2006 IEEE Conference on
Systems, Man, and Cybernetics
October 8-11, 2006, Taipei, Taiwan

3D Rough Reconstruction of Buildings from Streetscape by Synergetic Stereo Matching

Hideaki Kawano, Shinichiro Imamura, Hiroshi Maeda, and Norikazu Ikoma

***Abstract*— In this paper, a method to roughly reconstruct buildings in 3D space from a streetscape is proposed. The 3D reconstruction is based on a binocular stereo method called synergetic stereo matching. The proposed method is organized into three processes: extraction of buildings regions in each streetscape image, measurements of 3D distances for feature points in the buildings, and representation of a building by a matchbox model. The effectiveness of the proposed method was verified by experiments using actual streetscape images.**

I. INTRODUCTION

In recent years, products and services using 3D map of streetscape have rapidly increased, e.g. car navigation systems, townscape simulation systems, 3D games and so on. It, however, needs to consume a lot of efforts to make 3D maps, because most of production procedures for 3D map are achieved by hands. Thus, it is an important issue to automate the production process of the 3D map.

In this paper, a method to roughly reconstruct buildings in 3D space from a streetscape is proposed. In this study, measurements of 3D distances are achieved on the basis of binocular stereo. As compared to methods using a laser range finder[1], stereo-based methods are able to measure wider area at a time, and build up the system at a lower cost. Thus, it is considered stereo-based methods are adequate for the objectives. The proposed method is organized into three processes: Firstly, in order to limit the scope of reconstruction, areas of buildings in streetscape image are extracted. To achieve this procedure, an image segmentation method combining two different clustering methods is proposed. One is k-means clustering using color information. Another is dynamic binalization for spatial frequency data[2]. By combining these methods, buildings regions in streetscape are extracted with edges preserved. Secondly, the feature points are extracted from the segmented image, and stereo matching is performed for these points. In the estimation of 3D distance of buildings using by stereo method, the matching problem of the stereo image is contained. In the proposed method, the stereo matching problem is solved by using the pattern matching with synergetics[3], named Synergetic Stereo Matching[4]. The synergetic stereo matching is an effective and flexible method as compared with stereo matching methods using block matching and DP matching[5] used in general. On the other hand, accuracy is affected by the selection of features using for matching. Thus, it is necessary to select the features carefully. To apply natural scenes such a streetscape, the features using for matching is newly established. Finally, the positions of the building planes are estimated utilizing 3D distance data obtained by the stereo matching and 2D map information. After alignment and decision of building height, buildings are represented by matchbox models, i.e. 3D rough reconstruction of buildings can be achieved.

H. Kawano is with Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensui-cho Tobata-ku, Kitakyushu 804-8550, Japan kawano@sys.ecs.kyutech.ac.jp

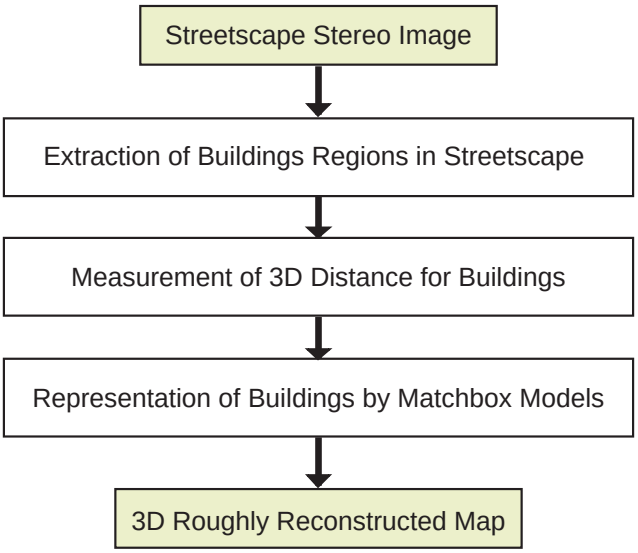


Fig. 1. Procedures of the proposed method.

The effectiveness of the proposed method was verified by experiments using actual streetscape images.

II. RECONSTRUCTION OF BUILDINGS FROM STREETSCAPE

In this section, the proposed method to roughly reconstruct buildings in 3D space from a streetscape is explained. The method is organized into three processes: extraction of buildings regions in each streetscape image, measurements of 3D distances for feature points in the buildings, and representation of a building by a matchbox model. The procedures of the proposed method is summarized as shown in Fig.1. Each procedure is discussed in more detail below.

A. Extraction of Buildings Regions in Streetscape

For the sake of buildings reconstruction, buildings regions in streetscape are extracted. In general, streetscape images contain the sky, road, street trees and so on. These regions other than buildings cause some errors in the following stereo matching process. Thus, a preprocessing to eliminate the regions other than buildings is required. The preprocessing is

achieved by newly proposed segmentation algorithm which integrates clustering with color information and wavelet analysis.

1) k-means clustering with color information: A streetscape image is converted from RGB color space to $L^*a^*b^*$ color space, and subdivided to small regions by k-means clustering using color information. After the clustering, integrations of grainy regions into the nearest region, integrations of small regions by majority decision filtering, and re-labeling process on the basis of cluster position are performed as post-processings.

2) Wavelet analysis: Considering that buildings regions include high-frequency components and that road and the sky include low-frequency components, an wavelet transform is applied to a streetscape image for the analysis of locally spatial frequencies included in the image. In the wavelet analysis, original color images are converted to gray-scale images. In this study, the multi-resolution analysis (MRA)[6] is used as a wavelet transform. In MRA, $d_k^{(j)}$ that meets the Eq.(1) for input waveform $f(x)$ is obtained, where $d_k^{(j)}$ is called as a wavelet coefficient.

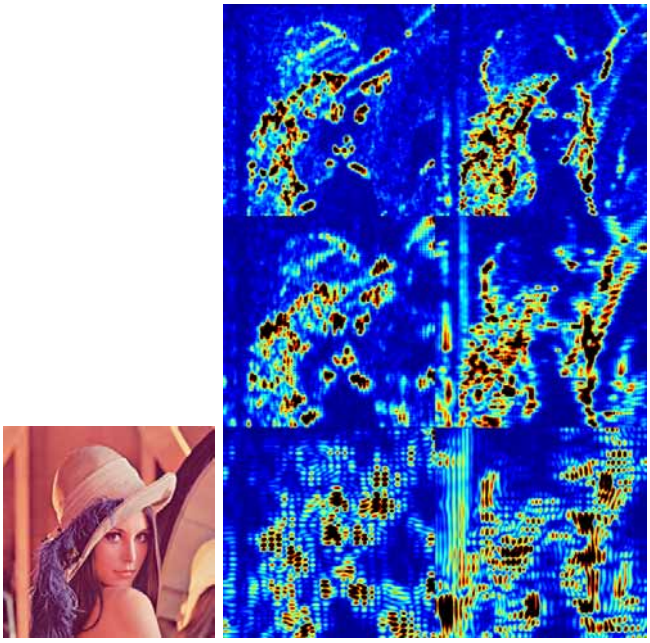
$$f(x) = \sum_j \sum_k d_k^{(j)} \psi(2^j x - k), \tag{1}$$

where j and k are level parameter and shift parameter, respectively. In this procedure, the segmentation method by wavelet analysis is based on the method proposed by Karino et al.[2]. In general, lower-level coefficients represent higher-frequency domain. For the purpose of extraction of buildings regions containing comparatively high frequency, the wavelet coefficients are calculated to the extent of level 5, and local energies of the wavelet coefficients are calculated by Eq.(2).

$$h(t) = \left|\tanh(\frac{t}{\mu})\right|, \tag{2}$$

where μ is a mean of absolute wavelet coefficients. As an example of locally spatial frequencies analysis, visual representations of local energies after 3-level MRA are shown in Fig.2. In the proposed method, the local energy at each level is summed up. The summed local energy image is binarized by Otsu's method[7] in order to separate low-frequency domain and high-frequency domain.

3) Integration of the results by two different segmentation method: Clusters obtained in the k-means clustering stage have well-preserved boundary between objects, however those tend to be too small. The separation result in the wavelet analysis can capture the feature of buildings regions, however the boundary between objects tends to be ambiguous. By integrating the results obtained by two different segmentation methods, acquisition of buildings regions with well-preserved boundary can be expected. In this integration procedure, each cluster obtained in the k-means clustering stage is labeled by dominant frequency-domain, i.e. high or low. Clusters labeled by high-frequency domain are interpreted as buildings regions. On the other hand, clusters labeled by low-frequency-domain are assumed as the regions other than buildings, i.e. the sky or road and so on.



(a) Original image

(b) MRA (3 levels) of original image. Left: vertical direction. Right: horizontal direction.

Fig. 2. Result of locally spatial frequencies analysis.

B. Measurement of 3D Distance for Buildings

The measurement of 3D distance used by the stereo method causes the matching problem of the stereo image. As the method of matching, the block matching is known. In the proposal method, the Synergetic Stereo Matching (SSM) is used as a solution of the matching. SSM is that applies the pattern recognition by the synergetics. In the proposal method, SSM is used as a solution of the matching. SSM is that applies the pattern recognition by the synergetics.

1) Pattern Recognition Based on Synergetics: Synergetics explains the system autonomously changes to another ordered state due to external controls or fluctuation forces, thus system called the complex system. The system explained synergetics receives an external environmental change, it prepares stable and unstable modes. Stable modes are dominated by unstable modes, and then vanished. This process is called the slaving principle. On the other hand, unstable modes competes among modes, and a specific unstable mode wins the growing competition, then forms a new ordered state in the system. The Equations representing this competition are called order parameter equations.

A certain autonomous system equation (3),

$$\dot{\boldsymbol{q}} = N(\boldsymbol{q}(\boldsymbol{x}, t), \alpha), \tag{3}$$

while $\boldsymbol{q}(\boldsymbol{x}, t)$ is a state vector (after this, $\boldsymbol{q}(\boldsymbol{x}, t)$ is written as $\boldsymbol{q}$ for simplicity) and α is an external control parameter. It is considered that the fluctuation force is none.

In the case that $\boldsymbol{q_0}$ is a stable solution at equation (3),

equation (3) is rewritten to (4)

$$\dot{q} = \boldsymbol{N}(\boldsymbol{q_0}) + \boldsymbol{L}\boldsymbol{w} + \hat{\boldsymbol{N}}(\boldsymbol{w}) \tag{4}$$

where $\boldsymbol{w}$ represents a small change and $\boldsymbol{x}$ is position vector. $\boldsymbol{N}$ is Expanded in a power series around $\boldsymbol{q_0}$. The third term of the right-hand side contains the second or higher powers of $\boldsymbol{w}$. Disregarding terms of more than second order and $\boldsymbol{N}(\boldsymbol{q_0})$=0, (4) is transformed into

$$\dot{w} = \boldsymbol{L}\boldsymbol{w}, \boldsymbol{L} = (L_{ij}) = (\partial N_i/\partial q_j|\boldsymbol{q} = \boldsymbol{q_0}) \tag{5}$$

The solutions of (5) are written in the general form

$$\boldsymbol{w}(\boldsymbol{x}, t) = e^{\lambda t}\boldsymbol{v}(\boldsymbol{x}) \tag{6}$$

where λ and $\boldsymbol{\nu}$ are eigenvalue and its eigenvector. In non-degenerate eigenvalues, assume state vector $\boldsymbol{q}$ is presented as

$$\boldsymbol{q} = \boldsymbol{q_0} + \sum_j \xi_j(t)\boldsymbol{\nu_j}(\boldsymbol{x}), \xi_j(t) = \boldsymbol{A}_j e^{\lambda_j t}. \tag{7}$$

λ_t and $\boldsymbol{\nu}_t$ are *t*-th eigenvalue and its eigenvector. Setting of adjoint vector $\boldsymbol{v_k^+}$ is introduced.

$$\langle \boldsymbol{v_k^+}\boldsymbol{v_i}\rangle = \delta_{ki}, \delta_{ki} = \begin{cases} 1 & for\ k = i, \\ 0 & otherwise. \end{cases} \tag{8}$$

Inserting (7) into (4) and multiplying (4) by $\boldsymbol{v_k^+}$,

$$\boldsymbol{v_k^+}\sum_j \dot{\xi}_j(t)\boldsymbol{v_j} = \boldsymbol{v_k^+}\sum_j \xi_j(t)\boldsymbol{L}\boldsymbol{v_j} + \boldsymbol{v_k^+}\hat{\boldsymbol{N}}(\sum_j \xi_j(t)\boldsymbol{v_j}). \tag{9}$$

From (8), the following is obtained

$$\dot{\xi_k} = \lambda_k\xi_k + \tilde{N}_k(\xi_j). \tag{10}$$

Equation (10) is divided into two sets of unstable modes and stable modes, and applying the slaving principle, stable mode are represented by unstable modes.

$$\dot{\xi_k} = \lambda_k\xi_k - \boldsymbol{B}\sum_{k'\neq k}\xi_{k'}^2\xi_k - \boldsymbol{C}(\sum_{k'}\xi_{k'}^2)\xi_k, \quad k = 1, ..., N \tag{11}$$

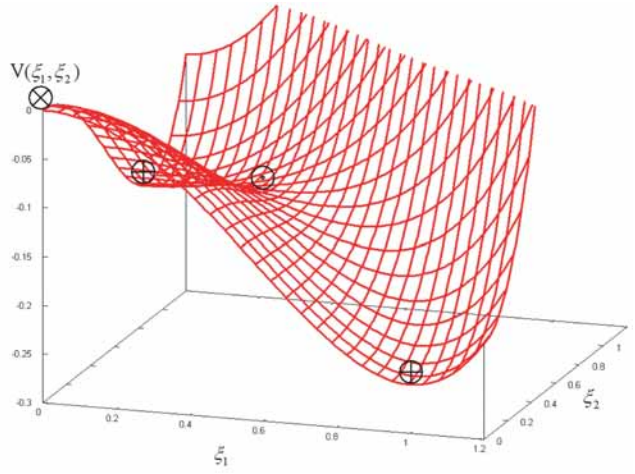
where B and C are positive constants, and ξ_k and λ_k are called the order and attention parameters. Equation (11) is called the order parameter equations of synergetics. In the application of synergetics to pattern recognition, $\boldsymbol{q}$ and $\boldsymbol{v_k}$ correspond to an input pattern vector and the *k*-th prototype pattern vector. ξ_k represents a similarity between $\boldsymbol{q}$ and $\boldsymbol{v_k}$. λ_k plays a role in controlling the growing of ξ_k. The initial value of $\xi_k(0)$ is obtained by inner product of $\boldsymbol{v_k^+}$ and $\boldsymbol{q}$.

$$\langle \boldsymbol{v_k^+}\boldsymbol{q}\rangle = \xi_k(0). \tag{12}$$

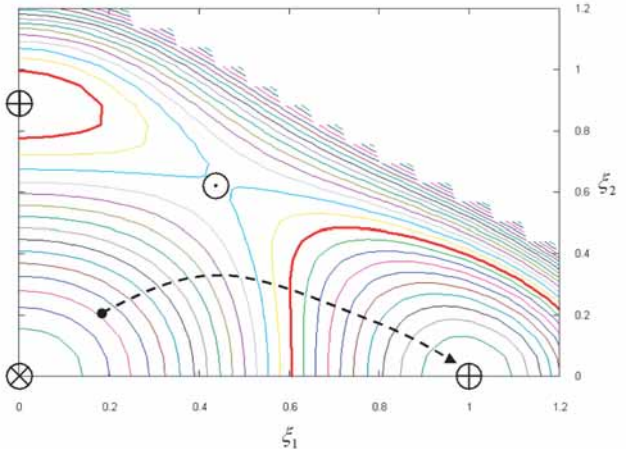
In competition of equation (11), a surviving ξ converges to a positive value and the others converge to 0. Therefore, the input pattern is recognized as a prototype pattern surviving its order parameter.

In this pattern matching, $B = 1, C = 1$. And if N=2, the potential function $\boldsymbol{V}$ is introduced as

$$V = -\frac{1}{2}(\lambda_1\xi_1^2 + \lambda_2\xi_2^2) + \frac{1}{2}\xi_1^2\xi_2^2 + \frac{1}{4}(\xi_1^2\xi_2^2)^2. \tag{13}$$



(a)



(b)

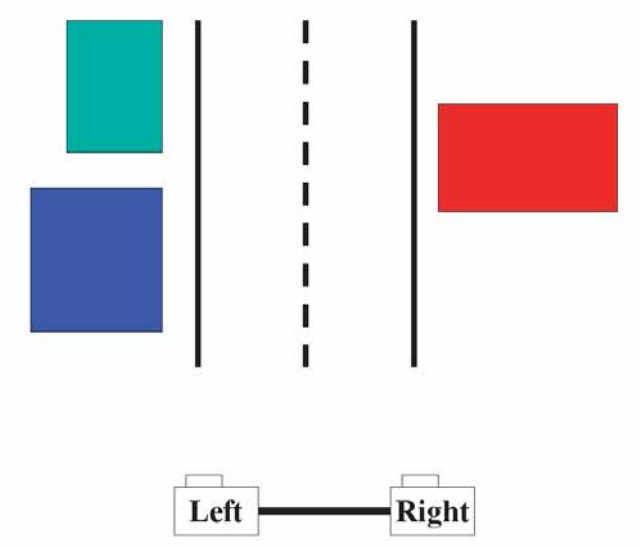
Fig. 3. Potential function $\lambda_1 = 1.0, \lambda_2 = 0.8$. Saddle point, stable focuses, unstable forcuses are shown as symbols $\odot$,$\oplus$,$\otimes$.

using (11). The contour map and landscape of potential (13) are shown in Fig.3, where $\lambda_1 = 1.0$ and $\lambda_2 = 0.8$. In this contour map, a certain stable focus correspond to a prototype pattern.

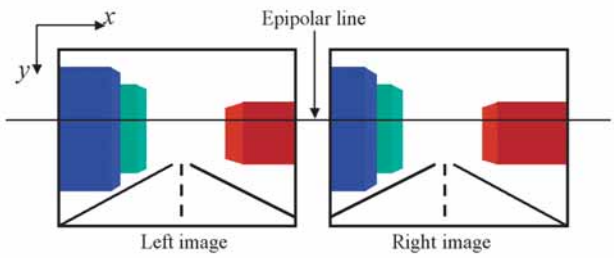
2) Synergetics Stereo Matching Algorithm: In a stereo matching problem, an input pattern vector on an feature point in either image is assigned to q, and the *k*-th prototype pattern vector on feature points in the other image is assigned to v k. The initial *k*-th order parameter psy k(0) is obtained by (10).

In this research, the stereo vision system is considered the following:

1) The geometry position of stereo camera is parallel as shown in Fig.4, so epipolar lines are parallel to the X-axis in a camera coordinate system.
2) The object of streetscape stereo vision is the buildings.



(a) camera geometry



(b) stereo image

Fig. 4. Stereo matching for streetscape.

3) Feature points are edges of texture of buildings and obtained using by a smooth filter , a Laplacian operator, and the thinning operator. These filter size is 3x3 pixels window.

Stereo images generally include two typical problems: occlusion and reversal position. Stereo Images with occlusion have feature points that do not match, and with reversal position may not be able to obtain correct correspondence. Streetscape stereo images are contained many occlusion, but few reversal position. This is attributed to the minimal parallax of stereo using by faraway objects.

3) Construction of Parameters: In the natural image like the streetscape, a definite element is few. In this research, the parameters are mainly set up by color information. For pattern vectors, Lab color information is used on the vicinity large area of the feature point, and brightness information on the vicinity small area of the feature point uses for attention parameters.

Pattern vector $\boldsymbol{v_k}$ is constructed with the following elements (5):

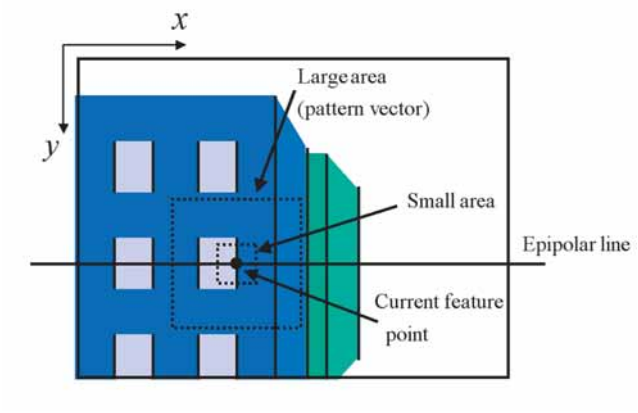


Fig. 5. The information elements. Large area constructs pattern vectors and small area constructs attention parameters.

Lab color values of pixels of large area window (71x71).

The following elements are adopted as the determination of an attention parameter:

The vector has pixel elements of Lab color values of small area window (9x9).

The vector vicinity of the *k*-th feature point of input pattern sets $\boldsymbol{X_k}$ and the vector vicinity of the *j*-th feature point of prototype pattern sets $\boldsymbol{Y_j}$. Thus, λ_k is vector correlation of $\boldsymbol{X_k}$ and $\boldsymbol{Y_j}$.

Let the number of edges on an epipolar line in the image at left and the image at right be N and M. Let the *i*-th edge of the image at left and the *k*-th edge of the image at right be input pattern vector $\boldsymbol{q_i}$ and prototype pattern vector $\boldsymbol{v_k}$. The algorithm of stereo matching is given as the following steps:

1) Calculation of an initial value $\xi_k(0)$ by equation (12).
2) Construction of λ_k.
3) Calculation of order parameter equation (11).
4) Determination of a matching edge to $\boldsymbol{q_i}$.
5) Repeat steps from 1 to 4 for $\boldsymbol{q_i}$ $(i = 1, ..., N)$.
6) Replace the image at right with input ad the image at left with progotype (cross-reference), and repeat steps from 1 to 5.
7) Define edges that are the same result for cross-reference as correspondence points.

C. Representation of Buildings by Matchbox Models

3D distance data is able to be calculated by triangulation from stereo matching. However it is difficult to reconstruct shapes of buildings out of only 3D information for the following reasons:

- In stereo method, 3D information is calculated based on stereo disparity, so depth-distance is obtained discretely by quantization error.
- 3D information from a stereo image has a lot of outliers by mismatching.
- When shape of the building is not included enough in the stereo image as only the side in the building can be

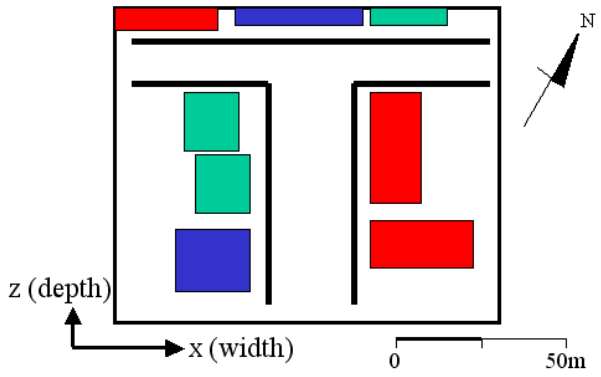


Fig. 6. 2D map information

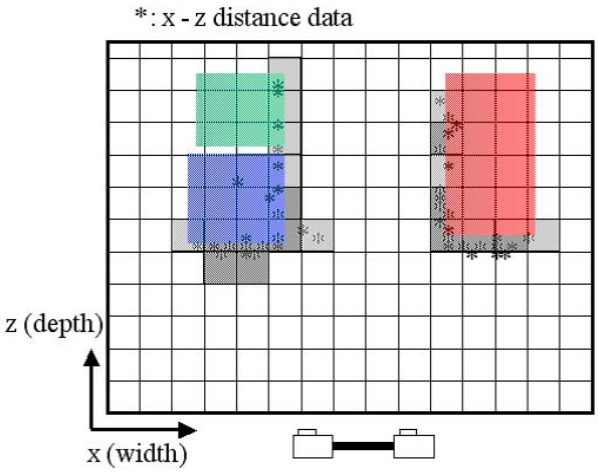


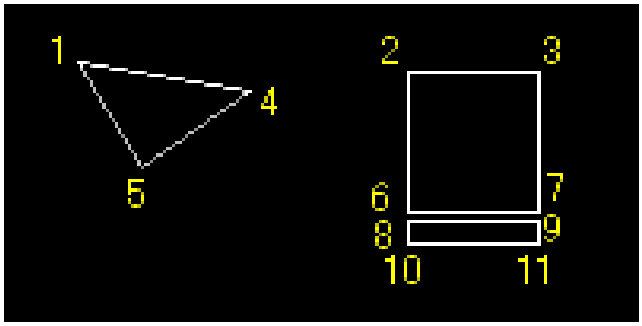
Fig. 7. 2D frequency distribution and 3D information

observed, it is necessary to complement a hidden plane for 3D reconstruction of buildings.

In this paper, a simple 3D map is composed by combining 3D information and 2D urban map. Now, it is assumed that the 2D map as shown in Fig.6 has the following characteristics: (1)Position and shape of the building except for height is known. (2)Map scale is also known. (3)Azimuth of the map is given.

At first, alignment between 3D information and 2D map is performed. 2D histogram of 3D data projected onto horizontal plane are considered. The alignment is conducted in such a way that the 2D histogram laps over the edges of 2D map as shown in Fig.7.

In such an alignment procedure, the photographing position on 2D map is known. For example, GPS can be used to know the photographing position. Moreover, an electronic compass can be used to know the direction of it. In this study, it is assumed to measure an exact position of photographing position. The photographing direction is searched by rotating map information between the position in which 2D frequency distribution of obtained 3D information overlaps with 2D plane information.



(a)

	1	2	3	4	5	6	7	8	9	10	11
1	−1	0	0	1	−1	0	0	0	0	0	0
2	0	−1	1	0	0	1	0	0	0	0	0
3	0	1	−1	0	0	0	1	0	0	0	0
4	1	0	0	−1	1	0	0	0	0	0	0
5	1	0	0	1	−1	0	0	0	0	0	0
6	0	1	0	0	0	−1	1	0	0	0	0
7	0	0	1	0	0	1	−1	0	0	0	0
8	0	0	0	0	0	0	0	−1	1	1	0
9	0	0	0	0	0	0	0	1	−1	0	1
10	0	0	0	0	0	0	0	1	0	−1	1
11	0	0	0	0	0	0	0	0	1	1	−1

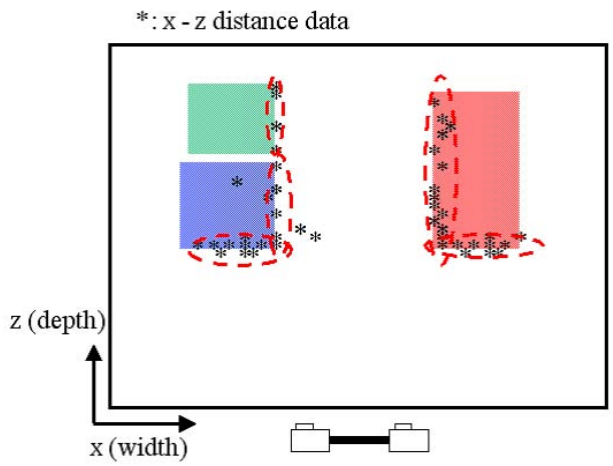
(b)

Fig. 8. Matrix of planes in test image

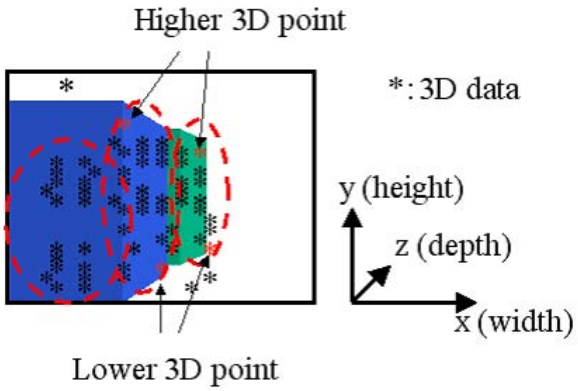
Next, plane combinations that belong to the building are found from 2D map. The extraction of the building plane from 2D plane extracts the corner in the building by the pattern recognition, and composes plane information from the combination of corners. Combinations for all the point of contact parts are put into matrix, and can be found plane combinations that composes one object by matrix operation. The top is extracted from the plane of the object, and matrix representation of the composition of the plane is shown in Fig.8.

Finally, a plane size that composes each building is determined. In this study, when each plane height that composes the building is different, highest height and lowest height of plane are selected. Therefore, the height can be decided from the image where only the side in the building is included.

The position on the x-z plane of the building plane can be expressed by positional identification with 3D information using the distance from the camera. The height of a building is determined from 3D information that exists in the neighborhood of all plane information that belongs to the building. 3D information on neighborhood is considered to be 3D information that composes the building plane, and the maximum value and the minimum value of the height of the information are assumed to be height of the building. Fig.9 shows the selection of 3D information that decides height of plane of the building. A building is schematically



(a)



(b)

Fig. 9. Height of buildings

(a) Example of streetscape image.

(b) Result of k-means clustering and post-processing.

Fig. 10. Clustering and integration

reconstructed using the shape of the matchbox by the above-mentioned procedure. Then, 3D simple map is made from building information expressed by the matchbox model.

III. EXPERIMENTAL RESULTS

In this section, experimental results in each procedure are presented in order.

A. Extraction of Buildings Regions

The size of stereo image used in this experiment is originally 3008 x 2000 pixels, but reduced size (752 x 500 pixels) of image is used from a viewpoint of computational costs in this stage. In the next stereo matching stage, the reduced images are expanded to the original size.

Experimental settings in k-means clustering are as follows: the number of cluster centers k is 20, the dimension of input vector is 3 (L^*, a^*, b^*). Experimental settings in integration process as post-processing of k-means clustering are as follows: size of grainy cluster is set to 1 pixel, size of filtering window in majority filtering is 3 x 3, threshold of size deciding a region as small region is set to 100 pixels experimentally. As a result of applying the k-means clustering to a streetscape image shown in Fig.10(a), cluster boundaries as shown in Fig.10(b) are obtained. In Fig.10(b), red line represents cluster boundary.

Figure 11 shows the result of wavelet analysis. In Fig.11, black regions and white regions represent low-frequency domain and high-frequency domain, respectively.

After the integration of results obtained by two different segmentation method processed so far, an image as shown in Fig.12 was acquired, i.e. buildings regions could be extracted from the given streetscape image.

B. Measurement of 3D Distance for Buildings

Figure 13 shows a stereo image after the extraction of buildings regions. As a result of synergetic stereo matching using the stereo image, 3D distance are measured as shown in Fig.14.

Fig. 11. High-frequency regions (white) and Low-frequency regions (black).

Fig. 12. Buildings regions extracted from the given streetscape image

C. Representation of Buildings by Matchbox Models

In this study, the 2D map for determination of the side planes of buildings is made manually. And it is assumed that the position taking the stereo image is known. The manually 2D map is shown in Fig.15.

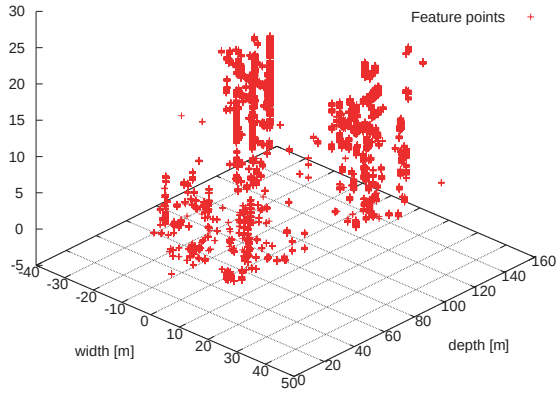
At first, positioning between 2D information and 3D distance data is performed by scaling and rotation of 2D map. The scales of 2D map and 2D frequency distribution of 3D data are set 0.5 [m/pixel]. Next, planes belonging to a certain building are obtained by matrix operation. Finally, heights of the buildings are determined by neighborhood 3D distance data, and shapes of the buildings are represented by matchbox models. Consequently, 3D map is constructed by these models. This 3D map is shown in Fig.16. As can be expected from Fig.16(b), relative height among three buildings is preserved well.

IV. CONCLUSIONS

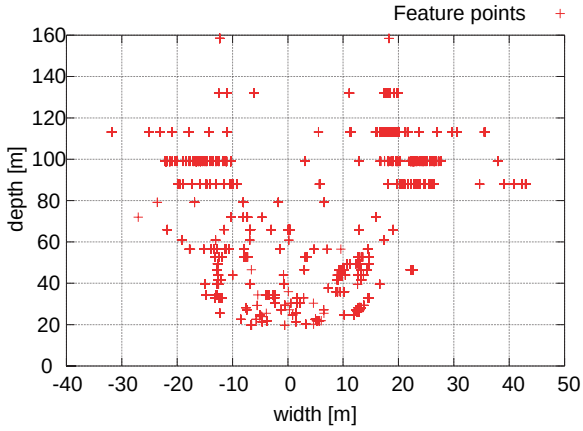
In these experimental results, an availability of automatically making 3D map is indicated. However, the accuracy of the 3D map is not examined enough. Examination accuracy of the 3D map and speeding up of these procedures are subjects for a future study.

(a) Left image (b) Right image

Fig. 13. Stereo image



(a) Birds-eye view of measured 3D distance data.



(b) Projected data onto x-z plane.

Fig. 14. 3D reconstruction

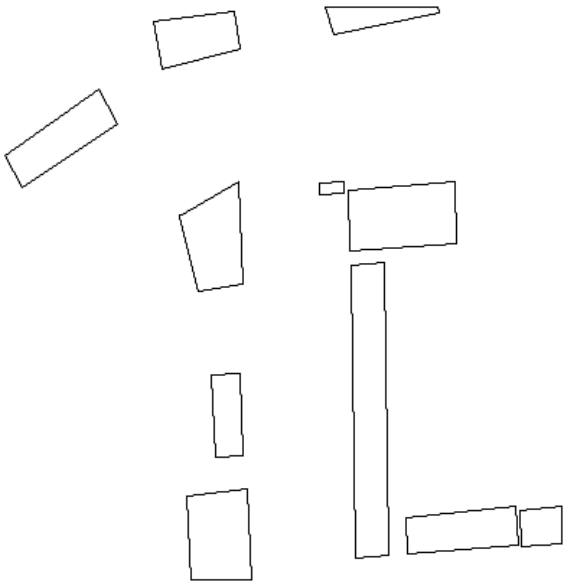


Fig. 15. 2D map information

And in this study, by using spatial frequency and $L^*a^*b^*$ color, building regions are able to separated. However, the objects in streetscape image (e.g. street trees, side walk, signaler, and so on) are not recognized. Recognition of these objects are necessary to making the 3D map more easily viewable.

The precision of the method depends on scenes or conditions. For instance, the intensity change of the sky is a small, there are a lot of cars and patterns on the road, and so on. Moreover, the extraction of the building area might not be enough accuracy according to the low-frequency component in the building that has wall without pattern and so on.

V. ACKNOWLEDGMENTS

This work was supported by the Japan Society for the Promotion of Science under the Grant-in-Aid for Scientific Research (C) (No. 16500133).

REFERENCES

[1] T. Asai, M. Kanbara, and N. Yokoya, "3D modeling of outdoor scenes by integration stop-and-go and continuous scanning of rangefinder," *Meeting on Image Recognition and Understanding 2005 (MIRU2005)*, pp.1630-1631, July 2005. (in japanese)

[2] Y. Karino, S. Omachi, and H. Aso, "Unsupervised Segmentation of Texture Images Using Feature Selection," *IEICE Trans. on Information and Systems D-II*, Vol.J86-D-II, No.7, pp.988-995, 2003.

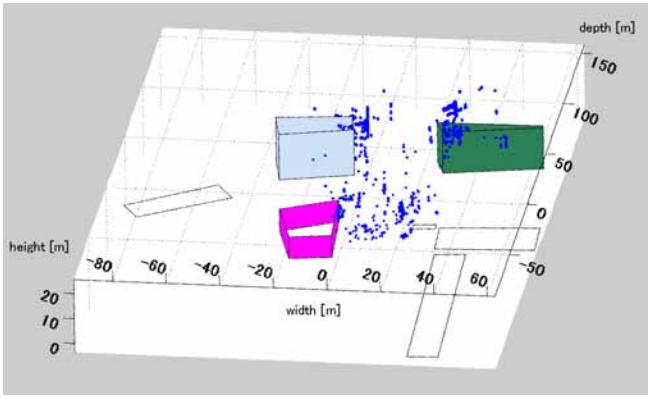
[3] H. Haken, "Synergetic Computers and Cognition," *Springer-Verlag*, 1991.

[4] T. Irie, H. Maeda, and N. Ikoma, "Synergetic Stereo Matching Algorithm for Occlusion and Reversal Position," *J. of Advanced Computational Intelligence and Intelligent Informatics*, Vol.7, No.2, pp.178-188, 2003.

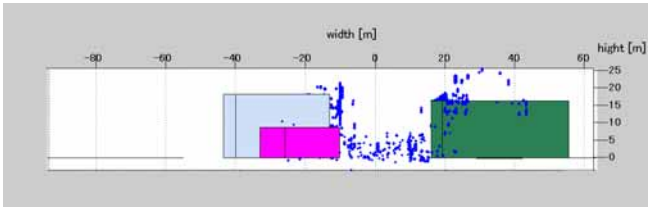
[5] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter- Scanline Search Using Dynamic Programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.7, No.2, pp.139-154, 1985.

[6] S. Mallat, "A wavelet tour of signal processing," *Academic Press*, 1998.

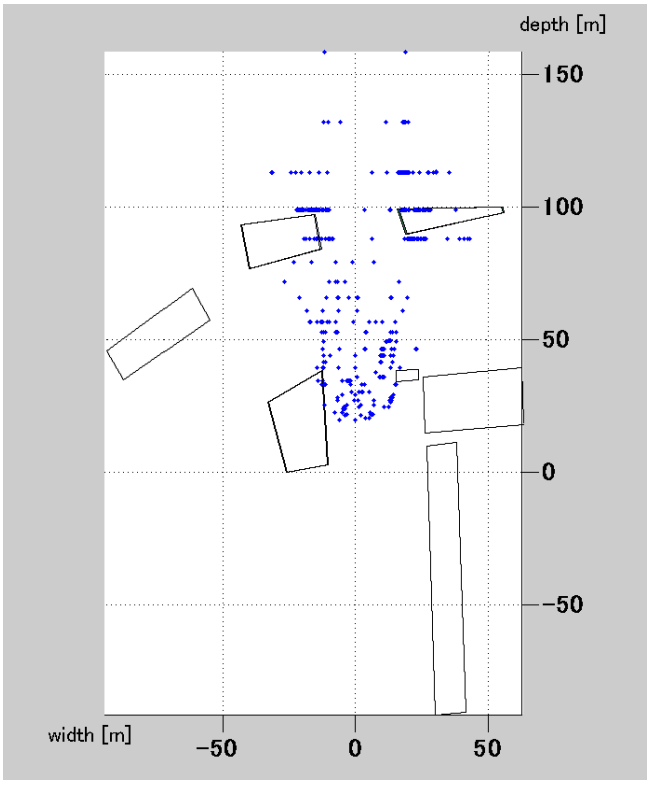
[7] N. Otsu, "An Automatic Threshold Selection Method Based on Discriminant and Least Squares Criteria," *IECE Trans.*, Vol.J63-D, No.4, pp.349-356, 1980.



(a) Birds-eye view of reconstructed 3D map



(b) Projected 3D map onto x-y plane



(c) Projected 3D map onto x-z plane

Fig. 16. Reconstruction of buildings and roughly 3D map