

## Data fusion of high-resolution satellite imagery and GIS data for automatic building extraction

GUO Zhou <sup>a, \*</sup>, LUO Liqun <sup>a, b</sup>, WANG Wanyi <sup>a</sup>, DU Shihong <sup>a</sup>

<sup>a</sup> Institute of Remote Sensing and GIS, Peking University, Beijing 100871

<sup>b</sup> Unit 61243 of PLA, Urumqi 830006

([guozhou2012@gmail.com](mailto:guozhou2012@gmail.com); [luo1128\\_2001@sina.com](mailto:luo1128_2001@sina.com); [wangwany1118@gmail.com](mailto:wangwany1118@gmail.com); [dshgis@hotmail.com](mailto:dshgis@hotmail.com));

**KEY WORDS:** high-resolution satellite imagery, building extraction, GIS, object-oriented, QuickBird

### ABSTRACT:

Automatic building extraction in urban areas has become an intensive research as it contributes to many applications. High-resolution satellite (HRS) imagery is an important data source. However, it is a challenge task to extract buildings with only HRS imagery. Additional information and prior knowledge should be incorporated.

A new approach building extraction is proposed in this study. Data sources are QuickBird imagery and GIS data. The GIS data can provide prior knowledge including position and shape information, and the HRS image has rich spectral, texture features. To fuse these two kinds of features, the HRS image is first segmented into image objects. A graph is built according to the connectivity between the adjacent image objects. Second, the position information of GIS data is used to choose a seed region in the image for each GIS building object. Third, the seed region is grown by adding its neighbor regions constrained by the shape of GIS building.

The performance is evaluated according to the manually delineated buildings. The results show performance of 0.142 in miss factor and detection percentage of 89.43% (correctness) and the overall quality of 79.35%.

### 1. Introduction

Automatic extraction of man-made objects in urban areas such as roads, buildings, bridges and other impervious surface has been intensive research for many years. As one of the most prominent features in urban environment, obtaining accurate building objects are essential to many applications, such as urban planning, landscape analysis, cartographic mapping, and map updating, etc.

In earlier days, the works on building extraction focused on aerial imagery due to its high spatial resolution. With high-resolution satellites successfully launching, HRS imagery has been gradually used for building extraction. The HRS imagery provides fine detail of urban areas. With the high geometric accuracy and high spatial resolution, it is possible to identify the individual buildings. Wei et al. (2004) used the clustering and edge detection algorithm to detect the edges of candidate buildings in QuickBird panchromatic imagery and then extracted buildings from the detected edges. Jin and Davis (2005) demonstrated a strategy combining structural, contextual, and spectral information to extract building

in the city of Columbia with IKONOS satellite imagery.

Although HRS imagery provides a more detailed description of the observed scene, the complexity and diversity make it more difficult to interpret (Duan et al., 2004). The task of automatic building extraction on HRS imagery is difficult mainly due to three reasons: scene complexity, incomplete cue extraction and sensor dependency (Sohn and Dowman, 2007).

It is a hard to extract buildings automatically with high accuracy solely based on HRS imagery. Additional data and prior knowledge should be incorporated. There have been a few recent studies extracting buildings with a combination of HRS imagery and other data sources. Sohn and Dowman (2007) combined IKONOS imagery and LiDAR data to extract buildings in Greenwich area and got fine extraction results. Duan et al. (2004) proposed a approach based on fuzzy segmentation. The input data were QuickBird imagery and digital GIS map. Vosselman (2002) combined Laser scanning data, maps and aerial photographs for building reconstruction, in which purposes of using maps were: 1) to locate buildings in both laser scanning data and aerial

---

\* Corresponding author

photographs, 2) to reveal the information about the structures of buildings.

The study aims at developing an efficient and accurate approach to extract buildings from HRS imagery and GIS data. The GIS data is acquired by mobile measuring vehicles. In order to take advantage of the HRS imagery and GIS data, the approach first segments the HRS imagery into objects. A graph is built to manipulate image objects more efficiently according to the connectivity between the adjacent image objects. Second, GIS data is used to provide reliable position information of buildings. For each building object in GIS data, an image object is selected as a seed respectively. Third, the seed polygon is grown by adding its neighbor image objects. The overlapping degree between image objects and GIS data, and the shape information of GIS data are used to determine whether the neighbor objects belong to the same buildings. Finally, the performance of the automated approach is evaluated using the manually delineated buildings.

## 2. Test dataset

For the current research, a pan-sharpened multi-spectral (PSM) QuickBird image is used. It covers a portion of Beijing City with the size of 1.7km×0.9km. The PSM image is produced by combining the multi-spectral data with the corresponding panchromatic data, and resampled with 0.61-metre ground pixel. Fig. 1 (a) is the natural color composite of the selected QuickBird image. As can be seen in Fig. 1(a), the QuickBird image shows a typical urban environment with schools, residential areas, commercial district and industrial park. The buildings are distributed densely and vary greatly in their sizes and structures.

In the same study area, a GIS vector map of buildings is also obtained. The map is produced by mobile measuring vehicles (Fig.1 (b)). The outlines of buildings in GIS map and QuickBird imagery cannot completely coincide with each other due to the following reasons: First, the representations of the same buildings in GIS data and HRS imagery are different. Several adjacent buildings which are functionally consistent may be represented as one, while in image, they are still recorded individually.

Second, the projections in two data are different. The GIS map is produced by using the upright projection, while imagery by oblique photography. Therefore, GIS data represents the bases of buildings, but imagery encodes the side wall of buildings. Third, GIS map and imagery have different accuracies, deformations and registration errors. The uncertainties are usually not more than 0.5 meter. Thus allowing a simple building location (Vosselman, 2002).

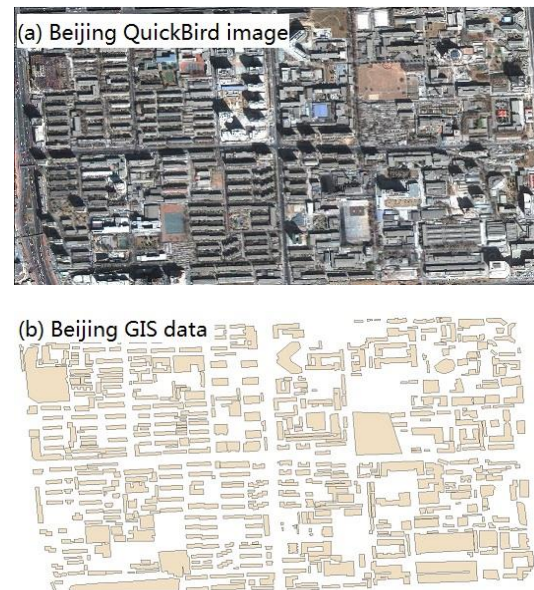


Fig.1 Beijing test datasets

## 3. Methods

Fig.2 illustrates the workflow of the automatic building extraction used in this study. It is pre-assumed that both data have been geo-referenced. The workflow consists of four major steps: segmenting QuickBird image, creating image object graph, selecting seed polygons and growing seed region.

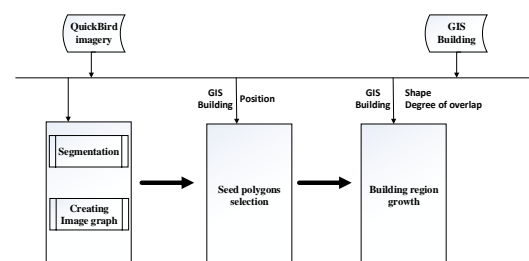


Fig. 2 The flowchart of building extraction

### 3.1 QuickBird image segmentation

Object-based image analysis (OBIA) methods are widely used in high-resolution image processing. Studies (Myint et al., 2011; Hay and Gastilla, 2008)

have showed that OBIA methods are more efficiently than per-pixel approaches in identifying urban land cover classes on HRS imagery. OBIA methods use image objects as the analysis units instead of single pixels. It can better utilize image features such as shape, size, texture and context.

The image objects is generated by using the multi-resolution segmentation in eCognition Developer 8.7. The segmentation starts from one pixel, and merges small segments into bigger ones until a heterogeneity threshold is reached (Benz et al., 2004). The heterogeneity thresholds are defined by users. The quality of segmentation is highly related to the input thresholds. In order to find an optimal scale, we carried out 10 segmentations by using different parameters ranging from 30 to 150. Earlier experiments showed that scale parameter smaller than 30 produces over-segmented objects, while scale parameter larger than 150 produces more under-segmented objects. The best segmentation is selected by visual interpretation. Color/shape weights are 0.7/0.3 instead of the default value 0.9/0.1, for considering more shape information. Smoothness/compactness are set to 0.5/0.5 so as to not favor either compact or non-compact segments. We carry out the image segmentation procedure using eCognition Developer 8.7, and export the image objects in the SHP file.

### 3.2 Image object graph

In order to manipulate the image objects efficiently, a graph  $G(V,E)$  is created, in which each image object  $V_i$  is represented as a node (Fig. 3). The edge connecting two nodes represents that two image objects are neighbors and share boundary. The building region growth process starts from one node and proceeds on its neighboring nodes, which will be discussed in detail in the Section 4.3.

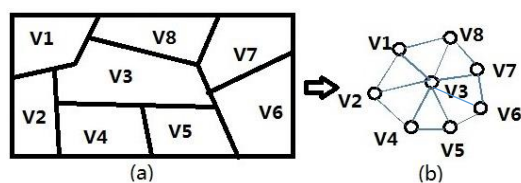


Fig. 3 Image objects and corresponding graph

### 3.3 Seed polygons selection

A seed polygon is an image object which belongs

to a building to be extracted. A building in the image can be represented by a seed polygon and other image objects in the segmentation image. Let  $P = \{V_i\}_{i=1}^n$  be the set of polygons segmented from the image. Let  $R = \{R_j\}_{j=1}^m$  be the set of building polygons in the GIS map. Each building in the image is represented by one  $R_j$  in the GIS map, and by a subset of  $P$  in the segmentation image.

The GIS data can be used to identify seed polygons. For each building in the GIS map, we select an image object whose centroid has the nearest distance with the centroid of GIS building. The centroid of a polygon  $V_i$  refers to the centroid of its minimum bounding rectangle (MBR), denoted by  $Cen(V_i)$ . Therefore, the distance between  $R_j$  and  $V_i$  is equal to the distance between  $Cen(R_j)$  and  $Cen(V_i)$ , denoted by  $D(R_j, V_i)$ .

For each  $R_j$ ,  $V_k$  is selected from  $\{V_i\}_{i=1}^n$  as the seed polygon using Eq. (1).

$$D(R_j, V_k) = \min\{D(R_j, V_i)\}_{i=1}^n \quad (1)$$

### 3.4 Seed region growth

A building in the image is represented by its seed building polygon and adjacent image objects in the segmentation image. The building region growth procedure starts from the seed polygon, and merges its adjacent objects into bigger ones until the merged objects fit best with the building objects in the GIS building map. The procedure of region growth is described as follows:

- i) For each building in the GIS map, two sets are created. The first is set T, it stores the image objects that have past the testing procedure and suitable for building region growth. The second is set S, which stores image objects that are candidates of building objects and need to be tested later.
- ii) Put the seed image object  $V_i$  into the set T, and the neighbors of  $V_i$  into the set S.
- iii) If set S is not empty, go to the iv); else end the procedure.
- iv) Fetch an object  $V_j$  from set S. If  $V_j$  meets the condition for region growth,  $V_j$  is put into the set T, and the neighbor objects of  $V_j$  into the set S; otherwise, go to the iii).

Criteria are helpful to determine whether to grow an image object. Here two indicators are used.

### 3.4.1 Overlapping degree

Overlapping degree is used to measure the similarity of area between image objects and vector polygons in GIS map. The degree of overlap (DOV) between an image object  $V_i$  and a polygon  $R_j$  is defined as follows:

$$DOV(V_i, R_j) = \frac{Area(V_i \cap R_j)}{Area(V_i)} \quad (2)$$

Where  $Area(V_i \cap R_j)$  denotes the overlapping areas between  $V_i$  and  $R_j$ . The larger  $DOV(V_i, R_j)$  is, the larger the similarity is. The value of  $DOV(V_i, R_j)$  ranges from 0 to 1.

### 3.4.2 Boundary distance

Boundary distance is to measure the boundary similarity between image objects and polygons in GIS building map. Let  $Q_i = \{Q_{ik}\}_{k=1}^{m_i}$  be the point set on the boundary of object  $V_i$ , and  $T_j = \{T_{jt}\}_{t=1}^{n_j}$  be the point set on the boundary of polygon  $R_j$ . The boundary distance (BD) between  $V_i$  and  $R_j$  is calculated as follow:

$$BD(V_i, R_j) = \frac{\sum_{s \in Q_i} \min_{t \in T_j} dis(s, t)}{m_i} \quad (3)$$

Where  $\min_{t \in T_j} dis(s, t)$  is the minimum distance value between the point  $s$  and  $T_j$ .  $m_i$  is the number of points in the boundary of  $V_i$ . In Fig.4,  $s$  is a point in  $Q_i$ ;  $t$  is a point in  $T_j$ , and has the nearest distance with point  $s$ . The distance from point  $s$  to the boundary of  $R_j$  equals one from point  $s$  to point  $t$ . The smaller  $BD(V_i, R_j)$  is, the larger the similarity is.

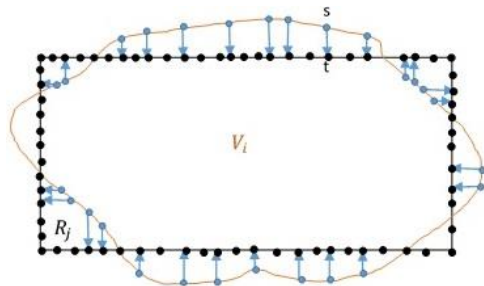


Fig. 4 Boundary distance between  $V_i$  and  $R_j$

### 3.4.3 Region growth criterion

The two indicators above are used to reveal the similarity between the image objects and the GIS objects. To determine whether an image object  $V_i$  can meet the condition for region growth, we first calculate

the overlapping degree between  $V_i$  and its corresponding GIS object  $R_j$ . If  $DOV(V_i, R_j)$  is too small (e.g. 0.2),  $V_i$  is abandoned. If  $DOV(V_i, R_j)$  is large enough (e.g. 0.8), we grow  $V_i$ , and put it into the identified set T. If  $DOV(V_i, R_j)$  is neither too large nor too small (e.g. larger than 0.2, and smaller than 0.8), the boundary distance is used to make a further judgment.

Before  $V_i$  is added into T, the image objects in T are merged into a larger object, called  $V_T$ . The boundary distance between  $V_T$  and  $R_j$  (i.e.  $BD(V_T, R_j)$ ) is calculated. Then,  $V_T$  and  $V_i$  is merged into a larger object, called  $V_{T+i}$ . The boundary distance between  $V_{T+i}$  and  $R_j$  (i.e.  $BD(V_{T+i}, R_j)$ ) is calculated as well. The value of  $BD(V_T, R_j)$  and  $BD(V_{T+i}, R_j)$  is compared. If  $BD(V_T, R_j)$  is smaller than  $BD(V_{T+i}, R_j)$ , it means that the boundary similarity between  $V_T$  and  $R_j$  is larger, and  $V_i$  is abandoned. Otherwise, the boundary similarity between  $V_{T+i}$  and  $R_j$  is larger, and  $V_i$  is added into set T.

## 4. Experiment results and discussions

The aim of this section is to assess the accuracy of building extraction results. Firstly, the building extraction results are presented. Secondly, the performance of the proposed building extraction approach is evaluated with reference to the manually delineated buildings. Finally, possible errors and inadequacies are analyzed.

### 4.1 Building extraction result

The developed approach is applied to the Beijing dataset and 515 buildings are automatically extracted. Fig. 5 shows extracted building results. As shown in Fig. 5, most of the buildings in the image are successfully extracted. The outlines of extracted buildings are accurate (Fig. 6).



Fig. 5 The extracted buildings extraction results.





Fig. 6. Outlines of extracted buildings (the red ones).

#### 4.2 Quality assessment

Manually delineated buildings are considered as reference building polygons for accuracy assessment. The automatically extracted and the reference buildings are compared pixel-by-pixel. Each pixel in the image fall into the four categories (Lee et al., 2003):

**True Positive (TP):** The pixel is labeled as building in both the automatically extracted and delineated maps.

**True Negative (TN):** The pixel is labeled as non-building in both the automatically extracted and delineated maps.

**False Positive (FP):** The pixel is labeled as building in the automatically extracted map, and as non-building in the delineated map.

**False Negative (FN):** The pixel is labeled as non-building in the automatically extracted map, and as building in the delineated map.

Based on the four categories above, the following statistical measures are employed to assess quantitatively the accuracy of the extracted buildings:

$$\begin{aligned}
 \text{Branching Factor: } & \frac{FP}{NP} \\
 \text{Miss Factor: } & \frac{FN}{TP} \\
 \text{Completeness (\%): } & 100 \times \frac{TP}{TP+FN} \\
 \text{Correctness (\%): } & 100 \times \frac{TP}{TP+FP} \\
 \text{Quality Percentage (\%): } & 100 \times \frac{TP}{TP+FN+FP}
 \end{aligned} \quad (4)$$

The branching factor indicates the rate of incorrectly labeled building pixels (FP), while the missing factor shows the rate of missed building pixels (FN). These two factors are closely related to the boundary delineation performance of the automatic building extraction algorithm (Sohn and Dowman, 2007). The completeness is the ratio of the correctly labeled building pixels (TP) with respect to the total building pixels by the manually delineated building map ( $TP + FN$ ). The correctness is the ratio of the correctly labeled building pixels (TP) with respect to the total building pixels by the automatically extracted building map ( $TP + FP$ ). These two factors indicate a

measure of building detection performance (Sohn and Dowman, 2007). The quality percentage combines aspects of both boundary delineation accuracy and building detection rate. It describes how likely a building pixel produced by the automatic approach is true.

Table 1 shows the pixel classification results by comparing the automatically extracted building map with the delineated building map. Based on the results, the statistical measures are calculated by Eq. (4). The proposed approach has accuracies of 0.118 in branching factor, and 0.142 in missing factor. It means that the number of incorrect pixels is less than the one of the missed building pixels. The automatic process detects building at the rate of 87.56% in completeness, while 89.43% in correctness. Finally, the overall accuracy of the proposed method is evaluated as 79.35%.

**Table 1** Pixel classification results

Pixel classification	Number of pixels
True positive (TP)	1273200
True negative (TN)	2626768
False positive (FP)	150512
False negative (FN)	180832

#### 4.3 Error analysis

The quality assessment shows a satisfactory result. It also produced a certain number of building extraction errors, including misclassified buildings and missed buildings, which should be reduced to improve the extraction result. The errors may be caused by the following factors:

(1) Image segmentation. Although a series of scale parameters were tested, the image segmentation result was still not perfect. Especially for buildings with complex structures and buildings with low contrast to the ground, it is difficult to obtain fine building objects with the multi-resolution segmentation algorithm. The result of image segmentation directly impacted the accuracy of building extraction. Building extraction result can be greatly improved with a better image segmentation algorithm.

(2) GIS building map error. The proposed building extraction method uses high-resolution imagery and GIS data as data sources. The GIS data cannot

completely coincide with the high-resolution image. There may have some errors existing in the GIS building map. The final building extraction results are highly relevant to the GIS building map accuracy. The higher accuracy the GIS building map is, the better extraction results we get.

## 5. Conclusions

This paper presented a new building extraction approach which combines HRS imagery and GIS data. First, the high-resolution image was segmented into image objects by using the multi-resolution algorithm. A graph was built according to the connectivity between the adjacent image objects. Second, the position information of GIS data was used to select a seed polygon for each GIS building object. Third, the seed polygon was extended by adding its neighbor objects. Two indicators, i.e. overlapping degree and boundary distance, were defined as criteria to grow seed regions. Finally, the accuracy of the building extraction result was evaluated with reference to manually delineated buildings. This evaluation showed that the branching factor was 0.118, the miss factor 0.142, the completeness 87.56%, the correctness to 89.43% and the quality percentage 79.35%. The overall quality of the result were satisfactory. According to the performance, it can be concluded that combining the HRS imagery and GIS data is an effective method to extract buildings. However, the error analysis shows that some errors can be reduced further. The image segmentation algorithm which combines high-resolution image and GIS data can be studied in the further research. Furthermore, the error analysis demonstrates that this technique can automatically detect the difference between the GIS data and the high-resolution image, which can be used in change detection and map updating.

## Acknowledgements

The work presented in this paper was supported by the Program for New Century Excellent Talents in University (NCET-10-0189).

## References

Benz, U.C., et al., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of*

*Photogrammetry and Remote Sensing*, 58(3), pp. 239-258.

Hay G J, Castilla G, 2008. *Object-Based Image Analysis*, Springer Berlin Heidelberg, pp.75~89.

Jinghui, D., P. Veronique and L. Hanqing, 2004. Building extraction in urban areas from satellite images using GIS data as prior information: *IEEE*, pp. 4762~4764.

Jin, X. and C.H. Davis, 2005. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Applied Signal Processing*, pp. 2196-2206.

Lee, D.S., J. Shan and J.S. Bethel, 2003. Class-guided building extraction from Ikonos imagery. *Photogrammetric Engineering and Remote Sensing*, 69(2), pp. 143-150.

Mayer, H., 1999. Automatic object extraction from aerial imagery—a survey focusing on buildings. *Computer vision and image understanding*, 74(2), pp. 138-149.

Myint, S.W., et al., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5), pp. 1145-1161.

Sohn, G. and I. Dowman, 2007. Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1), pp. 43-63.

Vosselman, G, 2002. Fusion of laser scanning data, maps, and aerial photographs for building reconstruction: *IEEE*, pp. 85~88.

Wei, Y., Z. Zhao and J. Song, 2004. Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection: *IEEE*, pp. 2008~2010.