CONSTRUCTION OF URBAN SCENE MODEL BY IMAGE SEQUENCE ANALYSIS USING DIGITAL MAP

Masafumi UEHARA, Takeshi SHIOSAKI, Heitou ZEN
Tokyo University of Mercantile Marine, Japan
{muehara, tshio, zen}@ipc.tosho-u.ac.jp

Working Group V/5

KEY WORDS: Motion Stereo, Epipolar Plane Image Analysis, Digital Map, DP Matching

ABSTRACT

Our group has researched about a method of the construction of urban model by Motion Stereo with motion images took from a camera on a vehicle. But in the real city environment, it was difficult to construct the model accurately, because a range of measured 3D data of buildings cluster and the slight motion of images due to the pitch, yaw and roll from the vehicle motion cannot be removed completely. This paper proposes a method of the construct of urban scene model with a large scale digital map and 3D measured data. A large scale map is expressed the topographical shape of buildings accurate comparatively. Then, we established the relation between 3D measured data by Motion Stereo and a digital map, and constructed the accurate and photorealistic city model adding height and texture data to each building.

1 INTRODUCTION

3D model of urban scene is made widely use in various fields, such as city engineering, traffic engineering and amusement. Under present conditions, these models are made manually with CG modeling tools. Hence, huge time and cost are wasted, and utilization of these models is restricted. Therefore, in order to construct model efficiently and automatically, many approaches have been suggested. These approaches are roughly divided into two methods. One is expression using stored 360 real images from observation point without 3D data[1][2]. This method is called IBR (Image Based Rendering), it is possible to realize photorealistic urban scene. But on the other hand, it is necessary to use a large quantity of data to utilize many images. The other is to make model using measured 3D date of buildings with the stereo method or a laser measurement[3]. This method enables to move a viewpoint freely and construct model with a small quantity of data. We have constructed urban scene model based on 3D data, too. Our group proposed a method of modeling using measured 3D data of buildings and texture data of target buildings with motion images took from a camera on a vehicle[4]. 3D data were obtained by applying Extended EPI (Epipolar Plane Image) analysis that had been improved over EPI analysis proposed by Bolles et al[5]. And texture data were obtained from the slit spatiotemporal plane image from image sequence. Figure 1 shows the result of modeling with the proposed method for our group. This method cannot separate every building because the construction result was only expressed the measured 3D data. And in the real city environment, it is difficult to measure 3D data of buildings because of the existence of objects except buildings, such as a thick growth of trees, guardrails and poles. For reasons mentioned above, it was difficult to construct the realistic model.

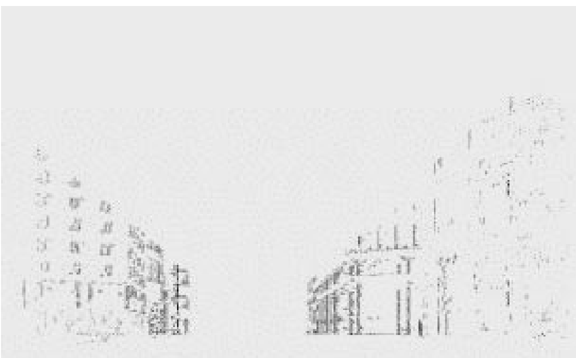


Figure 1: Previous result of modeling

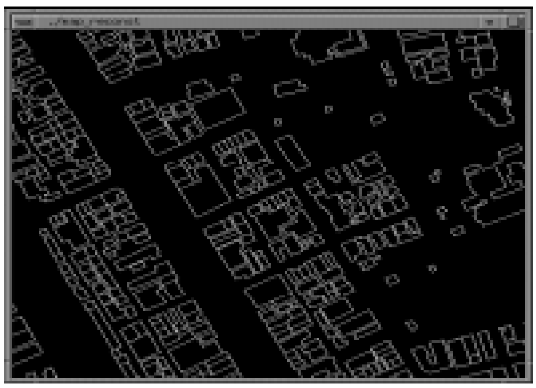


Figure 2: Example of digital map

We resolve the above problem by utilizing a large scale digital map. In the last few years, a digital map has been spreading remarkably. The digital map we used is expressed the topographical shape of buildings accurate comparatively (see Figure 2), so we know where buildings are separated. And, considered utilization of wide area city models, it is necessary to prepare data based on a digital map for modeling. Therefore, in this paper, we propose a method of the construction of urban scene model using a digital map adding 3D data of buildings by motion image analysis.
First, we mention the method how to obtain 3D data of buildings in Chapter 2, secondly Chapter 3 describes the matching method with measured 3D data and a digital map data. Experimental results are shown in Chapter 4, and Chapter 6 describes conclusion of this research.

2 IMAGE SEQUENCE ANALYSIS

2.1 Obtaining depth data by EPI analysis

3D data of buildings are obtained by EPI analysis. The traces of feature points appear on the Epipolar plane image when the camera motion is uniform linear motion (see Figure 3). These traces are straight lines if the optical axis of a camera is perpendicular to the camera path. The distance from the camera path to the target buildings is able to be measured the slants of these lines.
It is necessary to prepare many cameras for wide area observation in an EPI analysis. In an EEPI analysis, all flows of stationary objects appear from FOE (Focus of Expansion) are picked up by the camera whose the optical axis is parallel to the camera path (see Figure 4). Thus a wide region can be observed at once by a small number of cameras. The obtained EEPI (see Figure 5) can be translated to EPI with simple geometric formulas.

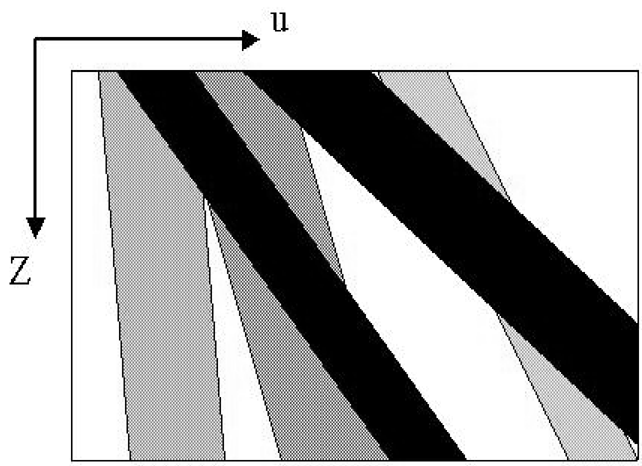


Figure 3: Trace of feature point on EPI

2.2 Construct of data

The texture data for expression of urban scene are obtained from the image sequence. The slit spatiotemporal plane image utilized as texture data integrate vertical lines of both sides of each frame of image sequence toward the temporal axis (see Figure 6).
Let this slit image corresponds to distance data obtained the EEPI analysis above, we call this depth map of urban scene.

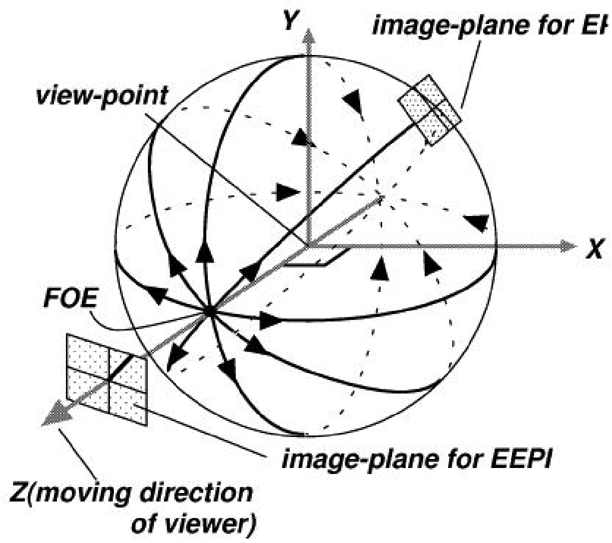


Figure 4: Flow of stationary object under straight moving

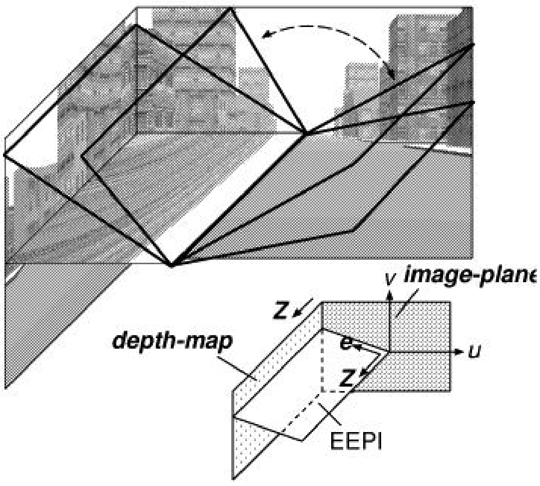


Figure 5: EEPI

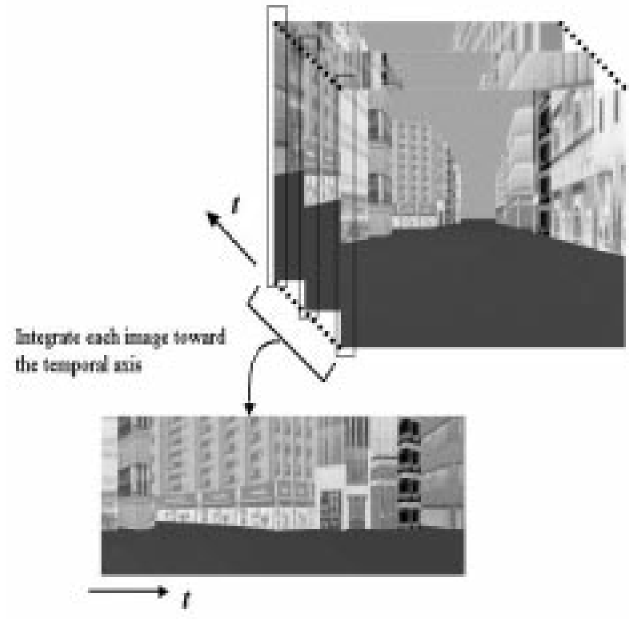


Figure 6: Slit image

3 MATCHING MEASURED 3D DATA WITH DOGITAL MAP

As mentioned in Chapter 1, it is difficult to measure 3D data of buildings because of the existence of objects, such as a tick growth of trees and guardrails in the real city environment. Moreover, the results of EPI analysis show that on the edge parts of texture alteration 3D data of buildings are densely measured meanwhile on the parts of less or no alteration 3D data are roughly measured.

For the reasons given above, it has been difficult to construct the city model accurately only from measured data by EPI analysis. Digital map covers the 3D measured points. In the following section, we explain the matching method utilizing boundary information between buildings in order to match 3D data with digital map.

3.1 Detecting boundary of buildings from depth map

As the white points show in Figure 6, the 3D measured points appear on the parts of the vertical edge on the depth map. These parts are equivalent to steep texture alteration, such as boundaries between buildings, windows and doors.

Therefore, when we make the histogram of measured points in the direction of the camera path (Z-axis), the peak of this histogram can be likely judged the prospective boundary of buildings (see Figure 7). At this time, we utilize the histogram of the measured points in the direction of the X and Z-axis with the measured points of buildings face the street.

Figure 6: Measured points by EPI analysis

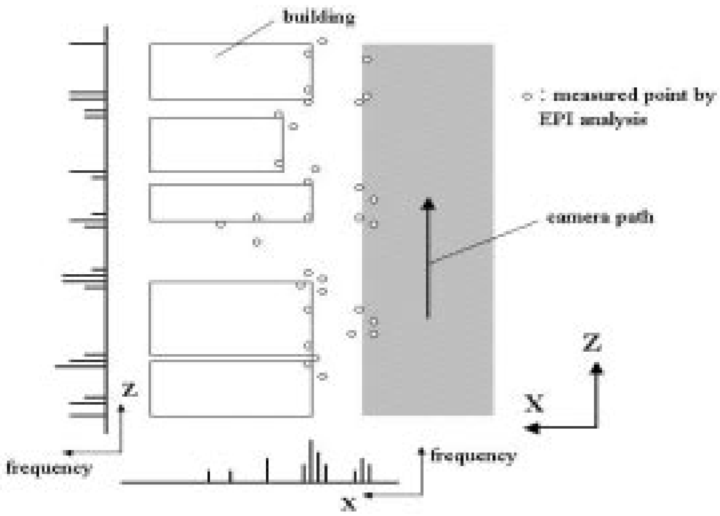


Figure 7: Histogram of measured points

On the other hand, we make the pattern of the building boundary from a digital map like Figure 8. Trace the track of the vehicle by GPS over the digital map. A perpendicular line is drawn from the building's edge faces the street toward this track. The record of intersection of this line and trace is utilized to make the pattern of buildings.
The next step of process is to match the histogram of measured points by EPI analysis with the pattern of the buildings made from a map, then judge the location of the building boundary in the peak of the histogram.

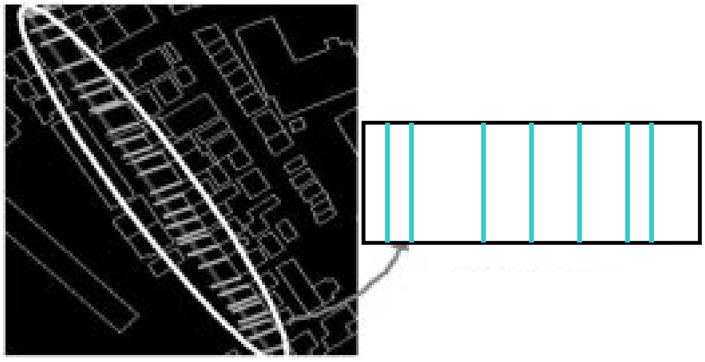


Figure 8: Building boundary pattern by digital map

3.2 DP matching

In this research, we apply DP matching method to match digital map data with 3D measured data by EPI analysis. The pattern of building boundary from digital map and the prospective pattern of building boundary from the histogram of the 3D measured points by EPI analysis are utilized as the feature vectors. Figure 10 shows an example of the path of DP matching. In this figure, the vertical line is the prospective pattern of building boundary from the histogram, and a horizontal line is the pattern from the histogram of the 3D measured data by EPI analysis. If these two patterns correspond completely, the path of DP matching in Figure 10 becomes straight. But windows or doors of buildings except building boundary are detected with 3D point histogram, the path of DP matching takes a zigzag course.
Matching urban scene information with a map data utilized DP matching method has reported in [6]. In that research, the boundary pattern made from obtained panorama image of urban scene has matched with the building boundary pattern made from a map. This research utilizes the depth-data of buildings by EPI analysis.

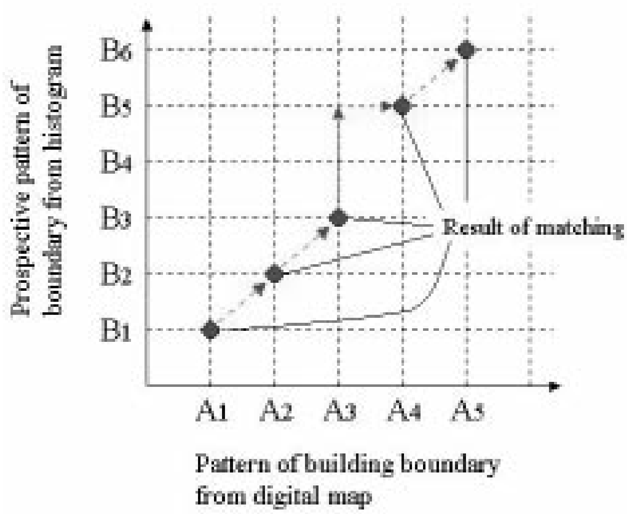


Figure 10: Example of path of DP matching

4 EXPERIMENTAL RESULTS

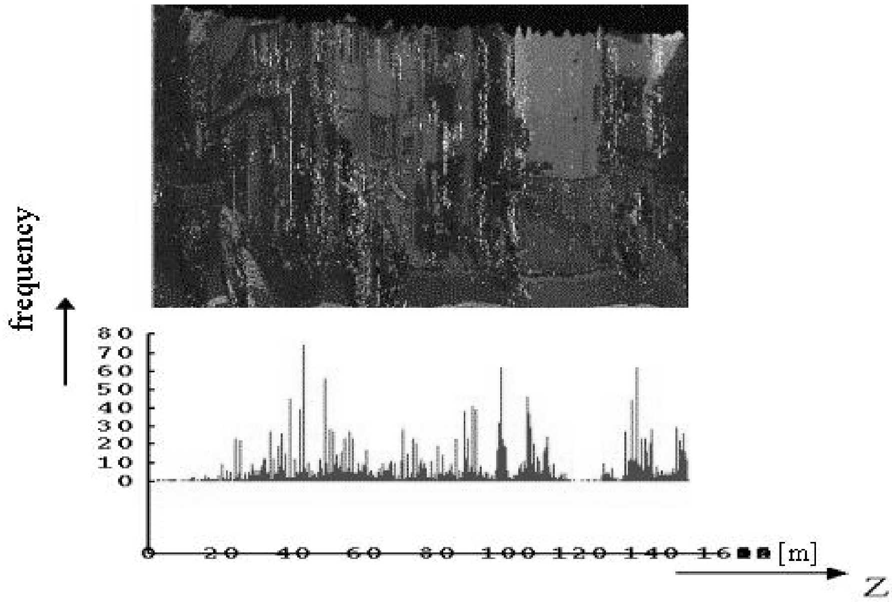
In these experiment, we used 600 consecutive input images built from the video image took by a car running along downtown (see Fig. 11). The car equipped with the gyro sensor and distance sensor in order to record vibration and to obtain the moving distance. GPS was used to record the location of the vehicle. Obtained image sequence was normalized using distance sensor.
Because of real environment, it is necessary to revise vibration for image sequence. As a result of measurement of gyro sensor, we had known that the vibration of a car running on the road is more influenced by a pitch than by a yaw and roll. Therefore, we shifted one of the two consecutive images up and down, and calculated the correlation between two images is maximized. Figure 12(a) showed the slit image before vibration removed, and Figure 12(b) showed the image after vibration removed.

Figure 11: Source image

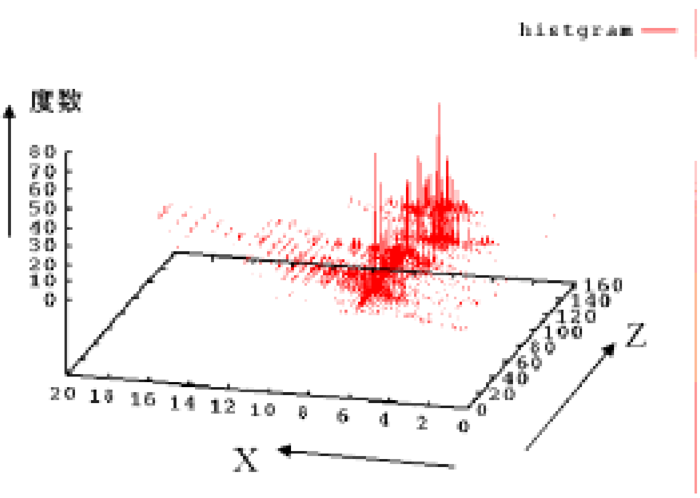
(a) Before vibration removed

(b) After vibration removed

Figure 12: Results of vibration removed



(a) 3D measured points on depth-map and histogram



(b) histogram on the X-Z plane

Figure 13: Results of histogram of 3D measured points

We analyzed the EPI to make the slit image, and build the depth map. Figure 13(a) showed the 3D measured points on the depth-map and its histogram. Figure 13(b) showed the histogram on the X-Z plane. We can judge that the peak is about 5 ~ 7 meters from Figure 13(b). This part is supposed that the building surface faces the street. And the peaks on the histogram are the prospective of building boundary in Figure 13 (a).
Using these results, we conducted the matching with the method from section 3.2. Figure 14(a) showed the image before matching. The straight lines on each image were signifying the pattern of building boundaries from a map. The lines in Figure 14(b) signify the building boundaries after matching.
By the analysis above, we constructed the urban scene model. The texture data of buildings were brought every building from Figure 14(b) using building boundary information. The shape of building was expressed a rectangular parallelepiped, and putting the texture on it. Figure 15 showed the urban scene model. This model was not expressed sky texture. And the parts of unable observation at mobile observation had no texture. These parts looked black in Figure 15. In this model, it is possible to walk through and to locate a viewpoint freely.

(a) Before matching

(b) After matching

Figure 14: Result of matching

(a)

(b)

(c)

Figure 14: Result of construct model

(a) showed the view at starting point of observation, (b) showed the view at moving point to the upper right from (a). (c) showed the view to the right forward after some advance.

5 CONCLUSIONS

This paper described the method of the 3D model construction from the motion images of urban scene by the camera on the vehicle and digital map data. In order to apply in real environment, the vibration was removed to obtain image sequence. A photorealistic urban model was realized as a result of construction model with this method.
For future research, we examine the method of expression from objects except buildings such as poles and guardrails. Also we will realize the construction of a wide area urban scene 3D map by expressing a crossing.

REFERENCES

[1] M.Hirose, S.Watanabe and T.Endo: "Generation of wide-range virtualspace using photograhics images", Proc. VRIS'98, pp.234-241, 1998.
[2] Y.Ohta and H.Tamura, "Mixed Reality", pp.183-197 (Chapter 10: "Building a Virtual World from the Real World"), 1999.
[3] H.Zhao and R.Shibasaki: "Simulation on Automated Reconstruction of Urban 3D Object using Laser Range Finder (LRF) data", Journal of the Japan Society of Photogrammetry and Remote Sensing, Vol.36, No.4, 1997.
[4]M.Notomi, S.Ozawa and H.Zen: "Construction of Urban Scene Model by Mobile Observation", Trans. IEICE, Vol.J81-D-II, No.5, pp.872-879, May, 1998.
[5]R.Bolles, H.Baker and D.Marimont: "Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion", International Journal of Computer Vision, 1, 1, pp.7-55 (1987).
[6]H.Kawasaki, T.Yatabe, K.Ikeuchi and M.Sakauchi: "Interactive System of Real-World Video Based on Self-Organized Data Structure," Trans. IEICE, Vol.J82-D-II, No.10, pp.1561-1571, Oct. , 1999.