AN INTEGRATED SYSTEM ON LARGE SCALE BUILDING EXTRACTION FROM DSM

Y. Li [a, *], L. Zhu [b], H. Shimamura[b], K. Tachibana[b]

[a] International Institute for Earth System Science, Nanjing University, 210093 Nanjing Jiangsu, P.R. China – liyan@nju.edu.cn
[b] Pasco Corp., 1-1-2 Higashiyama, Meguro-ku, Tokyo, 153-0043, Japan– (lin_zhu, hideki_shimamura, kikuo_tachibana)@ pasco.co.jp

Commission III, WG III/4

KEY WORDS: Building extraction, DSM, Morphology, Level set, Watershed

ABSTRACT:

Large scale automatic building extraction is difficult in Japan, because of the high density of buildings and various roof types especially in the residential area. We propose an automatic and integrated system for building extraction from DSM, in order to model the outlines of the buildings in 2D plane. In our previous research, the building extraction based on foreground/background marker gives good result and most buildings can be detected. However, it is still difficult to model the building because the shape is not accurate enough. The active contour is a kind of segmentation method widely used in medical image processing. It merges edge detection into gray segmentation so that the contour converges at the right edge of the certain object. It can be employed locally on the DSM to obtain the outline of a building if the initial contour is given to close to the building. Thus we developed a scheme of accurately locating the building by marker controlled watershed and accurately segmenting the building by active contour.

There are two basic processes in this scheme. First is the initial building segmentation, or the building location. Watershed of the gradient of the DSM can get the objects' contour. To avoid over segmentation, the markers of the terrain and the off-terrain must be labelled and set to local minimal. There are three types of markers to be labelled, including two of buildings and one of terrain. The methods of marker labelling for small houses and big houses are different. The markers of the small houses are decided by local maximum of DSM, while those of the big or high buildings are extracted using area morphology method. The terrain markers are detected by the watershed of the DSM not crossing the building markers. The marker controlled watershed segmentation on the gradient of the DSM gives the contours of the off-terrain objects. Then, each building object is used as a mask by enlarging its size, and active contour method is implemented to segment the masked DSM to obtain the subtle contour of the building. Level set as a typical solution of active contour is employed because of its fast iterating process. The contour of the level set can be converged exactly at the outline of the building for a local DSM including only one house. RGB image can be used instead of DSM, but the result is not as good as that of DSM. These methods and scheme construct a system of large scale and automatic building extraction. The experiment in Saitama, Japan is given as the evaluation of the proposed system.

1. INTRODUCTION

1.1 General Instructions

Feature detection and extraction from remote sensed data has been a hot point in geographic information science. 3D model have found its applications in urban landscape, environmental agencies (Narushige, 2001), 3D GIS (Köninger, 1998), architecture, urbanism and engineering (Eastman, 1999), ttraffic and road (Ravanbakhsh, 2009), disaster management etc. Building as the most important feature in the urban, the extraction methods and approaches of it have been studied and developed fast. In great lots of researches, imagery integrated with Digital surface model (DSM) are used as the data source for building extraction (Nakagawa 2002, Chen 2004). DSM only supply the height information for the objects thus no shadow or colour change will give difficulty on the algorithm. The disadvantage of DSM is its accuracy especially in distinguishing the fine structure of the house and the crowded houses. Imagery as the auxiliary data can compensate its shortcoming.

A technique of building detection is based on classification. The points of the DSM have to be classified according to whether they belong to the terrain, to buildings or to other object classes, e.g., vegetation (Rottensteiner et al. 2002).

Unfortunately, we found it is difficult to get a satisfying result to distinguish the vegetation from the building in DSM. Therefore, we decide using colour feature derived from the imagery instead of DSM to partition the vegetation from the building.

1.2 Outline of the proposed approach

In this paper, we will describe an integrated approach of multiple segmentations to extract the precise footprints of the buildings. It is a refining procedure from rough to accurate. First is the initial building segmentation, or the building location. Watershed of the gradient of the DSM can get the objects' contour. To avoid over segmentation, the markers of the terrain and the off-terrain must be labelled and set to local minimal. There are three types of markers to be labelled, including two of buildings and one of terrain. The methods of marker labelling for small houses and big houses are different. The markers of the small houses are decided by local maximum of DSM, while those of the big or high buildings are extracted using area morphology method. The terrain markers are detected by the watershed of the DSM not crossing the building markers. The marker controlled watershed segmentation on the gradient of the DSM gives the contours of the off-terrain objects. Then, each building object is used as a mask by enlarging its size, and active contour method is implemented to segment the masked DSM to obtain the subtle contour of the building. Level set as a typical solution of active contour is employed because of its fast iterating process. The contour of the level set can be converged exactly at the outline of the building for a local DSM including only one house. RGB image can be used instead of DSM, but the result is not as good as that of DSM. These methods and scheme construct a system of large scale and automatic building extraction.

2. BUILDING MASK EXTRACTION

DSM is a kind of height data in term of grid and can be taken as gray level image if we take the height of the grid as the gray scale of the pixel. The off-terrain objects including buildings and trees and other things have higher values than the terrain and show themselves bright in gray scale. Level-set segmentation (Osher 2002) as a kind of active contour can give a closing pixel-wise contour of the object. It is superior to general edge based segmentation. However, it is actually a binary segmentation. If there are more than one object in the image and the gray scale characters of the them are not consistent vs. the background, the segmentation may not certainly give each object a proper contour. To introduce level set in large scale remote sensing imagery for building extraction, a scheme of multiple segmentations is proposed in proposed in this paper. It is a refining procedure from rough to accurate. Watershed is employed to get the mask for each building, and level set segmentation is carried on to each windowed image including the building mask. The mask extraction procedure is shown in Figure 1.

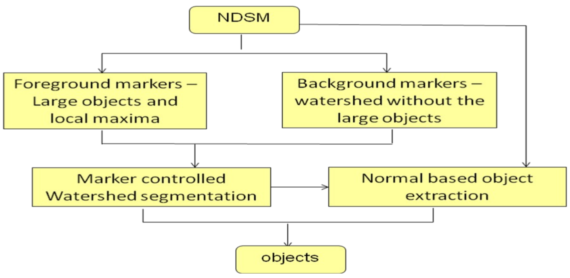


Figure 1 Flow chart of object mask extraction

Watershed to gray level image is suitable for the gray level images to separate the objects each others, but not exactly the object contours will be obtained. Watershed to gradient of the image can resolve this problem and give the contour of the object.

However, in practice, this transform produces an important over-segmentation due to noise or local irregularities in the gradient image. To avoid the over-segmentation, a marker controlled watershed is introduced (Gao 2001, Salembier 1994). Here, the watershed segmentation is implemented to the gradient of NDSM (GNDSM). A marker is an area which is the initial of a catchment basin. By giving each off-terrain object and the ground a marker, and making it the catchment bases, the building or tree etc. can be segmented from the DSM.

In large scale of urban area, there are various kinds of buildings. Central business district (CBD) and administration district usually have many great buildings, much higher or huger than most others. We define these buildings as large objects. On the other hand, in the residential area, there is a great amount of houses crowded with each other. They are generally lower and small as single objects. We define them as the small objects. Besides, there are some areas blending with these two kinds of buildings. The markers for the two kinds of buildings are given in different ways.

2.1 Marker extraction for large object

In our former work, a refined marker-controlled watershed segmentation was proposed (Li 2010). The large regions were detected in the first to avoid multi-marker of the woods and huge buildings. If not, there would be several catchment bases detected for a huge building and therefore it will be taken as several objects. The extraction procedure is improved and described as Figure 2.

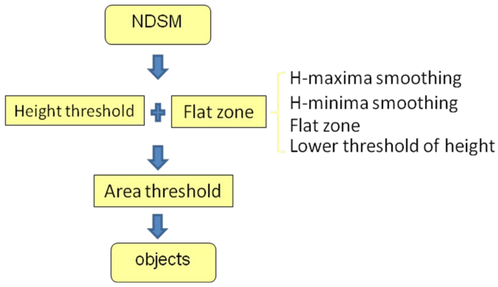


Figure 2. Flow chart of Large region extraction

Height threshold is used to segment the quite high buildings from the Normalized DSM (NDSM). Flat zone (Salembier 1995) refers to a region which has the same value in NDSM. They are detected for all height or gray level in NDSM, and a relevant low value is used as the threshold to extract the flat zones above a certain height. These zones sometimes correspond to the buildings with flat roofs, no matter high or not.

2.2 Marker extraction for small object

Buildings, trees, and other off-terrain objects are taken as the foreground objects and are assigned the foreground markers. The large objects include higher or bigger buildings or wood have been extracted by last step. The extraction of markers for those small houses is quite simple. The local maxima of the NDSM excluded with the large objects are taken the small object markers.

2.3 Vegetation detection

As we mentioned before, the large objects include higher or bigger buildings as well as wood. To detect wood from DSM is difficult and unreliable. Thus multi-spectral imagery is employed here for vegetation detection from the off-terrain objects. We use NDVI and image texture together to detect vegetations.

According to the reports, NDVI is an efficient measurement for vegetation detection in remote sensing imagery. NDVI values reflect the health of the vegetation. Healthy vegetation has a higher NDVI value than unhealthy vegetation (Xiao, 2005). We expect the NDVI can be of help in vegetation detection in aero base remote sensing. On the other hand, both of the resolutions of RGB and infrared imagery is 20cm in our case, thus the texture of trees is obviously distinguished from that of man-made objects. So we take texture as another measurement for vegetation detection.

2.3.1 NDVI parameter

NDVI is defined as the ratio of the difference of infer-red and red to the addition of them (Tucker, 1979). The values of NDVI range from 0 to 1. According to the literatures, for remote sensed multiple spectral image especially by the satellite, NDVI>0 correspond to the vegetation pixel. Generally the lower NDVI correspond to non vegetations. We have computed NDVI for our aero based imagery data and found that the NDVI distribution of the objects is a little different from reported. The NDVIs of the trees overlap with that of other objects. For some pixels of trees the NDVI are less than 0 or even lower. These trees are always the ones with brown colour which may lose their leaves in that season. In the mean while some houses may have NDVI greater than 0. Many of them are roofs in shadow. Obviously, lower threshold of NDVI will cause more trees detected and more buildings taken as trees. Higher threshold of NDVI will cause less trees detected but less buildings taken as trees.

2.3.2 Texture parameter

Now that NDVI is not sufficient to distinguish the vegetation from other objects, we have to consider other feature for this objective. Texture is a kind of feature usually used to classify various objects in image interpretation. Connective density of the singular edges (CDSE) of the gradient image is defined to describe the texture roughness of an object. After normalization of the RGB image to 0~1, the gradients of 0.05-0.1 correspond to strong edges, including trees' contours and buildings' contours. Most of that of 0.02-0.05 correspond to the singular edges within trees. Thus excluding of the strong edges and then binary segmenting the remained edges will get the pixels of trees.

So firstly the edge points that its maximum gradient component in row or column beyond a certain value G are assigned as 1, otherwise as 0. We use $I_o(i,j)=l$ denotes the pixel (i,j) within an object l. For the pixel of the terrain, $I_o(i,j)=0$. I_e is defined as the edge image.

$$I_e(i,j)=\begin{cases}1 \text{ for } g(i,j)>G \\ 0 \text{ otherwise}\end{cases} \tag{1}$$

where $g(i,j)$ refers to the maximum gradient of pixel (i,j). It is obviously that a lot of edge points locate at the boundary of the buildings and connect each other. There are some edge points within the tree which are discrete. Using hole filling operator to filter the edge image, the inner of the building will be totally full filled while the tree not. I_{ef} is defined as the filtered image of I_e. Define a connective density parameter C of object l as

$$C=\frac{N_c}{N} \tag{2}$$

$$N_e=\sum_{\{i,j|I_o(i,j)=l\}} I_{ef} \tag{3}$$

$$N=\sum_{\{i,j|I_o(i,j)=l\}} 1 \tag{4}$$

where Nc refers to the number of singular points within the object, and N refers to the area in terms of pixels of the object. A tree object is detected by the ratio of tree pixels less than 0.2.

3. REFINING OF BUILDING CONTOUR

Using the buildings extracted by watershed segmentation as masks, the windowed DSM is segmented again by level-set method to get more subtle shape of the building.

3.1 Level set

Active Contour which is also called Snake Model is proposed by Kass, Witkin and Terzopoulos in 1988 (Kass 1988). It is a minimization question:

$$E=\int_0^1 \frac{1}{2}\left[\alpha|x'(s)|^2+\beta|x''(s)|^2\right]+E_{ext}(x(s))ds \tag{5}$$

- α - tension
- β - rigidity

E_{ext} means external energy. A typical external energy coming from the image is positive on homogeneous regions while near zero on the sharp edges.

Level set is a kind of solute for active contour question. It is proposed by S.Osher and J. Sethian in 1989. It embeds the contour into a higher dimensional space. The advantage is that it automatically handles the topological changes. The disadvantage is the slower evaluation. $\phi(x,y)$ is a level set function. It gives the implicit contour C, inner and outer component respectively, as defined in the following equations.

$$\begin{aligned} C&=\{(x,y)\in\Omega:\phi(x,y)=0\} \\ in(C)&=\{(x,y)\in\Omega:\phi(x,y)>0\} \\ out(C)&=\{(x,y)\in\Omega:\phi(x,y)<0\} \end{aligned} \tag{6}$$

Chan and Vese model (Chan, 2001) is a solution of active contour using level set concept. The model is based on trying to separate the image into regions based on intensities. Define function

$$F(c_1,c_2,C)=\mu L(C)+\lambda_1\int_{in(C)}|u_0(x,y)-c_1|^2 dxdy+\lambda_2\int_{out(C)}|u_0(x,y)-c_2|^2 dxdy \tag{7}$$

It is also a minimization question. c_1 and c_2 are the average intensity levels inside and outside of the contour, which is separated by level set $\phi(.,t)=0$. $\phi(.,0)$ is initialized randomly and is evoluted in each iteration t. The contour is evolved implicitly by moving the surface ϕ.

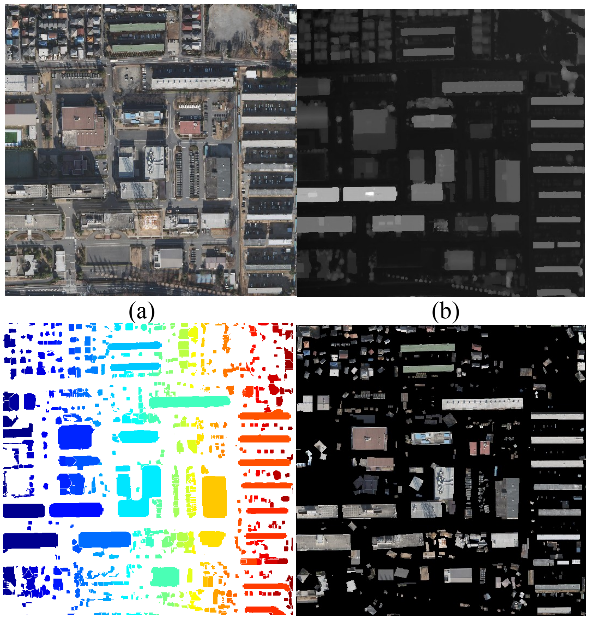
3.2 level-set segmentation on DSM and RGB

Level-set is equivalent with binary segmentation in the meaning of gray level. But general binary segmentation doesn't give any objects, instead, it gives pixel segmentation result. The advancement of level-set is that it segments the object from the background. So it has two characters, binary segmentation and object segmentation. It is usually used in medical imagery and physical imagery, because the object is sure and clarified from the background in gray level.

For remote sensing data, such as DSM, if a building is certain in a local region, level-set is useful for the segmentation of it. For colour image, level-set method sometimes doesn't give very good result. For example, some building roof is not uniform in colour due to lighting condition, and some building colour even is similar with the surrounding ground. In this situation, level-set segmentation will segment part of the building as the object or the whole building and the surrounding ground as one object. The level-set segmentation on DSM+RGB is very similar with that using RGB.

3.3 Experiments

The experiment data is UCD image and the derived DSM of Japan in 2007. Figure 3 (a) is a sample of RGB image, (b) is the associated DSM, (c) shows the building masks in colour map and (d) shows the building footprints by level set overlapped on RGB image. The level set segmented buildings persist in square shapes and sharp corners which make modelling easy in the following processing.



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

Figure 3. Building extraction experiments of a sample

This is an area mixing of big buildings and some residential. The extraction of the houses for the two kinds of buildings indicate that our approach is robust for complex urban distribution.

4. CONCLUSION AND FURTHER PROSPECT

Level-set segmentation as a kind of active contour can give a closing pixel-wise contour of the object. It is superior to general edge based segmentation. However, it is actually a binary segmentation. If there are more than one objects in the view field and the gray scale characters of the them are not consistent vs. the background, the segmentation may not certainly give each object a proper contour. To introduce level set in large scale remote sensing imagery for building extraction, a scheme of multiple segmentations to extract the precise footprints of the buildings in proposed in this paper. It is a refining procedure from rough to accurate. Watershed is employed to get the mask for each building, and level set segmentation is carried on to each windowed image including the building mask.

The building mask is derived from the DSM. Since DSM has one channel which presents the height signals for the terrain including the building etc. and is easy to detect buildings according to their spatial distribution characters in height. Level-set can be implemented on DSM or RGB imagery, or both. The experiments show that level set gives better building footprints on DSM than on RGB image.

References:

Chan T. F., and Vese L. A., 2001, Active contours without edges, *IEEE Transactions on image processing*, 10(2), pp.266-277.

Chen L.C., Teo T.A., Shao Y.C., Lai Y.C., and Rau J.Y., 2004, Fusion of LIDAR data and optical imagery for building modeling, In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXV(Part B4), Istanbul, Turkey, pp. 732-737.

Eastman C. 1999. *Building Product Models: Computer, Environments Supporting Design and Construction*. CRC Press.

Gao H.,Siu W.C. , and Hou C.H.. 2001, Improved techniques for automatic image segmentation, *IEEE Transactions on Circuits and Systems for video technology*, 11(12), pp.1273-1280.

Kass M., Witkin A. P. and Terzopoulos D., 1988. Snakes: Active Contour Models, *International Journal of Computer Vision*, 1(4): 321-331.

Köninger A., Bartel S. 1998. 3D-GIS for Urban Purposes, *Geoinformatica*. 2(1), pp.79-103.

Li Y., Zhu L., Gong P., Shimamura H., 2010. A refined marker controlled watershed for building extraction from DSM and imagery, *International Journal of Remote Sensing*, 31(6), pp. 1441-1452.

Nakagawa M., Shibasaki R., and Kagawa Y., 2002, Fusing stereo linear CCD image and laser range data for building 3D urban model, *in Symposium on geospatial theory, processing and applications*, *ISPRS*. Ottawa, Canada, pp.222-211.

Narushige S., 2001. Urban models: Recent developments in the digital modelling of urban environments in three-dimensions, GeoJournal 52, pp.263-269.

Osher, S., and Fedkiw, R., 2002. *The Level Set Method and Dynamic Implicit Surfaces*, Springer-Verlag New York, Applied Mathematical Sciences 153.

Ravanbakhsh M., Fraser C. S., 2009, Road roundabout extraction from very high resolution aerial imagery, *CMRT09*. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*,, Vol. XXXVIII(Part 3/W4), Paris, France.pp.19-26.

Rottensteiner, F. and Briese, C., 2002, A new method for building extraction in urban areas from high-resolution LIDAR data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV(Part 3A), Graz, Austria, pp.295–301.

Salembier P, and Pardas M., 1994, Hierarchical morphological segmentation for image sequence coding, *IEEE Transactions on Image Processing*, 3(5), pp.639-651.

Salembier P. and Serra J., 1995. Flat Zones Filtering, Connected Operators, and Filters by Reconstruction, *IEEE Transactions on Image Processing*, 4(8), pp. 1153-1160.

Tucker, C. J., 1979, Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sensing of Environment,* 8(2), pp.127-150.

Xiao Q. and McPherson E. G., 2005. Tree health mapping with multispectral remote sensing data at UC Davis, California, *Urban Ecosystems*, 8(3-4), pp.349-361.

Satellite imaging corporation, 2001-2010, http://www.satimagingcorp.com/svc/homeland_security.html(a ccessed 28 April. 2010)