A Wide Scope Modeling to Reconstruct Urban Scene

Shiro OZAWA, Mikito NOTOMI, Heitou ZEN
Tokyo University of Mercantile Marine
Etchujima 2-1-6, Koto-Ku, Tokyo, 135-8533
E-mail: sozawa, miki, zen@ipc.tosho-u.ac.jp
JAPAN

Commission V, Working Group V/III

KEY WORDS: Motion Stereo, Epipolar Plane Image analysis, ITS

ABSTRACT

3D model of city is one of most important and valuable in many areas, such as civil engineering, traffic engineering, education, amusement. Such models describes the shape, textures and locations of constructions i.e. buildings, bridges so on. Those models are made manually through CG modeling tools usually, hence, enormous costs must be paid and restricts a utilization of those models strongly. A more efficient or automated way must be considered to validate those models. One of promising way is to apply vision techniques, such as shape from stereo, motion. Especially, shape from motion analysis is efficient, as objects are existing widely, and it's easy to observe through vehicle running on a road. In this paper, we propose a method to modeling of city scene. We aim at sufficiently describe constructions, by their surface's depth from a road, textures, and locations. The technique we propose is based on Extended EPI (Epipolar Plane Image) that was improved over EPI that had been proposed by Bolles. The output of our method is like a range image with texture, along with the observer's locus i.e. the road that run through. We name this expression as "depth-map". The set depth-map apart from general range data is that depth-map has range from camera path. So depth-map has structure of data like a belt, and there are three data each upper, left and right that cope with one street. Conbining these depth-map and digital map, leads to many applications such as VR, walk-through systems.

1 Introduction

Modeling of urban scene is one of most expecting in many areas, such as civil engineering, traffic engineering education, amusement, etc. We introduce the motion stereo technique named Extended EPI (Epipolar Plane Image) analysis, to obtain depth. While, texture data is gathered by horizontal and vertical slits on the image plane, to form a spatiotemporal plane images. Interpolated depth values are mapped on these spatiotemporal images. These images are compared to a depth map from a path of observation, and different views are easily generated. Conbining these depth maps and digital maps, leads to many applications such as VR, walk-through systems. 3D model of city is one of most important and valuable in many areas, such as civil engineering, traffic engineering, education, amusement *etc.*

Such models describes the shape, textures and locations of constructions i.e. buildings, bridges so on. Those models are made manually through CG modeling tools usually, hence, enormous costs must be paid and restricts a utilization of those models strongly. A more efficient or automated way must be considered to validate those models.

One of promising way is to apply vision techniques, such as shape from stereo, motion *etc.* Especially, shape from motion analysis is efficient, as objects are existing widely, and it's easy to observe through vehicle running on a road.

In this paper, we propose a method to modeling of city scene. Our goal is to sufficiently describe constructions, by their surface's depth from a road, textures, and locations. In other words, it's sufficient to reconstruct a view as seen from a road and make it possible to reproduce motion parallax. The output of our method is like a range image with texture, along with the observer's locus i.e. the road that run through.

We employ EPI (Epipolar Plane Image) analysis as a basic technique to recover the surface's depth from a road, as observed from a running vehicle. Texture is collected simultaneously from same image sequence for EPI analysis. We also reorganize EPI method to be more convenient way for our observation environment. The original EPI restrict viewing angle to be perpendicular to observer's moving direction, while our modified method, named "EEPI"(Extended Epipolar Plane Image), coincides both. EPI needs multiple cameras to cover wide view, from lower portion of building to higher, our EEPI covers considerably. Hence EEPI just recover the depth of sparse points on objects' surface, we interpolate depth of the rest points.

In the following sections, we overview the needs for a city model, then explain our modeling technique with some experimental results for CG synthesized image sequence. A result with actual image taken from a vehicle is also shown.

2 Expression of City space

Modeling of urban space is becoming important with the progress of technology that indicates 3D objects, such as Virtual Reality. For instance, in civil engineering the conversely modeling that invoke a 3Dmodeler is carried out, to use simulation of scene. In the field of VR, it is necessary to model a vast space. Also needs for the numerical 3D map is increasing in the field of civil engineering, traffic engineering. In these fields, expressing a vast area is given a higher priority than expressing a limited area in full detail. But it needs a heavy cost to input for 3D modeling as things are stand, and it is important to create 3D model in efficiency term.

3D modeling of objects have been studied in the field of machine vision. Due to the progress of computer, we may be able to obtain data more than necessary through many techniques that use "a multi-view-point image" or "a series of image". And conversation systems that assort modeling and image analysis was proposed.

The 3D modeling of the 'real' objects has been a main issue in the field of machine vision, and we have seen many techniques for it. However, it is difficult to apply last techniques in the open air like in urban space, because these aim modeling of a single object.

Our study is to propose a method to get models that runs parallel to the street. The technique that we propose is based on the motion stereo technique. It is necessary to observe in many points of view, and it needs an invention to unite results that obtained at several points. In the case that objects are at a standstill such as urban scene, motion stereo is fit for use. We must obtain ego-motion in motion stereo, but we can control viewpoints such as on vehicle, or we can use some sensor except for image. The technique we propose is based on EEPI that was improved over EPI that had been proposed by Bolles, and we express city scene with texture information and range information [2][3]. We name our expression as "depth-map" that includes range data and texture data. The set depth-map apart from general range data is that depth-map has range from camera path. So depth map has structure of data like a belt, and there are three data each upper, left and right that cope with one street. Depth-map can be used to reconstruct city scene in optional view point. And depth map has structure of 2D image, then it can be a compact expression. The following chapters explain the technique to obtain depth information by the motion analysis, to interpolate depth-data in the after-processing, and to give an example of processing.

3 Modeling of city

3.1 Getting depth

There are a lot of methods to take 3 dimensions information, and an Epipolar Plane Image (EPI) analysis is one of them [2]. In this paper, we defined the camera coordinate system that a principal axis is *Z*-axis, a horizontal direction is *X*-axis, a vertical direction is *Y*-axis, and coordinate system of the image plane that the origin is center of image, horizontal direction is *u*-axis and vertical direction is *v*-axis. When we observe the stationary objects under camera moving along a track, we cut the spatiotemporal image parallel to raster, a pattern appears on the section (see Fig.1). There is a relationship between a slant of line and distance from the center of lens, as an object is far from the center of lens, so the line is close to a line that is parallel to *Z*-axis. $\varDelta D$ indicates travel between two view points on an EPI. When a slant of a line, distance from a view point to stational objects, focal distance and camera motion speed are indicated by $a = \varDelta Z / \varDelta D$, and Z_p, a next expression would be realized.

$$Z_p = f \cdot a \qquad (1)$$

By an EPI analysis, we need a lot of cameras to cover wide area, because of camera's field of vision. For example, we assume that the camera's field of vision is 36 degree, image plane size is 640x480[pix]. To get whole data around the camera path, we need ten cameras and the EPIs that we should dispose are 4800 sheets of image. Thus, in this paper, we have such a method in order to reduce the number of cameras and EPIs that we should dispose.

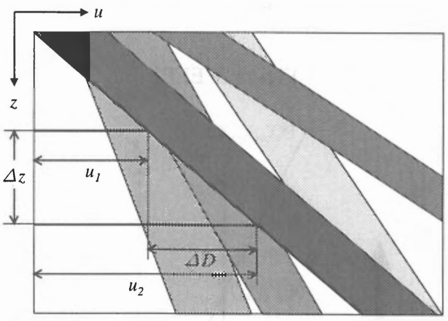

Figure 1. Feature path on EPI

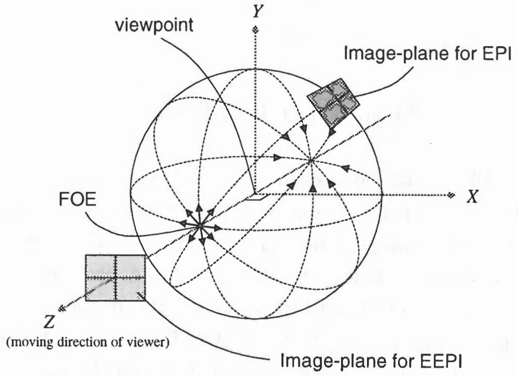

Figure 2. Flow of stationary object under straight moving

First, we move a camera under camera's principal axis is parallel to progress direction (see Fig.2). Therefore, radiate lines are extended from FOE (Focus of Expansion) on Image Plane. In this time, We can make images look like EPIs by accumulating pixels toward progress direction that is on radiate lines (see Fig.3).

The symbol 'Z' means travel distance and the 'e' means a distance from the center of image to each pixels. We call these slit spatiotemporal plane images EEPIs that are made with these radiating slits.

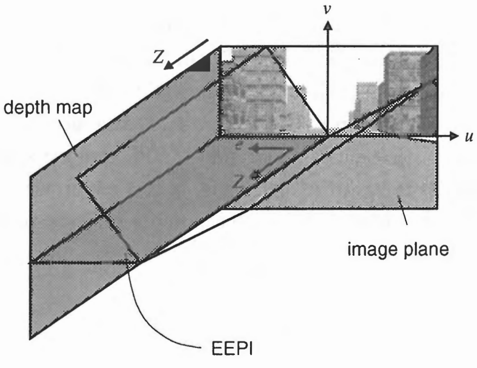

Figure 3. EEPI

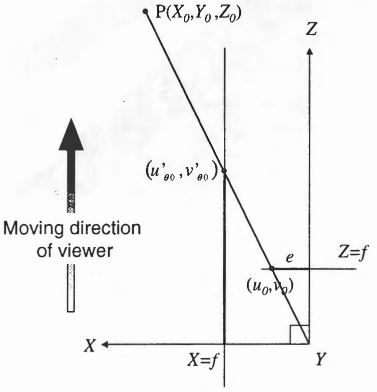

Figure 4. Translate EEPI to EPI

The EEPIs would be translated EPIs made with the camera which is fixed that principal axis is perpendicular to progress direction and parallel to the slit that makes the each EEPI. We measure distance from the camera-path to each feature by analyzing these translated EEPIs. We call this method Extended Epipolar Plane Image analysis (EEPI analysis). By EEPI analysis, the number of images which should be disposed is (640+480)×2=2240 sheets of image, thus the number of image by EEPI analysis is less than one by EPI analysis. On a coordinate system (X, Y, Z), there is a point 'P'(X_0, Y_0, Z_0).

The point P is projected a coordinate (u_0, v_0) on the image plane (u-v) of a camera whose principal axis is parallel to progress direction, the distance from the center of image plane to the projected coordinate is defined e. And an angle which the 'u'-axis and a line which links

a coordinate (u_0, v_0) to the center of image plane make, is θ. Still more, the point 'P' is projected a coordinate ($u'_{\theta 0}$, 0) on a image plane, (u'_θ - v'_θ), of an other camera whose principal axis is perpendicular to progress direction and rotate clockwise by θ toward 'X'-axis.

$$X_0 = \frac{e \cdot Z_0}{f} \tag{2}$$

$$X_0 = \frac{Z_0 \cdot f}{u'_{\theta 0}} \tag{3}$$

$$u'_{\theta 0} = \frac{f^2}{e} \tag{4}$$

Thus, equation (2) and equation (3) are lead equation (4), it can translate EEPI to EPI.

We make EEPI of each side and upper. The number of EPI is equal to the number of pixel of each side or upper end. In this paper, we dispose only the scene features whose feature paths reach to the end of EEPI. And, because equation (4) which translates EEPI to EPI is inverse proportion, as e is smaller, so the interval of $u'_{\theta 0}$ becomes weaker, and as e is larger, so the interval of $u'_{\theta 0}$ became dense. For these two features, we do EPI analysis with the area of translated EEPI those numerical value of $u'_{\theta 0}$ is nearly max. By EPI analysis, resolution of image is high. Because the objects are observed at nearest point. On the other hand, by EEPI analysis, resolution of image is low. Because the objects are observed from far to near by. Thus, from accuracy of view point, EEPI is inferior than EPI. But, we think the method is effective for the purpose that is refer to in section 1.

3.2 Making of depth map

We build up data that is the distance from view point to feature points. When we reappearance the urban scene, we need textures of building. Then, except the slits for EEPI analysis (slit for EEPI), on the ends of each side and upper, we make images by accumulating pixels that is on the three lines toward progress direction (see Fig.5). This is a left side image (see Fig.6).We call these pictures slit spatiotemporal plane images.

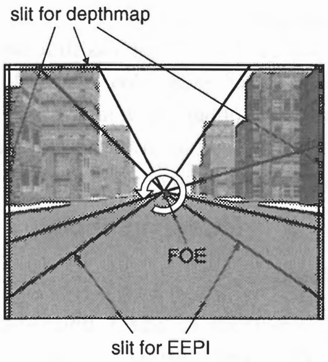

Figure 5. Slits for EEPI and depth map

We built up the data, those feature is that each distance of data correspond to each pixel on the slit spatiotemporal plane images. We built up the distance from a viewpoint to feature points are change to the distance from camera path to feature points. We call this depth map. We assume that depth maps have depth from camera path toward every pixel on slit spatiotemporal plane images.

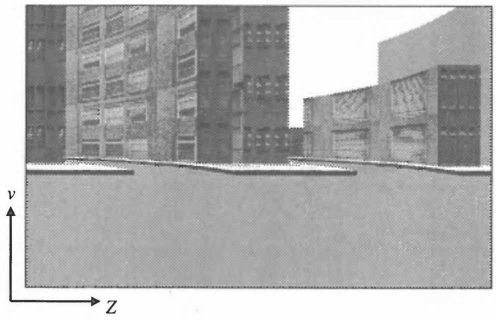

Figure 6. Slit spatiotemporal plane image for depth map

3.3 Interpolation of depth

By the method which we described in section 3.1 and 3.2, we can get the depth of whose features only appear feature path on EPI This means that we can only measure the depth of edge. So, we interpolate depthmap toward all pixels on slit spatiotemporal plane images by using the measured depth. In this paper, our aim is urban scene, so we can regard that objects are constructed of plane. So, we can interpolate it by such a method.

Firstly, segment slit statio temporal plane image by gray value of a picture. We assume that there are the pixels in a same segment, so the pixels are existing on the same plane. At this time, If there is less than three measured points in a segment, the segment connect in the neighborhood of a segment that has more than 3 measured points. Because to decide a plane, we need more than three points.

Secondly, we regard the measured data in each segment as a set of points in 3-dimension space. And, we presume a set of parameters of a plane toward a set of points. At this time, we consider that the measured data contain noise, so we need using a method, which is hard to influence by noise. Thus, we presume a set of parameters of a plane by using the LMedS (least median of squares) method, which is proposed by Rousseeuw.

Next, give depth to the pixel that don't have depth in each segment. We define that the world coordinate system is '*x*, *y*, *z*', the camera coordinate system is '*X*, *Y*, *Z*', the coordinate of pixel on the slit spatiotemporal plane image is '*i*[pix], *j*[pix]' ,the size of image plane is height *IMAX*[pix], width *JMAX*[pix],focal distance is *f*[m],a width of a pixel is *W*[m/s],camera's field of vision is 2 θ [rad], progress speed of camera is *v*[m/s],a height of camera is *h*[m],the point where the camera begins to move is (x_n, y_n+*h*, z_n), the point where the camera observe is (x_n, y_n+*h*, z_n+*jv*) .

At this time, we can say that each pixel (*i*, *j*) on a slit spatiotemporal plane image is shown by equation (5) toward world coordinate system. If (x_n, y_n, z_n) equal to (0,0,0), the equation that links a view point coordinate (x_n, y_n+*h*, z_n+*jv*) where each pixel is observed and each pixel on spatiotemporal plane image is shown by equation (6)(K is real number).

The cross point where these lines cross each presumed plane that every pixel is contained in every segment is indicated by equation (7). We interpolate the depth map with the value of cross point.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_n + f\tan\theta \\ y_n + h + W\left(\dfrac{IMAX}{2} - i\right) \\ z_n + jv + f \end{pmatrix} \tag{5}$$

$$\frac{x}{\tan\theta} = \frac{f(Y-H)}{W\left(\dfrac{IMAX}{2} - i\right)} = z - jv(= K) \tag{6}$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} K\tan\theta \\ KW\dfrac{\left(\dfrac{IMAX}{2} - i\right)}{f} + h \\ K + jv \end{pmatrix} \tag{7}$$

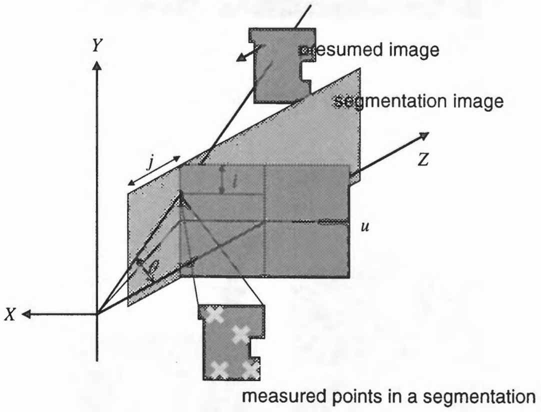

Figure 7. Method of interpolation

4 Experiment

4.1 Experiment with animation image

We assume that the images are taken by mounted camera on a vehicle. We experiment with 400 sheets of animation image whose size are 640×480 (see Fig.8). This picture (see Fig.9) is a sample EEPI and translated EPI which is translated from EEPI. Fig.10 and Fig.11 are results of 480 sheets of left side's EPI. White line shows detected lines on EPI (see Fig.10). Depth measured by detected lines is shown in Fig.11. In Fig.11, *Z*-axis indicates progress direction of camera, *X*-axis is left side, squares are buildings, 'Xs' are depth from camera path.

Table 1. compare measured value with real

segmentation	a	b	c
mesuared[m]	9.50	9.52	8.36
real[m]	8.00	8.00	8.00
error[%]	18.75	19.00	4.50
dispersion	3.39	3.23	3.67
points	681	403	465

Table 2. compare interpolated depth with real

segmentation	a	b	c
mesuared[m]	8.48	8.65	7.65
real[m]	8.00	8.00	8.00
error[%]	6.00	8.13	4.38
dispersion	8.00	16.30	2.65
points	8385	6048	4230

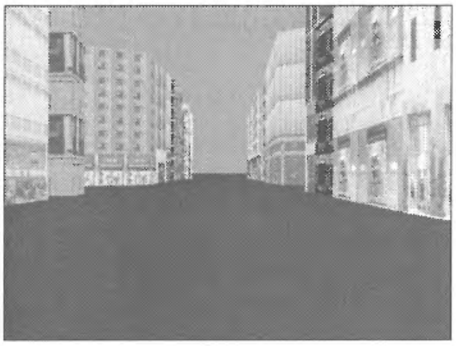

Figure 8. Animation image

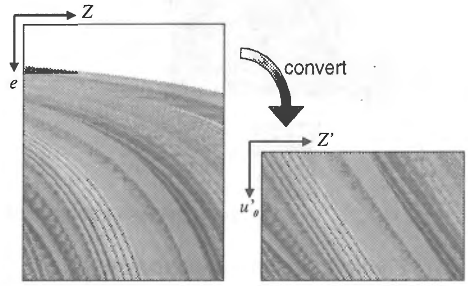

Figure 9. EEPI and converted EPI

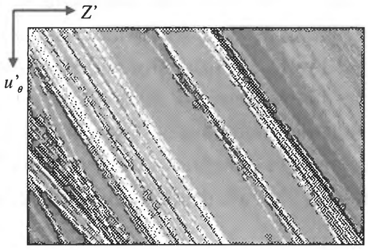

Figure 10. Detected lines on EPI

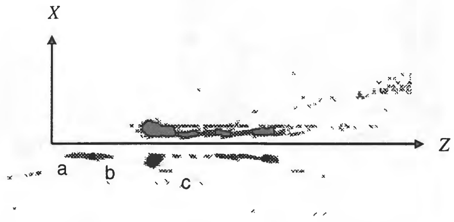

Figure 11. Result of distance measurement

We compared models with depth measured by EEPI analysis, and interpolated depth. Table.1,2 indicate errors at three positions(a, b, c). Compare Table.1 with Table.2, we think that LMedS method was useful. Because errors reduced 18.75%, 19.00% and 4.5% to 6.0%, 16.30% and 4.38%.

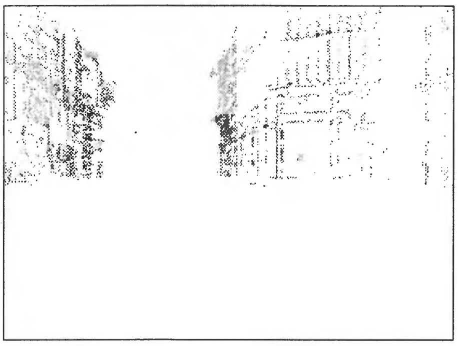

Figure 12. Rebuilt image

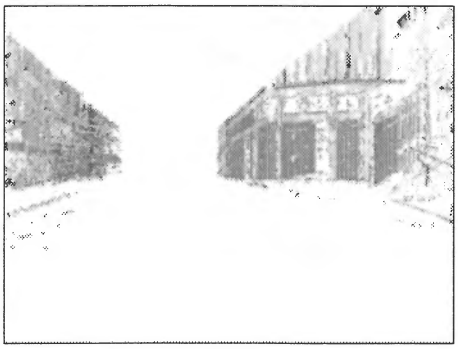

Figure 13. Rebuilt image with interpolated depth map

The rebuilt image made by the depth map that is made with only EEPI analysis is indicated with Fig.12. Furthermore, the result image made by the depthmap that is interpolated is indicated with Fig.13.

4.2 Experiment with real image

We experimented with real images that are taken by camera mounted on a vehicle. Fig.14 is a part of images. Detected lines on real image are indicated with Fig.15. Parts of result plotted on map are indicated with Fig.17. The locus of feature doesn't become a line, because at this time, we don't amend images from motion of vehicle. Then anytime it couldn't detect lines. It's notable that the feature becomes longer from camera path. Fig.16 shows a part of result. a, b and c correspond to street-lamps. Near the point of d, there correspond the left second building in Fig.14Fig.17 shows reconstructed image by measured data, and Fig.18 shows reconstructed image by interpolated data.

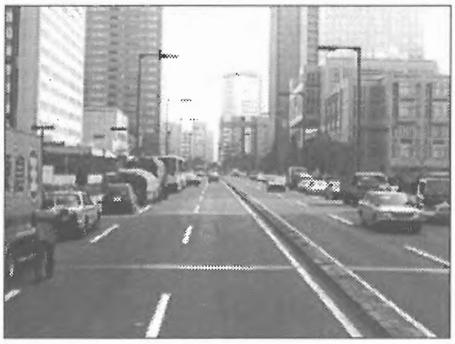

Figure 14. Real image by mounted camera on a car

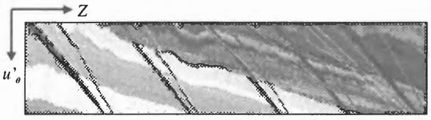

Figure 15. Detected line on EPI

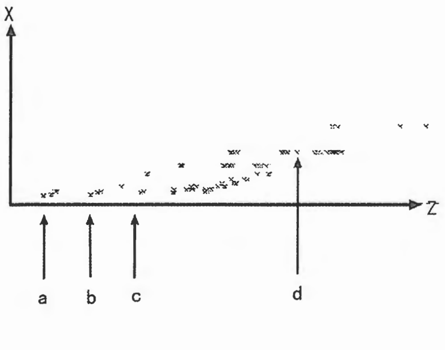

Figure 16. Result of distance measurement

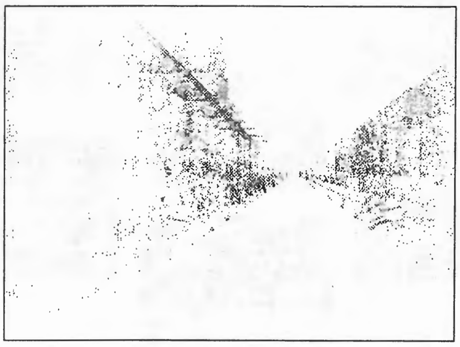

Figure 17. Rebuilt image

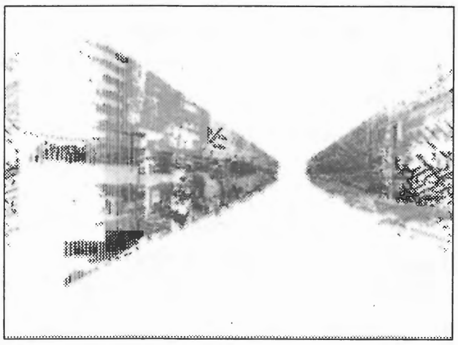

Figure 18. Rebuilt image with interpolated depthmap

5 Concluding remarks

Our study proposed the technique for modeling of urban space. We can obtain range data of objects by EEPI analysis to serial images from a vehicle. At the same time we can obtain texture data of objects by horizontal and vertical slits on the image plane. And we use divided spatiotemporal image to interpolate data that we couldn't obtain by EEPI analysis of depthmap. The experiment on animation images shows range errors from 4.5% to 19.0%. After this, we will cope with some problems as processing real images and raising accuracy. In particular it is necessary to obtain ego-motion by some way in processing real image because it is difficult to uniform velocity linear motion in real scene. Currently we use reconstructed vehicle (Fig.19) that carry sensors of translation (rotary encoder) and velocity of angle (gyro sensor)to obtain ego-motion. But it is difficult to obtain ego-motion only these sensors, then we will develop a new technique that can permit some measure of tremors. And we will improve process of interpolation with Voronoi tessellation (Delaunay net).

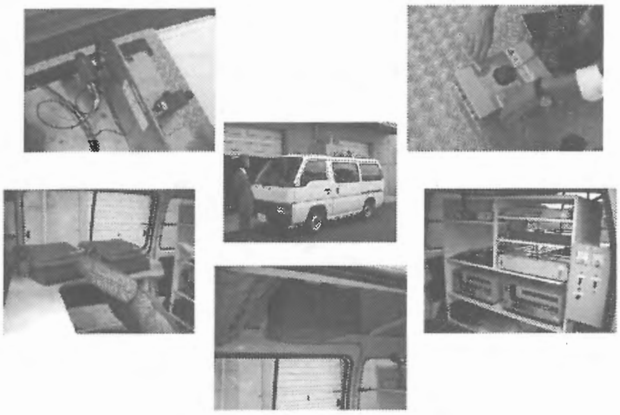

Figure 19. Vehicle for EEPI analysis

References

[1] C. Toamsi and T. Kanade: "Shape and motion from image streams under orthography: A factorization method", International Journal of Computer Vision,9,2,pp.137--154(1992).

[2] 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).

[3] H. Zen,M. Nohtomi and S. Ozawa: "Virtual City Space:A Construction of 3D City Scene through Moving Image Analysis",Technical Rreport of IEICE,(PRMU96-126),96,436,pp.~59--65(1996).