of ground objects after calculation and combining with the actual situation, MNDBI threshold was chosen as −0.50 in order to retain the scope of urban built-up area to the greatest extent. At this time, the arable land was well removed, and would not be considered in the next extraction.

After visual inspection, part of woodland, water and bare soil were still mixed in the extracted built area. These ground objects in the binary images obtained only through MNDBI calculation and classification were difficult to be distinguished, so NDVI and MNDWI was introduced to extract and remove them in the next step.

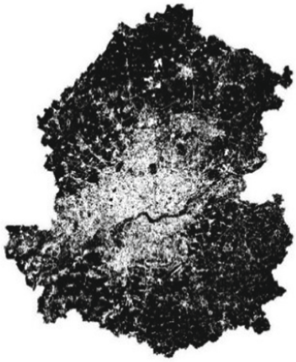
According to the statistical results after NDVI calculation, 0.56 was selected as the threshold value, as a result, not only was the built-up area preserved completely, but also the forest land was removed well, the situation that part of the non-built-up area of the forest land was misclassified into the built-up area was reduced effectively, and the fine patches were reduced. Through the calculation of NDVI, the vegetation mixed in urban built-up area was removed well. After MNDWI calculation, according to the statistical results, a threshold value of 0.10 was selected to conduct binary classification



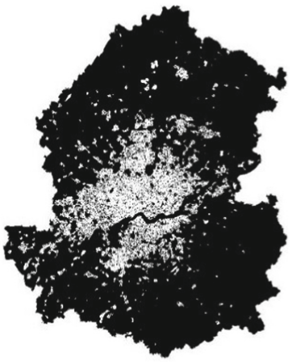
a. MNDBI extraction results

b. Result after removing vegetation

c. Result after removing water

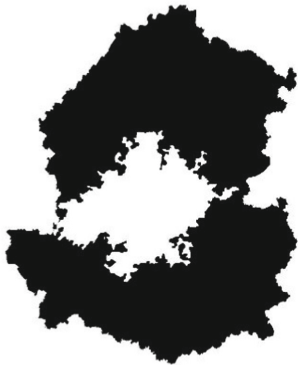


d. Result after removing bare soil



e.Majority analysis result

f.Clump result bare soil



g. Result after thinning

Fig. 3. Extraction results of built-up area in each step

of the calculated results. The index could well remove the water mixed in the built-up area and the small patches outside the built-up area, while retaining the urban built-up area to the greatest extent.

3.2 A Group of Mineral Indexes Are Used to Remove Bare

There was still a situation that bare soil was misclassified into built-up area after exponential extraction, which was not conducive to the post-processing of data and the evaluation of the final accuracy. In this paper, a group of mineral indexes were introduced to identify and eliminate the misclassified bare soil in the urban built-up area. Firstly, Iron, Clay and Ferrous were calculated, and the calculated results were superimposed. K-means unsupervised classification algorithm was used to classify the mineral indexes. The classification results were visually identified, and the bare soil part was selected to be removed from the results obtained by index extraction. The comparison of extraction results of each step is shown in Fig. 3.

4 Conclusion and Analysis

According to the definition of urban built-up area, the built-up area includes the areas that have been developed and constructed in the urban administrative area, the municipal public facilities and public facilities have been fully equipped in the urban

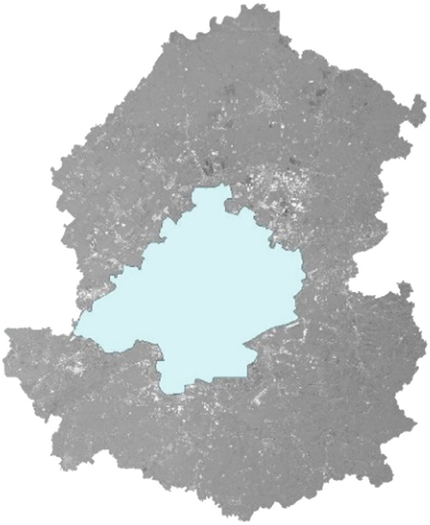


Fig. 4. The reference range of built-up area (hand-extracting result (blue)) (Color figure online)

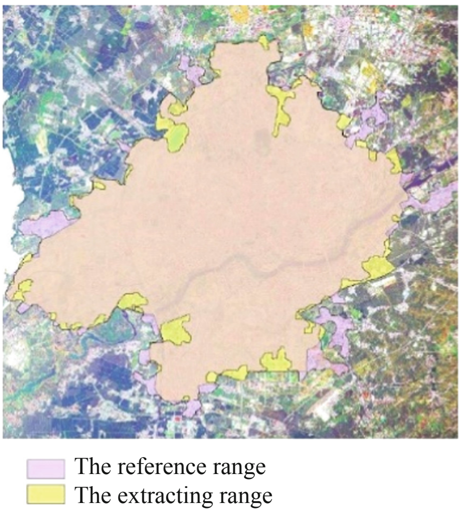


Fig. 5. The reference range and the extracting range of built-up area

administrative area, includes the buildings, highways, bridges, lakes, rivers and other infrastructures that have been built and improved in the urban area. Because of the variety and complexity of the ground objects species, it was not typical to use the Confusion Matrix Using Ground Truth ROIs function of ENVI 5.1 to evaluate the accuracy of ground objects by using pixels based on their spectral characteristics. This paper used area as evaluation criterion to establish confusion matrix to evaluate the classification results, which could minimize subjectivity and reduce the impact of other objects within the built-up area.

In 2017, the built-up area of Shenyang was 633.8 km^2 [16]. Taking the pre-processed remote sensing data of Shenyang in August 2017 as the base map, referring to the remote sensing images of the same period in 2015 and 2016, and combining the image data and information provided by Google Earth and Auto Navi Map, a map patch representing the urban built-up area of Shenyang was plotted. As shown in Fig. 4, its area is 633.8 km^2. Figure 5 shows the comparison between accuracy assessment area and built-up area. The "combined" superposition analysis tool in ArcMap 10.2 was used to combine the experimental results with the property sheet of map patches in the built-up area of Shenyang. By comparison, the correct classification, misclassification and missing area of built and non-built were calculated. Confusion matrix was made and the User's Accuracy, Producer's Accuracy, Overall Accuracy and Kappa coefficient of each step were calculated and were used to evaluate the Accuracy, the results are shown in Table 2.

Table 2. The extracting accuracy chart of the built-up area in each step (User's accuracy, Producer's accuracy, Overall accuracy and Kappa coefficient)

Precision index	MNDBI extraction	Mineral index	Post classification	Proper refining
User's accuracy	89.04%	66.96%	81.43%	92.46%
Producer's accuracy	26.49%	45.46%	68.64%	91.91%
Overall accuracy	52.845	79.28%	89.55%	97.14%
Kappa coefficient	0.1180	0.4140	0.6818	0.9091

The results show that the extraction effect was poor only through MNDBI, and only the area of the User's Accuracy meets the needs. However, many ground objects in the built-up area were misclassified into non-built-up areas, resulting in low overall accuracy and poor reliability. The results showed that it was not nearly enough to extract the built-up area only through MNDBI because of the complexity of the ground objects included in the urban built-up area, and the elimination and post classification steps of other interfering ground objects must be carried out.

Eliminating the influence of water, vegetation, bare soil and other ground objects could reduce the area of non-built area misclassified in the built-up area significantly, but it also increased the voids in the built-up area, reduced the area, and reduced the user accuracy. However, the overall classification accuracy and reliability were improved significantly, but the available state has not been reached yet. Further post classification was needed to increase the built-up area, and improve user's accuracy and overall accuracy.

The post classification filled the voids existing in the built-up area, effectively increased the area of built-up areas, and significantly improved the user's accuracy. The area of the non-built area which was misclassified into the built-up area and the area of the built-up area which was misclassified into the non-built area were also significantly reduced. The overall accuracy and Kappa coefficient were improved, which basically achieved the desired results of the experiment.

The refinement of built-up area effectively removed the hard-to-remove villages and large areas of land to be developed in the periphery of the city, and filled in the large and hard-to-fill voids within the built-up area. Compared with the results before refinement, the accuracy of the built-up area has been improved, the accuracy index has reached more than 90%, and the extracted data has reached the available state, which also proved the necessity of refinement of the built-up area.

The various remote sensing indexes adopted in this paper have the advantages of convenient calculation, low learning cost, easy realization and the built-up area can be extract fast. In addition, combined with the processing steps of removing vegetation, water and bare soil mixtures, post-classification and refinement of built-up areas, the final extraction results have high classification accuracy and location accuracy, which all meet the statistical requirements. This method is suitable for rapid extraction and long-term change monitoring of built-up area, and can play a very good role in the monitoring of geographical conditions.

Acknowledgments. This study was supported by Natural Science Foundation of Liaoning Province of China: Study on Remote Sensing Monitoring Method for Maize Planting Area in Liaoning Province (No. 20180550479).

References

1. Li, G.D., Fang, C.L., Wang, S.J., Zhang, Q.: Progress in remote sensing recognition and spatio-temporal changes study of urban and rural land use. J. Nat. Resour. **31**(04), 703–718 (2016)
2. Xu, Z.N., Gao, X.: A novel method for identifying the boundary of urban built-up areas with POI data. Acta Geogr. Sin. **71**(06), 928–939 (2016)
3. Xu, H.Q., Du, L.P., Sun, X.D.: Index-based definition and auto-extraction of the urban built-up region from remote sensing imagery. J. Fuzhou Univ. (Nat. Sci. Ed.) **39**(05), 707–712 (2011)
4. Song, J.C., et al.: A method of extracting urban built-up area based on DMSP/OLS Nighttime data and Google Earth. J. Geo-Inf. Sci. **17**(06), 750–756 (2015)
5. Li, Z., Yang, X.M., Meng, F., Chen, X., Yang, F.S.: The method of multi-source remote sensing synergy extraction in urban built-up area. J. Geo-Inf. Sci. **19**(11), 1522–1529 (2017)
6. Zhong, S.Y., Li, X.X., Bai, Y.H., Feng, J.: The method of extracting built-up areas based on multi-scale segmentation and spectral features. Softw. Guide **17**(09), 180–184 (2018)
7. Tong, B., Shen, W.: Object-oriented Landsat 8 image of the city proper extraction method research. J. Liaoning Prov. Coll. Commun. **19**(02), 21–25 (2017)

8. Zha, Y., Gao, J.Q.: Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int. J. Remote Sens. **24**(3), 583–594 (2003)
9. Hu, Y.Y.: Investigation of a modified normalized built-up index and a post processing scheme for BUILT-UP extraction in urban area. Geomat. Sci. Technol. **5**(3), 83–92 (2017)
10. Xu, H.Q.: A study on information extraction of water body with the modified normalized difference water index (MNDWI). J. Remote Sens. **9**(5), 589–595 (2005)
11. Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring vegetation systems in the great plains with ERTS. NASA Spec. Publ. **351**, 309 (1973)
12. Douglas, W., Stuart, R.P., Alan, T.M.: Monitoring growth in rapidly urbanizing areas using remotely sensed data. Prof. Geogr. **52**(3), 371–386 (2000)
13. Xu, H.Q., Tang, F.: Analysis of new characteristics of the first Landsat 8 image and their eco-environmental significance. Acta Ecol. Sin. **33**(11), 3249–3257 (2013)
14. Mac, Q.J.: Some methods for classification and analysis of multivariable observation. Comput. Chem. **4**, 257–272 (1967)
15. Duan, M.X.: Research and Application of Hierarchical Clustering Algorithm. Master, Central South University (2009)
16. Liaoning Statistical Department: Liaoning Investigation Team of National Statistical Bureau: Liaoning Statistical Yearbook 2018. China Statistical Press, Shenyang (2018)