Site Model Generation Based on 3D Line Segments from Aerial Images

Dong-Min Woo

***Abstract*—**This paper presents 3D site modeling methodology based on 3D line data. 3D lines are evaluated by using line fitting of digital elevation map (DEM) on 2D line coordinates in an image space. Region of interest for site model is generated from the disparity map to segment the interested objects and consequently reduce unnecessary line segments extracted in the low level feature extraction step. Hypothesis selection is carried out by using an undirected graph, in which close cycles represent complete rooftop hypothesis. We test the proposed method with the synthetic images generated from Avenches data set of Ascona aerial images. The experiment result shows that the extracted 3D line segments of the reconstructed buildings reflect real site model and our method can be efficiently used for the task of building detection and reconstruction from aerial images.

***Keywords*—**Site modeling, 3D line digital elevation map, rooftop, feature extraction.

I. Introduction

3D site modeling is one of the most popular tasks in aerial image application. Generation of 3D rooftop model requires 3D features, such as corner, 3D line, surface, etc. Among them, 3D line segment can be regarded as one of the most important features, especially in generating a polygonal form of model.

There are two main problems in any building detection approach based on 3D line. The interested objects must be segmented from the background and the fragmented line segments of the interested object's boundaries should be grouped to human-made structures. These are challenges because the objects of interest could be partly occluded. Also, 3D lines of objects are often fragmented and missed due to improper 2D line extraction.

Mohan and Nevatia [1] proposed an approach for detecting and describing buildings in aerial images using perceptual grouping. They demonstrated the usefulness of the structural relationships called collated features which can be explored by perceptual organization in complex image analysis. All reasonable feature groupings are first detected and the candidates are then selected by a constraint satisfaction network. But this approach involves all extracted line segments in the image. Consequently, it requires a significant computational effort. It also depends on the accurate extraction of line segments.

Some approaches such as Lin [2] and Noronha [3] use hypotheses and verify paradigms based on perceptual grouping to solve the second problems. Hypotheses are generated by a hierarchical perceptual grouping process and verified by the evidence of visible walls and expected shadows. But the system needs to make several decisions in the selection and verification process based on simplicity and intuitive judgments that affects much of the final result. In monocular analysis, Jaynes [4] proposed task driven perceptual organization for extraction of rooftop. Features such as corner and line segments are first extracted and assigned a certainty value. Then features and their grouping are stored in a feature relation graph. Close cycles in the graph represent the grouped polygon hypotheses. The independent set of closed groups that have maximum sum of certainty values of its parts is the final grouping choice. This approach is limited on rectangular buildings and tends to have false hypotheses in complexity images.

In this context we propose a new method for rectilinear building detection using two overlapping aerial images. We use hypothesis generation and selection based on perceptual organization strategy to solve the building detection task. The key idea is that we use the proposed suspected building regions extracted from the disparity map for obtaining the location of interested objects in the image. This building location information helps to remove the unnecessary line segments in the low level feature extraction result and thus reduces computational complexity and false hypotheses in later steps. Additionally, hypothesis selection is carried out by graph searching for close cycles in an undirected graph. Compared with Jaynes's approach, our method detects corners from filtered low level features before constructing the graph, whereas corners are extracted using pre-defined corner masks and each corner can take part in many different close cycles in his approach. So our system can significantly reduce computational complexity and false hypotheses. Moreover, we expand the condition required for a link between two corners to be formed and thus enable our system to detect the rectilinear buildings that Jaynes's approach does not detect.

The remainder of this paper is organized as follow: Section 2 describes the generation of epipolar images, disparity map, DEM and area of interest. In Section 3, we present the principle of 3D line extraction, and Section 4 introduces procedure of rooftop hypothesis. Section 5 presents experimental results on an aerial image data set. At last, conclusion and future work are given in Section 6.

Dong-min Woo, is with the Electronics Engineering Department, Myongji University, Gyeongido, Korea 449-728 (e-mail: dmwoo2000@gmail.com).

II. DEM Generation

The goal of this step is to generate the epipolar images from the two original aerial images. For each point in one of the epipolar images, we only need to search for the points in the same row index of the other image for the corresponding point. In this way, matching can be reduced to a one dimensional search instead of a two-dimensional search. Consequently, the system can significantly reduce computational complexity in solving the correspondence problem in the disparity map step.

The goal of the image matching process is finding a match between the pixels in the first (reference) R and second (warp) W image such that the pixel located at (i, j) in the R image and a pixel located at $(i + I(i, j), j + J(i, j))$ in the W image view the same point in object space, i.e., $W(i + I(i, j), j + J(i, j)) \rightarrow R(i, j)$. The index i (column index) is measured along scan lines and the index j (row index) is measured across scan lines. In this paper, we use resampled epipolar images as the input so that $J(i, j) = 0$ for all i and j, and the relation reduce to $W(i+I(i, j), j) \rightarrow R(i, j)$.

Considering the correspondence problem, there are two popular approaches. The first one is Normalized Cross Correlation (NCC), which is an area-based matching typical metric approach and the second one is non-parametric technique with census transform [5]. We employ the census transform due to its preservation of the edges and simple computational complexity.

To find the accurate disparity map, we employed a multi-resolution scheme, referred to as hierarchical, or pyramid processing. For each resolution scheme, the correspondence problem is solved by first computing a census transformed image and then using Hamming distance correlation on that image. The census transformation maps the local region surrounding a pixel to a bit string with pixels having lesser intensities. For example, in a window surrounding a pixel, if a particular pixel's value is less than the center pixel, the corresponding position in the bit string will be set to 1; otherwise it is set to 0. After that, two census transformed images will be compared using a similarity metric based on the Hamming distance, which is the number of bits that differ in the two correlation window bit string. The Hamming distance [6] is summed over the window, shown as follows:

$$\sum_{(u,v)\in W} Hamming(I'_1(\mathrm{u},\mathrm{v}), I'_2(\mathrm{x}+\mathrm{u}, \mathrm{y}+\mathrm{v})) \quad (1)$$

where I'_1 and I'_2 represent the census transforms of I_1 and I_2, and W is the correlation window.

It is usually difficult to separate interested objects from 2D line segment collection obtained in low level features extraction. The boundary of interested objects, the buildings, can be partly occluded by vegetation, shadows, and other objects. In the rooftop hypothesis process, these fragmented boundaries and the presence of roads, vehicles, etc., can make false hypotheses including unwanted rooftops and rooftops of the wrong shape. This causes not only significant computational effort in processing but also erroneous final results. To solve this problem, the system should be able to detect line segments that are within or near buildings in the image. Here, we use suspected building regions that are extracted from the disparity map. The suspected building regions are areas in which pixel values change in comparison with the surrounding area. The difference in pixel values between the suspected building region and surrounding areas indicates the difference of elevation values. It specifies the existence of higher objects such as buildings, trees, etc., in those regions. In other words, these regions can give us the information of where the buildings are located.

These regions could be extracted by using a simple height threshold technique. Their boundaries are extracted by convolving the disparity map with a Laplacian-of-Gaussian filter and then employing connected component analysis to achieve zero-crossing pixels' coordinates in the convolution output.

III. 3D Line Extraction

Our 3D line extraction scheme begins with the 2D line detection, so that we can locate proper coordinates of 3D line. To detect 2D lines, edge detection is carried out first and then 2D lines are formed from edges. We employed Canny edge detector, since it is optimal accord-ing to the criteria where edge is defined and comes up with thin edges. To obtain 2D line segment, we use Boldt algorithm [7] based on token grouping. The method extracts a basic line element, token, in terms of the properties of line, and constructs 2D line using grouping process. It is efficient in detecting 2D lines of large structure appeared in urban image.

The preliminary requirement for the accurate extraction of 3D line segment is that the elevation data used in 3D line fitting should be highly reliable. To determine the reliability of elevation data, we adopt the concept of self-consistency, which reflects our expectation that reversing the target and reference images will lead to similar result when the image matching algorithm find correct correspondence and come up with reliable elevation. Therefore, the difference of the elevations provides a measure of consistency and reliability, and equation (2) indicates the condition for the reliable elevation data.

$$|Z_{ab}(i,j) - Z_{ba}(i,j)| < THD_{height} \quad (2)$$

where $Z_{ab}(i,j)$ and $Z_{ba}(i,j)$ are two elevation data obtained by reversing the reference and target images, a and b. Also, THD_{height} is the threshold to determine the degree of self-consistency. The elevation data determined as reliable by equation (2) can be used in 3D line fitting. Since we have two elevation for each (i, j), the average value of two elevations are actually participated in line fitting such as

$$(Z_{ab}(i,j) - Z_{ba}(i,j))/2 \quad (3)$$

3D line segments are extracted by line fitting of reliable elevation data, as shown in Fig. 1. To get elevation data, we transform DEM to image space by using inverse projection.

In order to minimize the error of the estimated 3D line as small as possible, in this paper, we adopt a line fitting in a sense of LSE(least squared error). In this paper, we define the fitted 3D line as the intersection of two plane equations.

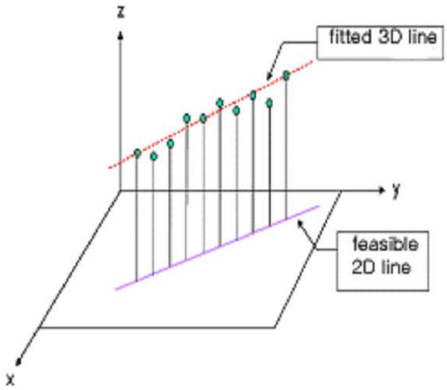

Fig. 1 3D line fitting

IV. 3D Rooftop Model Generation

Corners are calculated as the intersection of two line segments that have an angle from 80° to 100° with one of them having the nearest distance to another one. We define four types of corner. They are labeled as I, II, III, and IV. Each corner has an attribute to indicate whether it is L-junction or T-junction. This attribute is used to decide whether two different corners have a connection or not. For example, if a corner's label is I and type is L-junction, it connects to any type of corner. However, it prefers connecting to a corner which label is II or IV. If that corner is T-junction, it can only connect to a corner which label is II or IV. This rule is used in hypothesis generation to build collated features.

A collated feature is a sequence of perceptually grouped corners and line segments. Here, collated features are constructed from filtered line segments and corners obtained from the filtering and grouping process. This reduces computational effort and false hypotheses.

Hypotheses are formed by the alternation of corners and line segments that form collated features. In a collated feature, two corners have connectivity only if they satisfy the corner relation condition and they are the nearest appropriate corner to each other. Beside, every corner connects to only one corner on each of its line segment directions. Hypothesis generation is performed by constructing the feature graph. Construction of the graph can be seen as placing corners as nodes and edges between nodes if there is the relation between the corresponding corners in the collated features. When a node is inserted into the graph, the system looks into the remaining nodes to determine whether any node has a relation with the inserted node. If some nodes satisfy the connectivity relation rules, those nodes are inserted into the graph and the system creates an edge between them.

The graph is the place to store features and their groupings. Feature as corner is node in the graph and relations between corners are represented with an edge between the corresponding nodes. Closed cycles in the graph represent the rooftop candidates. Hypothesis selection can be seen as a simple graph search problem. The close cycles in the graph are rooftops that we need to detect.

V. Experimental Result

The experimental environment was set up based on Ascona aerial images. Since this area's 3D model is supplied as ground truth data, we can evaluate the quantitative accuracy for the 3D rooftop model generated by the proposed method. Two aerial images as revealed in Fig. 2 are used as a set of stereo images for the experiments.

Fig. 2 Two aerial images used in the experiment.

The first step to carry out the experiment is to reconstruct DEM from overlapped images. Area-based stereo is performed on synthetic images with non-parametric technique with census transform. We employ the census transform due to its preservation of the edges and simple computational complexity.

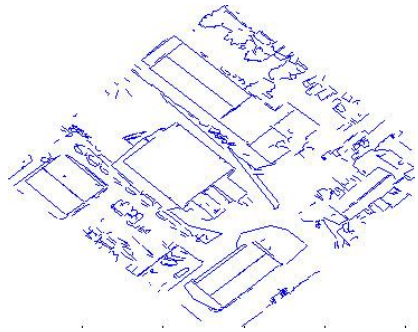

Fig. 3 Example of low level feature extraction result

To find 3D line segments, we begin with Canny edge detector and Boldt line finder to find 2D line segments, as shown in Fig. 3. To get the feasible 2D line segments, the feasibility test is carried out for filtering of all the 2D line segments detected. The threshold for the feasibility test is set to 1m. 3D line segments are obtained by line fitting of elevation data. The elevation data is generated from DEM by using inverse projection into image space. To remove unnecessary line segments, we use the suspected building regions extracted from the disparity map as shown in Fig. 4

Fig. 4 Example of suspected building regions

After the unnecessary line segments are removed, the perceptual filtering and grouping process is employed to obtain line segments which can be part of any U-structure group. The close parallel line segments that are inside their folding space

will be grouped into one representation line. The line segments that are part of a collection of line segments forming the U-structure will be used to generate hypotheses in the next step.

The corners are calculated from the intersection of the line segments that satisfy two conditions: their angle is from 85° to 95° and one of them has the nearest distance to another one. Fig. 5 provides extracted corners from the line segment collection. We can use the obtained corners and line segments from the previous steps to build the collated features. In order to have a link between each other, two corners must satisfy the connecting relation of corner type and the required condition of their distance. Another important rule that helps to define the corner connectivity is that on each line segment of a corner, there is only one corner that has connection with it.

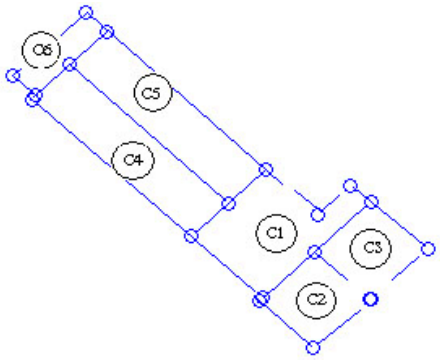

Fig. 5 Detected corners and collated features

The collated features are used to construct a graph by placing a corner as a node and two corners of a line segment as an edge between two nodes if there is a relation between the corresponding corners in the collated features. Closed cycles in the graph represent the possible rooftops. Hypothesis selection becomes the searching of close cycles in the graph.

Fig. 6 presents 3D rooftop detection result. There is a building located near the border of the epipolar image that the system cannot detect correctly due to missing line segments in low level extraction step. The result of the remaining building is very good. From the detected rooftop and the known geometric parameters of image acquisition, we reconstructed 3D building using 3D triangulation.

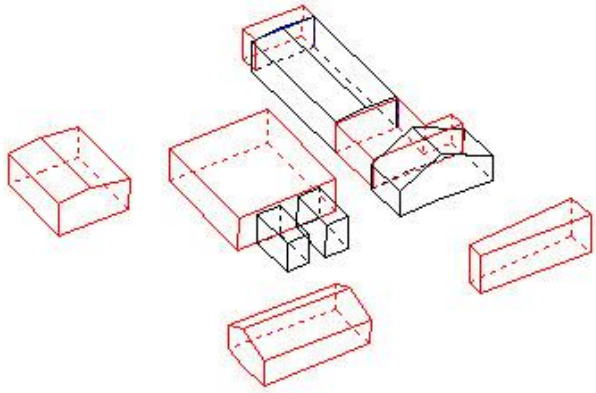

Fig. 6 Example of 3D building reconstruction

VI. CONCLUSIONS

In this paper we suggested 3D rooftop detection method based on 3D line segment by using DEM and ortho-image which can be obtained by conventional area-based stereo method. Our approach to detect and reconstruct buildings using perceptual organization from two aerial images has been suggested. Low level feature extraction is not applied in the original images but from the epipolar images. Using the undirected feature graph, the selection of rooftop hypotheses becomes a simple graph searching for close cycles. The experimental result shows that the proposed method can be very effectively utilized for the rectilinear structures of an urban area.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. NRF-2012R1A1A2004950).

REFERENCES

[1] R. Mohan and R. Nevatia. Using Perceptual Organization to Extract 3D Structure. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 11, 1989, pp. 1121-1139 http://dx.doi.org/10.1109/34.42852

[2] C. Lin and R. Nevatia. Building Detection and Description from a Single Intensity Image. Computer Vision and Image Understanding, Vol. 72, 1998, pp. 101-121 http://dx.doi.org/10.1006/cviu.1998.0724

[3] S. Noronha and R. Nevatia. Detection and Modeling of Buildings from Multiple Aerial Images. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, 2001, pp. 501-518 http://dx.doi.org/10.1109/34.922708

[4] C. Jaynes, F. Stolle and R. Collin. Task Driven Perceptual Organization For Extraction of Rooftop Polygons. IEEE Workshop on Application of Computer Vision, 1994, pp. 151-159 http://dx.doi.org/10.1109/ACV.1994.341303

[5] J. Banks, M. Bennamoun and P. Corke, Non-parametric technique for fast and robust stereo matching, Proceedings of IEEE TENCON - Speech and Image Technologies for Computing and Telecommunication, 1997, pp. 365-398

[6] R. Zabih and J. Woodfill. A non-parametric approach to visual correspondence. Proc. ECCV, 1994, pp. 82-89

[7] M. Boldt, R. Weiss and E. Riseman. Token-based Extraction of Straight Lines. IEEE Trans. Systems Man Cybernetics. Vol. 19, 1989, pp. 1581-1594 http://dx.doi.org/10.1109/21.44073