TOWARDS AUTOMATED HOUSE DETECTION FROM DIGITAL STEREO IMAGERY FOR GIS DATABASE REVISION

Zhongchao SHI Ryosuke SHIBASAKI
Institute of Industrial Science, The University of Tokyo
Email : shizc, shiba@shunji.iis.u-tokyo.ac.jp
Commission IV, Working Group 3

KEY WORDS : GIS, Urban, Database Revision, Automation.

ABSTRACT

To keep GIS database current, it is crucially important to capture the changes of man-made structures, especially houses, from digital urban imagery. In this paper, we describe a method aiming at extracting houses automatically from stereo imagery of urban areas for GIS database revision. We begin with the discussion on feature extraction based on wavelet analysis. Image segmentation techniques involving region segmentation and line feature extraction are then discussed. Region-based and line-based stereo matching algorithms for the purposes of disparity estimation, house detection, and change detection of houses for GIS database revision are addressed, respectively. Experimental results are demonstrated at the end of this paper.

1. INTRODUCTION

One of the most challenging approaches in the domain of GIS database revision is the automated 3D structure extraction, especially the house and road extraction, from complex photographs of urban or suburban scenes.

Since the pioneering work of Nagao and Matsuyama (1980) in which structural analysis is applied for establishing fully automated system, a multitude of automated structure extraction methods have been proposed and tested, such as shadow-analysis based algorithms(Liow and Pavlidis, 1990) and information-fusion based systems(Mckeown, 1991). Many of them, however, employed image shadows as a obvious information for predicting the presence of houses and estimating the height of houses. Moreover, hypotheses are often exploited in these existing studies, such as the roof of each house should have corners of 90 degrees(i.e., the even edges of a house should be parallel with one another and similar for odd edges, while the even edges have to be perpendicular to the odd edges), there should exist shadows on photographs, and so on. The question is that : when there are no explicit or complete shadows or/and there exist very complex and irregular roofs, can these algorithms work perfectly ? It seems that most of them are not robust enough to resolve such problems. This is the major motivation for the research to be introduced later in this paper.

In order to solve above problems, we believe that the knowledge of *depth* of the scenes in aerial images is significant in the absence of shadows, the presence of irregular roofs and occlusions. In practice, the depth information is generally obtained through stereo analysis(Marr and Poggio, 1977; Grimson, 1985). It indicates that stereo matching will be the major technique for developing our system. In addition, for the purpose of man-made structure extraction, such as houses and roads, from complex aerial imagery in urban scenes, we also believe that the feature-based stereo-matching techniques are superior to the area-based ones because of the higher matching efficiency and accuracy are expected.

Moreover, we assume that house change has the state of either emergence or demolition. It indicates that two kinds of house change detection algorithms corresponding to emergence and demolition detection will be studied in this research.

According to above considerations, a practical system was developed which involves mainly following three parts : *feature extraction*, *stereo matching and house extraction*, as well as *house change detection* for GIS database revision.

2. IMAGE PREPROCESSING

Image preprocessing which includes feature detection and image segmentation, is a very important step in our system. In this chapter, both the region and line segmentation techniques will be addressed.

2.1 Region Segmentation

Region boundaries are derived by employing Shi and Shibasaki's algorithm(1994), which consists of *multiresolution decomposition* of images based upon wavelet transform, *edge and corner detection* as well as Modulus Based Image segmentation(MOBIS).

In addition, in order to resolve the correspondence problem in region-based stereo matching, a metric for measuring the similarities or dissimilarities is needed. According to Marapane and Tridedi(1989), following six attributes are critical for assessing the similarities or dissimilarities between regions : mean gray level, area, perimeter, width of principal axis(PA, denoting an axis which is parallel to X axis as well as pass through the centroid of a region), height perpendicular to the PA, and width-to-height ratio. In our system, however, we exploit area-to-perimeter ratio in stead of width-to-height ratio. Experiments show that it can greatly reduce the ambiguities between irregular regions.

2.2 Line Feature Extraction

Line segmentation in our system, consists of following procedures :

1) Wavelet transform supported edge detection, involving edgels(edge pixels) and their orientations.
2) Contrast sign or zero-crossing sign computation.
3) Line segmentation based on Hough transform(see Duda and Hart, 1972).

The major property of our algorithm is that the zero-crossing sign is also taken into account besides the constraints of orientation and average contrast. That is, only these edgels, which have the same contrast signs of zero-crossings, are able to the candidates for forming a line. Hence, two distinct voting arrays will be produced corresponding to different contrast sign. One of the major advantages of the use of Hough transform is that the disconnectivity of lines can be resolved. In fact, there exist abundant occlusions in images from urban aerial scenes.

3. STEREO-MATCHING ALGORITHMS

It has been widely recognized that no single feature based stereo matching(e.g. either region-based matching or line-based matching) can provide enough information for 3D object reconstruction from complex urban scenes. A matching system which integrates both the region and line matching schemes is proposed and its basic framework is shown in Fig.1.

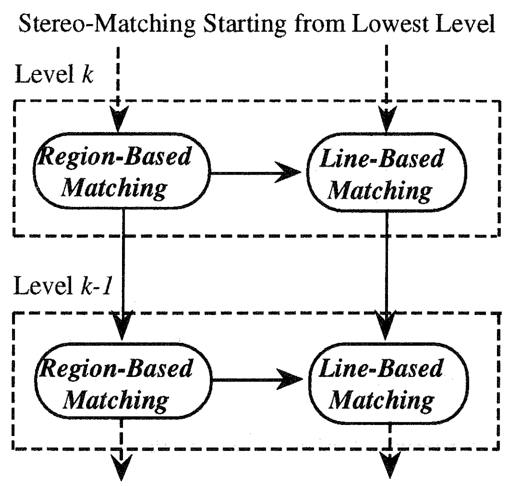


Fig.1. A framework of multi-level stereo matching algorithm.

According to Fig.1, the stereo-matching process starts from the regions at lowest level of pyramid because there exist the least information as well as noises, and therefore smallest ambiguities in matching scheme. The matching results at a lower resolution will be used to guide and fasten stereo-matching at a higher resolution. On the other hand, the results of region-based matching are applied to reduce the matching ambiguities of line-based matching at same scale.

Note that in order to impose an epipolar line constraint, it is reasonable to assume that the images used in our system have already been rectified, by no loss of generality.

3.1 Region-Based Stereo Matching

We would like to give some definitions at first :

1) Let I_l and I_r denote, respectively, left and right images of a stereo pair. R_l and R_r denote total regions in left and right images respectively, while p_i and q_j denote any individual region of left and right images, respectively. It indicates that $R_l = \{p_i\}$ and $R_r = \{q_j\}$.

2) d_{max} denotes the maximum value of disparity of the stereo pair. For each region p_i in I_l, the candidates of match in I_r are searched in a space defined by a parallelogram-shaped window $w(p_i)$ similar to which is defined and used by Medioni and Nevatia(1985).

Stereo matching is then performed as follows. For each region p_i in the left image I_l , the region(s) lying in searching window $w(p_i)$ in right image I_r are picked up as the candidate(s) for matching. A region with the "closest" attributes to the p_i is selected as the match. In case more than one candidates are close enough to a candidate, ambiguities should be resolved. Here is a simple method for sovling such a problem : 1) sliding each candidate along a direction parallel to the epipolar line; 2) computing the maximum overlapping area with in finite sliding range; 3) the candidate that leads to largest overlapping area is thought to be the match. If there still exist ambiguities after that process, the geometric order is suggested to be used to resolve them.

3.2 Line-Based Stereo Matching

Line-based matching is carried out for two purposes. One is to estimate the *ground height* (see the next section), while the other is to extract 3D lines above the ground, which can be important due for house extraction.

The line-based matching is performed by comparing the voting arrays by Hough transform between left and right images. The searching window for candidate selection is similar to what was defined in section 3.1. Candidates are then chosen in corresponding searching window. One will see in Chapter 4 that, for the purpose of ground height estimation, it is no need to search the ***unique*** match for each line. That is, the presence of ambiguities in line matching does not affect the accuracy of ground height very much.

4. GROUND HEIGHT ESTIMATION AND HOUSE HYPOTHESIS EXTRACTION

Last two chapters described how to segment images into meaningful regions and lines, and how to get their match pairs through stereo matching. Since the houses are generally located above the ground surface, if the height of ground surface is known, we can easily capture the house by simply comparing the height of each region pairs or 3D lines with the height of ground. It is then crucial to estimate the height of ground surface in order to extract houses from stereo images.

4.1 Ground Height Estimation

For the purpose of ground height estimation from the stereo images of urban areas, it is reasonable to assume that : if all pixels of extracted region boundaries(many of them are houses) are removed from the edge images, most of the left features should approximately lie on a flat plane(with certain tolerance), i.e., ground surface.

Based on that assumption, a computational algorithm aiming at the ground height estimation is proposed and developed by the authors. It can be described as :

1) Finding the minimum and maximum values of disparities d_{min} and d_{max}. Let $\Delta D = d_{max} - d_{min}$.

2) The range of disparity ΔD is equally divided into k intervals($k \in Z, k > 0$, generally k=10). Let $\Delta D = \sum_{i=1,k} \Delta d_i$.

3) A vote is given to a interval Δd_i when the disparity of a matched lines lies in disparity interval Δd_i.

4) The disparity of the "ground surface" is thought to be Δd_i where a voting peak is reached.

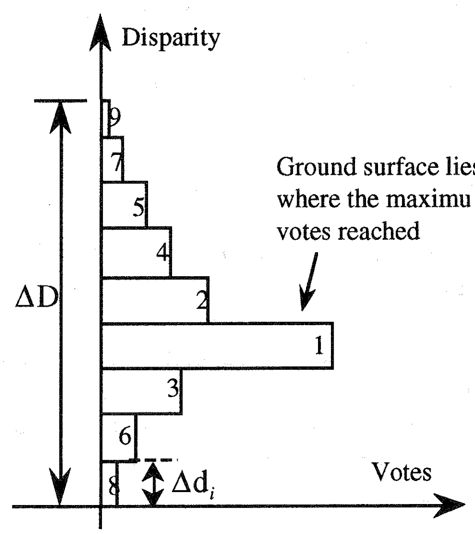


Fig.2. "Voting" for ground surface estimation : a new concept.

4.2 House Hypothesis Extraction

As mentioned before, after the ground surface was conformed, houses are able to be extracted simply by comparing the disparity or height of each region or line pair with that of the ground surface. These pairs which lie above the ground are thought to be the candidate elements constructing house hypotheses. House hypotheses can be extracted from following three models :

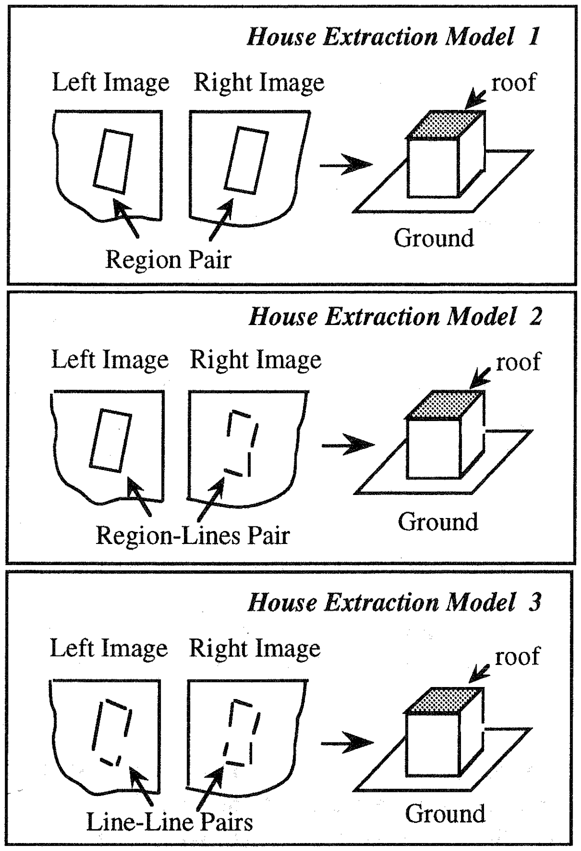


Fig.3. House hypothesis extraction models.

It is obvious that the possibility and reliability(P&R) of house extraction of above three models are different. Since the region-based process has the lowest ambiguity in stereo matching, model 1 has the highest P&R among above three models.

5. HOUSE CHANGE DETECTION FOR GIS DATABASE REVISION

As mentioned in Chapter 1, house changes are limited to two states in our current research : emergence and demolition. It means that we will not take into account the state of continuous change.

5.1 New-Built House Detection Model

These house hypothesis extracted from above three models which can not find their correspondences in existing database, are thought to be new-built houses. Since the new-built houses have no records in existing GIS database, the possibility and reliability(P&R) of new-built house detection is basically relied on the quality of house hypotheses extraction. Furthermore, as the house hypotheses captured from the reign-region pairs have the best quality and such hypotheses are generally affected by the threshold values in edge and region segmentation processes, the iterative process is generally needed for improving the detection possibility and reliability.

5.2 House Demolition Detection Model

For an existing house in GIS database, if there exists corresponding *house-like* region or region pair and their overlayed area is larger than a threshold T, we may say that the house is still there. Otherwise, we say that such a house has already disappeared. It is then easy to find the demolition houses by comparing the house hypothesis extracted from model 1 and model 2 with existing database.

However, it is unreasonable to try to find all demolitions from region pairs. Hence, 3D lines should be also an important source for demolition detection. A 3D line based extinction detection algorithm was then developed. It is summarized as follows :

1) seeking all possible candidates of 3D lines relevant to existing house A in GIS database. The possible candidates here are these 3D lines that close enough to an edge of a house(e.g., average distance < 3 pixels) and having relative angle less than 25 degrees.

2) matching each edge of existing house A with extracted 3D lines from step 1). In this matching scheme, the height of each 3D line is the major constraint for best match selection.

3) shifting house A with a finite tolerance and repeat steps 1) and 2) to find the best matches.

After above iterative process finished, for the existing house A, if there exists enough corresponding house-like 3D lines, we can say house A is still there. Otherwise, we can get the decision that "house A has disappeared".

6. EXPERIMENTAL RESULTS AND CONCLUSIONS

Experiments were performed to examine the accuracy of the methods described above. As the page limitation, only one of the testing results is demonstrated in this paper. The original photographs used in this experiment were taken in 1988 scaled about 1:5000. The images utilized here are 1000X1000 pixels with an range of intensity level from 0 to 255. The sequential experimental results are illustrated in Fig.4. The results show that the possibility and reliability of house extraction and house change detection is affected by the threshold values in edge extraction and about 90-95% of the houses are extrected without interative processes whereas the extraction rate may be increased at rate of 3-

5% when iterative process is performed. The details may also be seen from Shi and Shibasaki(1995).

We have presented some computational algorithms for the purpose of automated house change detection for GIS database updating, referring to image segmentation, multi-feature based stereo matching, statistical disparity voting for ground height estimation and so on. The major contribution of our research is that we employed the disparity differences or surface discontinuities in imagery to detect the significant man-made structures like houses and created a computational algorithm for estimating the ground height, which opened up a new path for automated object recognition.

The experimental results also show that many of the 3D lines which give votes for forming the "ground surface" in voting algorithm were found lying on the streets or roads. It indicates that we may detect the changes of roads by overlaying these 3D lines with existing road database. It will be one of our research topics in future.

7. REFERENCES

Duda, R. O. and Hart, P. E., 1972. Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM , vol. 15, no. 1, pp.11-15.

Grimson, W. E. L., 1985. Discontinuity detection for visual surface reconstruction. CVGIP 30, pp.316-330.

Haralick, R. M. and Shapiro, L. G., 1985. Image segmentation techniques. CVGIP 29, pp. 100-132.

Lew, M. S., Huang, T. S. and Wong, K., 1994. Learning and feature selection in stereo matching. IEEE Trans. on PAMI , vol. 16, no. 9, pp.869-881.

Li, D.R. and Shao, J. L., 1994. The wavelet and its application in image edge detection. ISPRS J. Photogramm. Remote Sensing , vol 49 , pp. 2-11.

Liow, Y. T., Pavlidis, T., 1990. Use of shadows for extracting buildings in aerial images. CVGIP, vol. 49.

Mallat, S. G., 1992. Singularity detection and processing with wavelets. IEEE Trans. IT , vol. 38, pp. 617-643.

Marapane, S. B. and Trivedi, M. M., 1989. Region-based stereo analysis for robotic applications. IEEE T. Sys. Man. and Cyber. vol. 19, no. 6, pp.1447-1464.

Marr, D., Hildreth, E., 1980. Theory of edge detection. Proc. Roy. Soc. London, vol.207, pp. 187-217.

Marr, D., and Poggio, T., 1977. A theory of human stereo vision. Artifi. Intell. Lab., Mass. Inst. Technol. Cambridge , Mass..

McKeown, D. M., 1991. Information fusion in cartographic feature extraction from aerial imagery. in : Digital Photogrammetric Systems, Wichmann, pp.103-110.

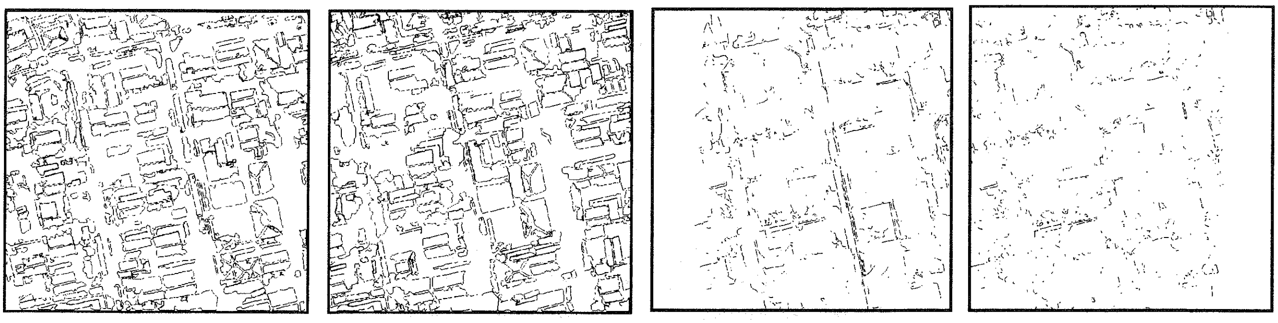
Medioni, G. and Nevatia, R., 1985. Segment-based stereo matching. CVGIP, vol.31, pp.2-18.

Nagao, M. and Matsuyama, T., A Structural Analysis of Complex Aerial Photographs, Plienum Press, New York and London, 1980.

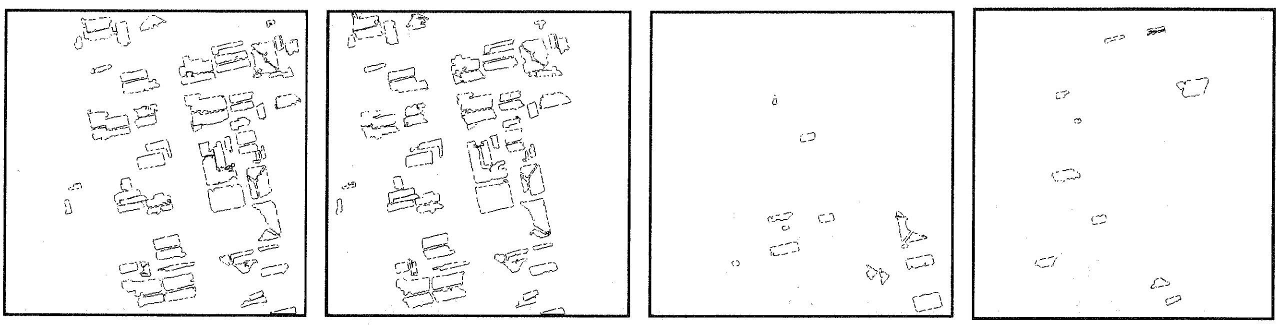
Shi, Z. C. and Shibasaki, R., 1994. Wavelet transform supported image segmentation. Proc. of the 15th ACRS, vol. II, Bangalore, India , pp.I-3-1 to I-3-6.

(a) Original left image

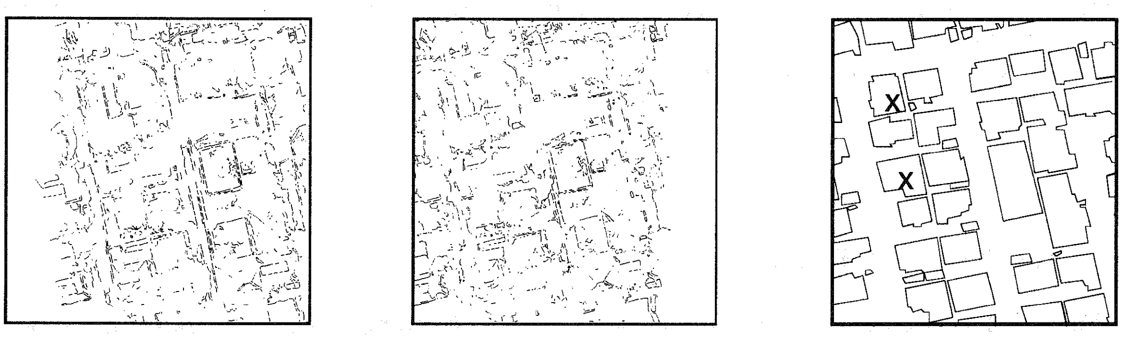
(b) Original right image



(c) Segmented regions(L) (d) Segmented egions (R) (e) 3D lines on the ground(L) (f) 3D lines on the ground(R)



(g) 3D regions above the ground (L) (h) 3D regions above the ground (R) (i) House extracted from region-lines pairs(L) (j) House extracted from region-lines pairs(R)



(k) 3D lines above the ground(L) (l) 3D lines above the ground(R) (m) Corresponding map data

Fig.4. Experimental results of house extraction(L--left, R--right). (a) and (b) are the original images involving a lot of buildings, roads, and a few trees as well as cars. (c) and (d) present the totally segmented regions from the left and right images, respectively. (e) and (f) show the 3D lines which lie on the ground surface. (g) and (h) illustrate the house hypotheses extracted from 3D regions, i.e., these 3D regions which lie above the ground. (i) and (j) show the house hypotheses extracted from region-lines pairs which are located above the ground. (k) and (l) are the 3D lines above the ground. (m) is the map data at the experimental area in which two houses with X sign were unable to be detected in this experiment.