OBJECT-ORIENTED BUINDING EXTRACTION BY DSM AND VERY HIGH-RESOLUTION ORTHOIMAGES

N. Jiang [a,b,*], J.X. Zhang [a], H.T. Li [a], X.G. Lin [a,c]

[a] Chinese Academy of Surveying and Mapping, Beijing 100039, P.R. China
[b] Shandong University of Science and Technology, Qingdao 266510, P.R. China
[c] Wuhan University, Wuhan 430079, P.R. China - jiangna123321@163.com

Commission III, WG III/4

KEY WORDS: building extraction, high resolution, segmentation, DSM, object-oriented, eCognition

ABSTRACT:

High-resolution remote sensing images accelerate the development of information extraction and 3D city reconstruction. This paper concentrates on object-oriented methods to solve building extraction problems. With object-oriented methods, not only the spectral information but also the shape, contextual and semantic information can be used to extract objects. The object-oriented building extraction typically includes several steps: data pre-processing, multi-scale image segmentation, the definition of features used to extract buildings, building extraction, post-processing and accuracy evaluation. Among the features of buildings extraction, we consider the height of the buildings as the most useful information. In this paper DSM is used to extract high objects as trees and buildings; then the work of distinguishing trees and buildings will be applied. With the removal of trees, building information is extracted. We achieve this goal in the software package eCognition. The result shows that the method works well.

1. INTRODUCTION

1.1 General Instruction

High-resolution imagery provides an important new data source for building extraction and 3D reconstruction. How to extract topographic objects as buildings, roads, trees and pipes in urban areas form images automatically, rapidly and accurately, now has become a hot spot of imagery information extraction and application. The need for 3D structure of buildings has evolved most rapidly in the last years (Förstner, 1999).

Detection and description of buildings from aerial and spatial images is a practical application of 3-D object description. Building extraction is also the key problem in urban information updates construction of digital cities, large-scale mapping, and also motivated by the importance of geographic information systems (GIS): the need for data acquisition and update for GIS.

A couple of years ago the main input data for the production of building extraction and 3D city models were aerial images, terrestrial images, map data, and data derived from classical surveying (Fuchs et al, 1998), so the main feature used is the spectral and textural information in the images. However, with the development of data acquisition methods, multi sensory data, e.g. SAR, infrared, stereo or laser scan images, is available as additional information; the image processing methods have developed, so different cues as color, texture, semantics, edges and color edges elevation data features can be used to detect and reconstruct buildings. Different input data combining with different processin [1] g means lead to different models. Parametric models and generic models are used to describe the buildings; DSM and DEM also appear to solve these extraction and reconstruction problems. Both semiautomatic and automatic methods are applied to building extraction; the semiautomatic way that allows for efficient human interaction can meet the high precision but the automatic procedures seem to be the only way to satisfy the developing trend in the future.

The development tendencies of the building extraction according to Jiang(2004) are: From single image to multiple images; From gray information to color information; Interaction between 2-D and 3-D; From single image information grouping to grouping with multiple images; More imaging geometry, object knowledge and spatial reasoning are used; Building model develops from single to complex building of plane patches; Multiple sources information integration, such as with LIDAR data or map data.

1.2 Overview of Related Work

A brief literature overview concerning building extraction from image data is given here. Many of the early building extraction systems have used a single intensity image. Nevatia (Nevatia et al, 1997)described a method for detecting rectilinear buildings and constructing their 3-D shape descriptions from a single aerial image of a general viewpoint. They use the geometric and projective constraints to make hypotheses for the presence of building roofs from the low-level features and to verify by using 3D cues. Shadows, wall vertical and base line are important cues in these methods. Then stereo or multi-view analysis is focused because of widely available data. In(Roux et al, 1994)3D descriptions of buildings are generated from matched lines and junctions.

DSM, DEM and LIDAR data are used in the extraction of buildings. Gerke(Gerke et al.) have done a lot of work on automatic detection and extraction of trees and buildings from aerial CIR orthoimages and normalized digital surface models. They use a hierarchical strategy to solve the complex models and complex images problem. Orhner and Descombes (Orhner

* Corresponding author.

et al., 2002) extract buildings from Digital Elevation Model(DEM); Huber et al(Huber et al.) use LIDAR (LIght Detection And Ranging) as an attractive alternative for the building extraction and acquisition of 3D information.

The object-oriented building extraction methods have developed quickly, and a lot of work has been done. Mϋller (Mϋller et al, 2005) presented a building extraction approach that divides the method into a low-level image processing step and a high-level feature extraction and final classification step. They use seeded-region growing algorithm to segment the entire image, then take area, mean hue angle as preselection features to reduce calculation time, finally classify based on robust features as size, form, hue angle, neighbourhoods, and shadow. Pasko and Gruber (1996) gave an approach that incorporates a priori knowledge for buildings extraction. In Jiang(2004) multiple high-resolution images and Digital Surface Model(DSM)are combined to extract the urban road grid in complex, stereotypical, residential areas. In our approach, we proposed an object-oriented building extraction method by DSM and orthoimage. In Section2, the main steps of building extraction are given. Section 3 illustrates an example for building extraction and precision evaluation. Conclusions are given in Section4.

2. OBJCET-ORIENTED BUILDING EXTRACTION

Pixel-oriented processing approaches deal with each pixel, so when the resolution improved, the number of pixels consisted in the building increased sharply. Object extraction is time consuming. Also, Salt and Pepper Noise phenomenon in classification and extraction is evident and not easy to get rid of. Of all the disadvantages, the utmost one is it has excessive dependence on the spectral information. Object-oriented method can solve these problems well. It can take the full advantages of high-resolution of remote sensing images, integrate the semantic, contextual, and spectral information, and mine the hidden information sufficiently. At the same time, it deals with objects that are consisted of pixels with the same grey level. It reduces time consuming of computation. What's more object-oriented methods work more in line with people's cognitive thinking habits and characteristics. In this paper, object-oriented approach is selected to extract building information from urban areas.

The object-oriented building extraction from high-resolution images typically includes several steps: data pre-processing, multi-resolution image segmentation, the definition of the characteristics used to delineate the buildings, building extraction, post editing and accuracy evaluation. The flow chart of the paper is shown as below:

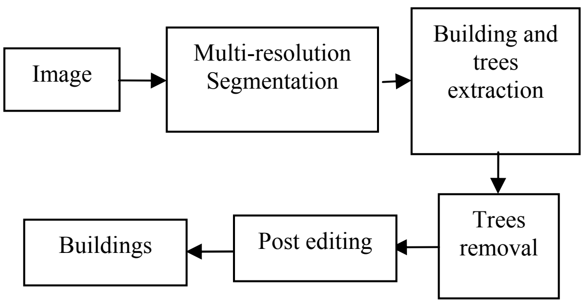


Figure1. The flow chart of building extraction

2.1 Image Segmentation

Detection, segmentation and classification of specific objects are the key building extraction blocks of object-oriented approach in image analysis area. Image segmentation is the first step and also one of the most critical tasks of image analysis (Zhang, 2006). In image processing, segmentation is the subdivision of a digital image into smaller partitions according to given criteria. The fundamental step of image analysis is a segmentation of an image—into image object primitives.

Segmentation algorithms are used to subdivide the entire image represented by the pixel level domain or specific image objects from other domains into smaller image objects. For segmentation, the basic algorithms like chessboard, quadtree, or multi-resolution segmentation usually are used. We emphasize multi-resolution segmentation. It is a bottom-up segmentation algorithm based on a pairwise region merging technique, and it is an optimization procedure which, for a given number of image objects, minimizes the average heterogeneity and maximizes their respective homogeneity.

The segmentation procedure works according the following rules, representing a mutual-best-fitting approach (Definiens, 2007):

A. The segmentation procedure starts with single image objects of 1 (one) pixel size and merges them in several loops iteratively in pairs to larger units as long as an upper threshold of homogeneity is not exceeded locally. This homogeneity criterion is defined as a combination of spectral homogeneity and shape homogeneity. Scale parameters can be modified manually through calculation. Higher values for the scale parameter result in larger image objects, smaller values in smaller image objects.
B. As the first step of the procedure, the seed looks for its best-fitting neighbor for a potential merger.
C. If best-fitting is not mutual; the best candidate image object becomes the new seed image object and finds its best fitting partner.
D. When best fitting is mutual, image objects are merged.
E. In each loop, every image object in the image object level will be handled once.
F. The loops continue until no further merger is possible.

The homogeneity criterion is calculated as a combination of color and shape properties of both the initial and the resulting image objects of the intended merging. Here the color homogeneity is based on the standard deviation of the spectral colors. The shape homogeneity is based on the deviation of a compact or a smooth shape.

Figure 2. Weighted components of the homogeneity criterion (Definiens, 2007)

The new image objects created by segmentation are stored in what is called a new image object level. Each image object is defined by a contiguous set of pixels, where each pixel belongs to exactly one image object. Each of the subsequent image object related operations like classification, reshaping, re-segmentation, and information extraction is done within an image object level. Simply said, image object levels serve as internal working areas of the image analysis.

For successful image analysis, defining object primitives of suitable size and shape is of utmost importance. As a rule of thumb, good object primitives are as large as possible, yet small enough to be used as building blocks for the objects to be detected in the image. Pixels are the smallest possible building block, however pixels have limited information. To get larger building blocks, different segmentation methods are available to form contiguous clusters of pixels that have larger property space. (Definiens, 2007). The segmentation scale parameter should be decided through the repeated practice.

2.2 Extraction of Building and Trees by DSM

Through the segmentation, we avoid to deal with pixel level but to deal with characteristic level. After the segmentation of images, now we deal with objects, not pixels. On this basis, there are many features can be used to extract buildings, like shape, form, orientation of the building and the ratio of length/width. But they often can not recognize buildings very well. DSM appears to solve the building extraction problems due to the data source requires much less model knowledge (Förstner, 1997). We use DSM as a key factor to extract buildings from surroundings.

A Digital Surface Model (DSM) is an image consisting of height values including vegetation, buildings and other objects (Gerke et al.). The DSM supplies a good way to solve 3D problems in terms of building extraction, because the DSM has the information about the height, so it can distinguish buildings and roads, other low objects very well.

Trees, buildings and other objects with large height, which in the DSM image appear high brightness, will be classified into the same group. Experiments are needed to find the optimal height parameter (in the DSM height usually appears as intensity) as threshold to classify. It works very well though some trees are also involved in.

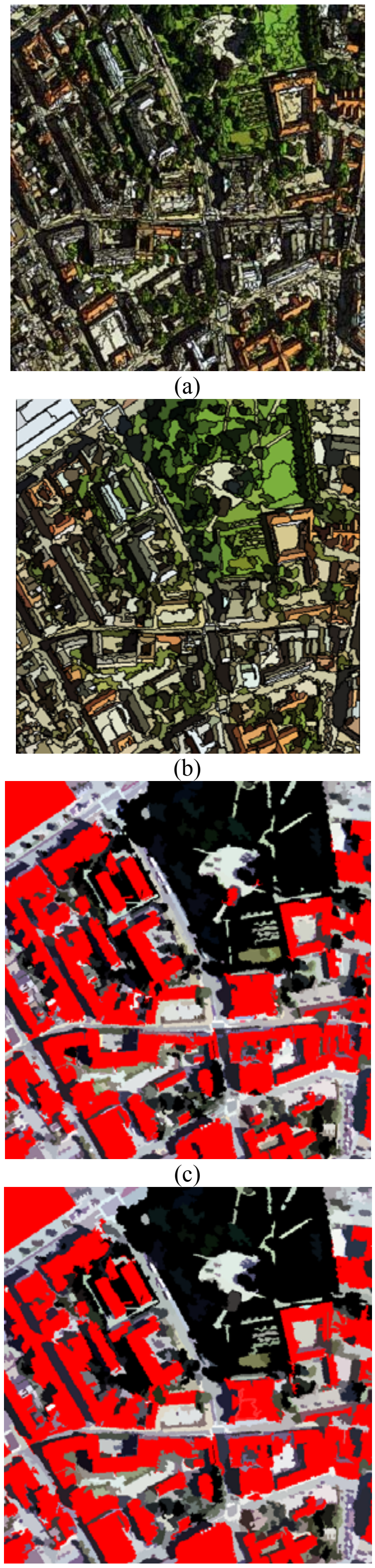
2.3 Building Extraction

After using DSM to delineate the objects with high elevation, buildings and trees are grouped together. To reach a better delineation of the building areas, a ratio factor is needed. NDVI is the most useful factor to extract trees, e.g. in (G. Markus et al.). When the NDVI is not available (there is no NIR band), VI or other factors are involved. Through experiments, we found that the ratio of Green Band/Red Band is quite useful. Some other combinations are also available, e.g. the ratio of (Green Band layer value- Red Band layer value) / (Green Band layer value +Red Band layer value). With this, trees and buildings are separated. Then use mask to filter trees, only buildings are left.

3. RESULTS

In our experiments, we use the software package eCognition to extract buildings. Definiens Developer is the product of Definiens Corporation with a powerful Integrated Development Environment (IDE) for rapid image analysis solution development. It has done quite well in object-oriented segmentation and classification.

3.1 Multiresolution Segmentation



(a) (b) (c) (d)

Figure 3. Different scale parameters leading to different results. (a) scale parameter: 45; (b) scale parameter: 60; (c) result of a; (d)result of (b). The red area delineate building. (The original data supplied by Definiens Corporation)

The first step is multi-resolution segmentation. The right scale of parameter is crucial. When segmenting in Color mode with high radiometric resolution, it's better to choose a larger scale parameter for generating image objects. In our example, a high-resolution aerial photograph was used for segmentation. Image objects like buildings, roads and types of land covers required large scale parameters. If the scale is too small, the segmentation will generate many fragmentations while the scale is too large to omit some useful details. Through trials, we chose 60 as final decision. The edge of DSM usually is quite rough, so it did not involve in the segmentation step. Nevertheless, the information contained in the elevation image can be well used for discriminating elevated rooftops from low lying roads. Therefore, DSM just can be utilized in the classification. In this step, its weight was set to 0.

The result of multi-resolution is shown in figure3. From figure3 we can see, the large-scale segmentation can describe buildings well.

3.2 Classification

In this step, we take DSM as thematic information to extract buildings and trees by the classification function of eCognition. Height information is used as a filter. Based on this, image objects can be analyzed according defined criteria and assigned to classes that best meet these criteria (Definiens, 2007). In "Class Hierarchy" view we define a new class named *con&tree*, which mainly contains buildings and trees. In eCognition, in the "*Feature View*" window, take layer *means->layer 1*(this is the DSM layer), and then find the right scale parameter. Use "*class description*" to insert the classification expression, then classify. After the procedure of classification, each image object is assigned to "*con&tree*" or not—so classes connected with the class hierarchy.

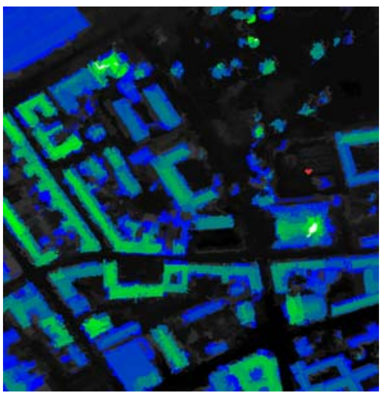


Figure 4. The elevation information in DSM (Green represents high value while blue represents low)

3.3 Building Extraction

In this step we separate trees from "*con&tree*". As we know, kinds of ratio index are quite useful to extract vegetation. For the lack of NIR band, we use the ratio Green Band layer value / Red Band layer value instead of NDVI to extract trees. The *building*'s expression uses "similar to" *tree* but in the "invert to" form. The entirety of image objects is organized into a hierarchical network, and the "*tree*" and "*building*" classes inherit from "*con&tree*".

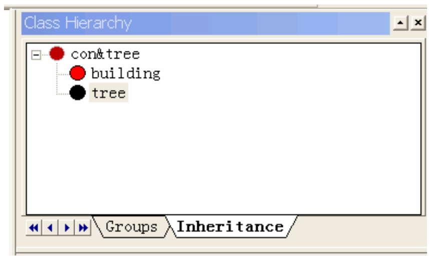


Figure 5. Class Hierarchy window

In the end, as we can imagine, the green area represents the trees, and the red area represents buildings. The exhausted work of distinguishing buildings, roads and other similar objects are avoided in this approach.

3.4 Evaluation

Two measurements for a detection evaluation as described in (Lin et al, 1998) were made:

$$\textit{Detection percentage} = 100 \cdot TP/(TP + TN)$$

$$\textit{Branch factor} = 100 \cdot FP/(TP + FP)$$

The two measurements are calculated by making a comparison of the manually detected buildings and the automated results, where TP(True Positive) is a building detected by both a person and the program, FP(False Positive) is a building detected by the program but not a person, and TN (True Negative) is a building detected by a person but not the program. A building is considered detected if any part of the building is detected; an alternative could be to require that a certain fraction of the building be detected. The analysis is based on the image of 35 buildings. Among these, 35 buildings are detected with 1mistaken involved, and 1 omitted. The DP is 97.1% and the BF is also 97.1%. The figure is relatively high in contrast with the data obtained by other methods.

4. CONCLUSIONS

This paper proposes an object-oriented method to extract building information by DSM and orthoimage in eCogniton. It extracts buildings in only two main steps and reaches quite high accuracy. Also there are some points should be paid attention to: The object-oriented method proposes a way leading us to think in a new manner. Within this, the spectral, semantic and contextual information can be combined together to mine the image information. Although many methods to extract buildings are developed in recent decades, the most useful information in the extraction is spectral information. With segmentation, the structure of level and inheritance is constructed, and the time consuming of computation has been cut down.

Building extraction with DSM has quite a lot of limitations. It is required flat terrain, and when the area is large, the extraction result is not good. To solve this problem, we can use a normalized DSM (nDSM) which is the difference DSM-DTM. Here, a DSM is an image consisting of height values including vegetation, buildings and other objects. A DTM consists of only those points lying directly on the terrain. Another method is to

subset the image into several small images and deals with them one by one.
In the segmentation step, the select of parameter scale is very important. The weight of DSM layer is set to 0. It means that the DSM layer doesn't involve in segmentation. That's because the edge of buildings in DSM layer usually has low contrast with the background, and also because the low resolution of DSM. If with DSM in segmentation layer, the edges of the buildings in the extraction become rather rough.

Because we have adopted the object-oriented method, we deal with the objects; the edge of extraction is not as smooth as pixel-based method. But this can be overcome through post-processing, like using rectangle restriction to make the edge regular. Also, building extraction by DSM and high-resolution imagery is robust, but its popularity is not so good as the model-based building extraction.

ACKNOWLEDGEMENT

This work was developed within the National Key Basic Research and Development Program under grant No.2006CB701303.

REFERENCE

Definiens, UserGuide, 2007

Fu, K. S., Mui, J.K. 1981. A survey on image segmentation. Pattern Recognition, 13, 3-16.

Förstner, W.1999. 3D-City Models: Automatic and Semiautomatic Acquisition Methods, Photogrammetric Week, Wichmann Verlag, Heidelberg, pp291-303.

Gerke, M., Straub, B. M., Koch, A. Automatic Detection of Buildings and Trees from Aerial Imagery Using Different Levels of Abstraction. http://cmsv021.rrzn.uni-hannover.de/uploads/tx_tkpublikationen/gerke_straub_koch_dgpf.pdf (accessed 10 Sep. 2007)

Huang, H.P. 2003.Scale Issues in object-oriented Image Analysis, Ph.D dissertation.

Huber, M., Schickler, W., Hinz, S., Baumgartner. A. Fusion of LIDAR Data and Aerial Imagery for Automatic Reconstruction of Building Surfaces, http://www.sanborn.com/Pdfs/Article_Fusion_of_Lidar_Schickler.pdf(accessed 10 Sep. 2007)

Jiang, W.S. 2004. Multiple Aerial Image Matching and Automatic Building Detection, Ph.D dissertation

Lin,C., Nevatia, R. 1998. Building detection and description from a single intensity image. *Computer Vision and Image Understanding*: CVIU 72(2), pp. 101–121.

Mathisa, O., Xavier, D., Josiane, Z. 2002. Building Extraction from Digital Elecation Model. INRIA Report.

Müller, S., Zaum, D. W. 2005. Robust Building Detection in Aerial Images, IAPRS, Vol.36, part3/W24. Vienna,Austria, August29-30

Nevatia, R., Lin, C., Huertas, A. 1997. A System for Building Detection from Aerial Images, Monocular Building Detection Ascona Workshop ,Ascona Switzewrland, Apr. 1997.

Pasko, M., Gruber, M., 1996. Fusion of 2D GIS Data and Aerial Images for 3D Building Reconstruction. IAPRS, Vol. 31, Part B3, pp. 257-260.

Roux, M., McKeown, D. 1994. Feature Matching for Building Extraction from Multiple Views. Proc. IEEE *Computer Vision and Pattern Recognition*, Seattle, WA, pp. 46-53.

Zhang, Y.J. An Overview of Image and Video Segmentation, 2006,http://image.ee.tsinghua.edu.cn/Papers/E123=AIVS-06b.pdf(accessed 15 Sep. 2007)