MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

An Approach to Automated House Recognition from Digital Stereo Imagery

Zhongchao SHI * Ryosuke SHIBASAKI *

Institute of Industrial Science, University of Tokyo
7-22-1 Roppongi, Minato-ku, Tokyo 106 , Japan
Email : (shizc, shiba)@shunji.iis.u-tokyo.ac.jp

Abstract

One of the most critical steps in 3D GIS database establishment and revision as well as automated digital photogrammetry is to capture the 3D man-made structures from stereo images automatically and accurately. The key problem for fulfilling such a task is to resolve the *correspondence* problem between image pairs. The most challenging problems exist in the large scale imagery of dense urban or suburban areas where the structures are very complex, unsmooth and quite different in size. In this paper, we address an reliable and effective system aiming at reconstructing the ground surface and extracting houses with relative disparity from the complex aerial imagery in urban areas.

1. Introduction

In the domain of digital photogrammetry, people have moved their interest from the nonsemantic information collection to the automated photograph interpretation. One of the most challenging approaches in this domain should be house recognition from aerial photographs. 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 (Huertas and Nevatia, 1988; Liow and Pavlidis, 1990) and information fusion based systems (Mckeown, 1991).

* Address: 7-22-1 Roppongi, Minato-ku, Tokyo 106 Japan.

In this paper, we would like to introduce a new algorithm aiming at house extraction from complex imagery in urban areas by integrating both region and line based stereo matching. The system involves three parts : low level image processing, stereo matching, and house extraction with 3D features.

2. Low Level Image Processing

Low level image processing takes very important roles in our research. It includes wavelet based feature detection and image segmentation. Since both the region-based and line-based stereo matching will be performed in our system, the core of this subsystem is to segment the images into *meaningful* regions and lines.

2.1. Region Segmentation

Region boundaries are derived by employing the algorithm proposed by Shi and Shibasaki (1994), which consists of multi-resolution decomposition of images based upon wavelet analysis, feature (edge and corner) extraction, and modulus based image segmentation (MOBIS), etc.. In order to save pages, we will not discuss it in this paper.

2.2. Line Segment Extraction

Line segmentation in our system, consists of following procedures :

1) Wavelet transform supported edge detection, involving edgels and orientations.

2) Contrast sign or zero-crossing sign computation(see Grimson, 1985).

3) Hough transform based line segmentation (Duda and Hart, 1972), in which the contrast signs of zero-crossings is also taken into account except for the ρ and θ. That is, only these edgels, which have the same contrast signs of zero-crossings, are able to construct a line. Hence, two distinct voting arrays will be produced corresponding to different contrast sign. The definition of contrast sign is referred to Grimson(1985).

3. Stereo Matching Algorithms

In order to impose a epipolar line constraint, it is reasonable to assume that the images used have already been rectified by no loss of generality.

The primitives for matching in our system are regions, lines and corners. Fig.1 shows a proposed hierarchical stereo matching scheme. The stereo-matching starts from the regions at lowest level because there are the least information as well as noise attributions, and therefore smallest ambiguities at that case. The matching results at a lower resolution will be used to guide and fasten the performance of stereo-matching at a higher resolution. On the other hand, the results from a high level feature based matching(e.g. region based matching) will be used in the matching processes based on its low level features(e.g. line based matching) for resolving the match ambiguities.

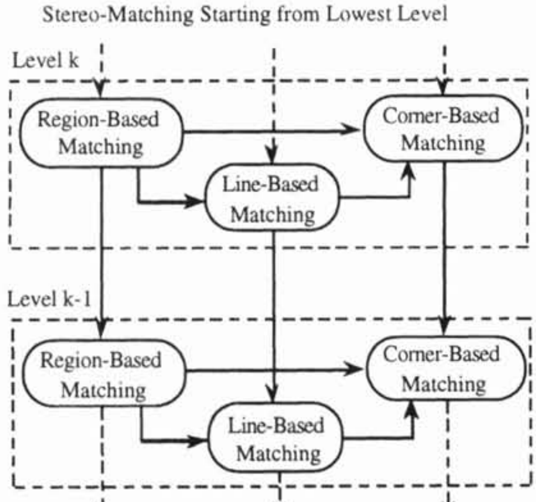


Fig.1 Hierarchical stereo matching algorithm.

4. House Recognition Models

Since the digital data sources utilized in our system are the complex photographs in urban areas and all the images have already been rectified, it is then reasonable to assume that : if all pixels of extracted region boundaries are removed from the edge images, most of the left features should approximately lie on a flat plane(with certain tolerance). Let us call such a flat plane the "ground surface". It is then obvious that the roofs of houses should locate above the "ground surface" in 3D space. Hence, the house recognition problem in this case can be simplfied as a problem to estimate the disparities of both the "ground surface" and feature matches.

Based on above consideration, an algorithm for estimating the disparity of "ground surface" was proposed firstly. The algorithm is 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.

After the ground surface was estimated, the houses can be then detected from following three basic models :

model 1 : 3D regions above the ground;

model 2 : 3D region-line pairs above the ground;

model 3 : 3D lines above the ground.

It is possible to improve the possibility and reliability of house detection by applying also the results from area-based matching.

5. Experiments

The algorithms introduced above were implemented on a workstation system. Experiments were performed to examine the efficiency and accuracy of the methods described above with real complex stereo images. We give out only one of the testing results in this paper.

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.2(the results at resolution 0 are exhibited in order for reducing the number of pages). Fig.2 (a) and (b) are the original images from an complex urban scene with a lot of houses, roads, and a few trees as well as cars(they were enlarged for seeing clear). The voting for ground surface estimation was performed with $K = 10$. The peak of voting was found at the disparity near -124. Fig.2 (e) and (f) show the 3D lines on the ground. According to the models given in Chapter 4, we can then extract the house hypotheses by just comparing the disparties of 3D features(3D regions etc.) with estimated disparity of the ground and picking out these 3D features whose disparities are smaller than that of the ground surface.

6. CONCLUSIONS

In this paper, we have presented an computational house recognition system by integrating hierarchical stereo matching algorithm. The major idea of the algorithm is that it employs the disparity differences or surface discontinuities between the ground surface and object surface in imagery to detect the significant man-made structure, which opens up a new path for automated approach.

The experimental results show that more than 95% of the houses in stereo images are able to be automatically extracted very low uncertainties. Furthermore, as a by products, one can find that many of the 3D lines on the ground are the borders of the roads which provide rich information for our further research on road extraction.

6. REFERENCES

P. J. Burt and E. H. Adelson , "The Laplacian pyramid as a compact image code," IEEE Trans. Communication vol. COM-31 , April 1983 , pp. 532-540.

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

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

A. Grossmann and J. Morlet , "Decomposition of Hardy function into square integrable wavelets of constant shape," SIAM J. Math. Anal. 15 , 1984 , pp. 723-736.

M. Herman, T. Kanade, "Incremental reconstruction of 3D scenes from multiple complex images," AI, vol. 30, 1986.

A. Huertas, R. Nevatia, "Detecting houses in aerial images," CGIP, vol. 41, no. 2, 1988.

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

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

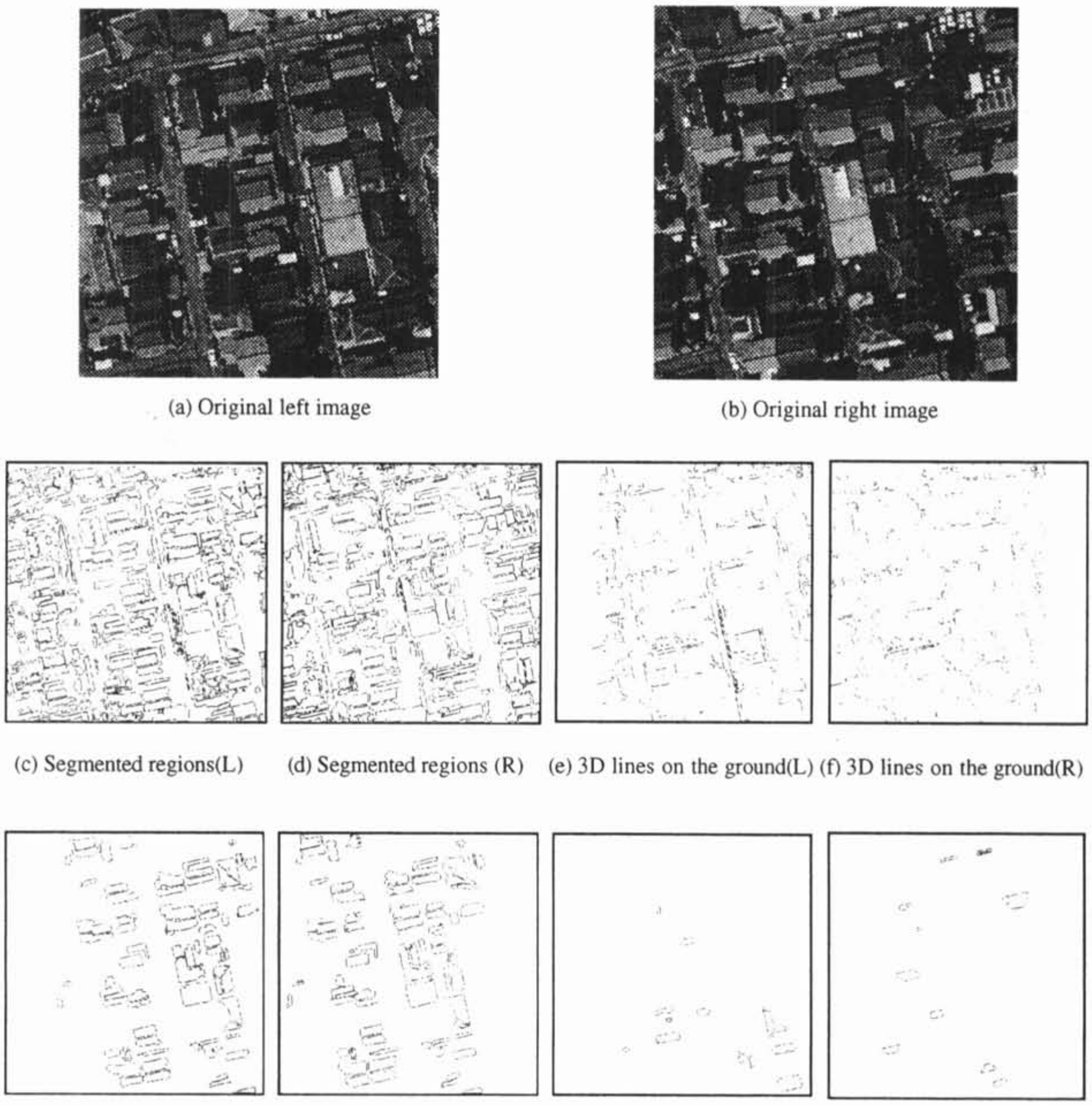
S. B. Marapane and M. M. Trivedi, "Region-based stereo analysis for robotic applications," IEEE T. Sys. Man. and Cyber. vol. 19, no. 6, 1989, pp.1447-1464.

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

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

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

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



(a) Original left image (b) Original right image

(c) Segmented regions(L) (d) Segmented regions (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)

Fig. 2. Experimental results of house extraction with real images.