AUTOMATIC 3D CITY CONSTRUCTION SYSTEM USING OMNI CAMERA

Hiroshi Kawasaki　*Katsushi Ikeuchi*　*Masao Sakauchi*

Institute of Industrial Science, University of Tokyo
7–22–1 Roppongi, Minato-ku, Tokyo, 106-8558, Japan

ABSTRACT

Video analysis and mixed reality (MR) systems, which integrate the virtual world and the real world, are important topics of multimedia research. In this paper, we propose an efficient 3D city construction system from omni video data as a practical application of MR systems. To develop this system, two main methods are proposed for analyzing omni video data. The first one is omniEPI analysis, which makes it possible to obtain 3D information from omni video data robustly and efficiently. The second method is omniPVI analysis, which makes it possible to obtain not only front images of the buildings but also side images of them. To demonstrate the effectiveness of these methods, we present some experimental results and the prototype systems using outdoor environment omni video data.

1. INTRODUCTION

Video analysis has been an important topic of multimedia research and, with the recent progress in the computational efficiency and expansion of storage capacity, it has become more useful and realistic. As a result of these improvements in video analysis, many projects were proposed and subsequently realized, for example, the INFORMEDIA project at CMU[1]. Recently, mixed reality systems, which integrate the virtual world and the real world, have come to be regarded as promising multimedia applications; for these, many new technologies and devices have been developed. With this in mind, we have proposed and developed automatic 3D city construction systems from video data as mixed reality applications [3]. To make a 3D city map from video data, video analysis is needed, especially a 3D information retrieval technique and an efficient video database structure; we have developed both of them.

Although these work well, some problems still remain, for example, we are to obtain images of the tops and sides of buildings because the camera's usual field of view is limited to a narrow range and the assumed 3D information does not have sufficient accuracy.

In this paper, we describe how we solved the former problem: we adapted the omni-directional camera's field of view to be wide enough to obtain images of the entire buildings. We also describe how we solved the latter problem by developing a new 3D information retrieval technique which is based on both EPI and model-based analysis.

Section **2** describes how to obtain 3D information from omni-images using EPI analysis and Section **3** presents an actual process for acquiring texture images from omni video data. Section **4** shows our 3D city construction system to demonstrate the effectiveness of our method in acquiring 3D information as well as in constructing virtual city maps from those images. Section **5** concludes the paper.

1.1. 3D information retrieval

In the past, various researchers have investigated the possibility of the acquisition of three dimensional (3D) information from video data. These methods are classified into two types: the EPI analysis [2] and the factorization method [6].

The EPI method recovers depth information from known motion such as speed and route of the camera and, in the actual analysis, it is usually assumed that the camera speed is constant and the route is straight.

In contrast, the factorization method has no restrictions on camera motion. This method uses a feature point and collects these feature coordinates into the measurement matrix to compute shape and motion simultaneously.

However, when we apply these methods to a real-world environment, it is difficult to acquire 3D information due to noise and many obstacles. And the object itself sometimes consists of complicated structures instead of plane surfaces and also may contain many complex textures. These factors often result in difficulties in stable extraction of the feature points.

Based on the facts: that extracting feature points in the real-world is usually very difficult and that the target video satisfies the constraints of EPI (constant speed and a straight route), we adopted the EPI analysis for our research rather than the factorization method.

1.2. Omni directional camera

The omni directional camera is a camera designed to take an omni-directional view of the environment all at once by taking the reflected view on a symmetrical mirror [5] [7] [4]. Many kinds of omni cameras are now available; we selected the paraboloidal mirror for our analysis.

The omni camera with a paraboloidal mirror (Fig.2(a)) has only one focus and can make an orthogonal projection image from the mirror and can easily reconstruct the perspective view. The paraboloidal omni camera can take one hemisphere of the environmental view at a time.

2. RETRIEVAL OF 3D INFORMATION

In this section, we propose an efficient method to obtain 3D information from omni video data using EPI analysis. However, there remain some significant difficulties in applying this technique to the real-world environment as opposed to using this in the experimental indoor laboratory environment. So, we propose the model-based analysis in addition to the EPI analysis to achieve robust and accurate results. In the following section, we describe both EPI and model-based analysis, then show the actual process of the retrieval of 3D information from omni video sequence(s).

2.1. Model-based EPI analysis

Under the circumstances that the camera speed is constant and the camera is moving along a straight route, the traces of objects describe straight lines on EPI. The coefficients of these straight lines represent the depth of the objects; we can determine the depth of the objects by assuming the incline of these lines.

On the other hand, the EPI analysis is sometimes very difficult to apply to real world data, because keeping the camera speed almost constant and moving along the route in a straight direction is difficult in reality. Also, objects in the real world have complicated shapes and textures.

Further, we can obtain models for most objects in the real world, especially models of buildings. And in the actual situation, when we use the video data taken from the vehicle as it travels along the street, we can use map data as a model.

Considering these situations, we propose a model-based analysis to avoid the difficulties of image processing caused by real world complexity. To put this analysis concretely, we perform matching between video data and models, and then retrieve 3D information from video data using the matching results.

2.2. Omni EPI

In this section, we show how to make EPI from omni video data. Usual EPI is a cross section of spatio-temporal volume made from ordinary video data and horizontal plane. To employ the same method to omni video data, we first make spatio-temporal volume from omni images; then we cut this spatio-temporal volume at a curved plane which is equivalent to the horizontal plane in the perspective image. In the following section, we describe the characteristics of this spatio-temporal volume made from omni images (STVO) and actual shape of curved plane. We also show omni EPI made from omni video sequence.

2.2.1. Spatio-Temporal Volume

By accumulating omni images along the time axis, we can obtain a spatio-temporal volume of omni images (STVO) as shown in Fig.1(a). For the analysis of this 3D volume, we can generate several different kinds of cross sections in this volume, and extract useful characteristics of traces of image features. In this paper, we describe how we cut this STVO at the stream line and the radius line; each line is equivalent to the horizontal and vertical lines in the perspective image, respectively. We will discuss the stream line in this section and the radius line in Sec. **3**.

2.2.2. Locus on Omni

The paraboloidal omni camera, the use of which is described in this paper, has a paraboloidal mirror, a solid of revolution generated from a parabolic curve. The equation of a paraboloidal mirror is represented as:

$$\frac{x^2+y^2}{h^2}+\frac{2\cdot z}{h}=1 \tag{1}$$

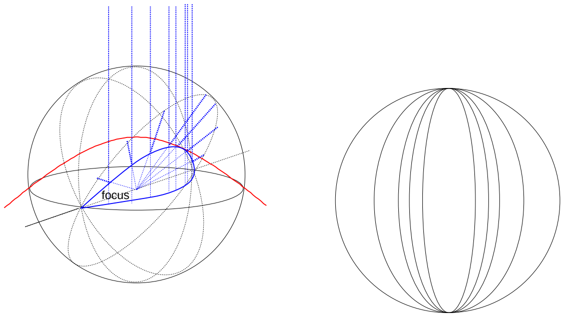
Here, we use the same parameterization as [5]. As shown in Fig. 2, under this parameterization, all the incoming rays toward the focus of the paraboloidal mirror become parallel and orthogonal against the image plane of the camera after being reflected by the mirror.

The shape of a stream curve is also an elliptic curve and its equation is expressed as:

$$\frac{x^2}{h^2\sin(\theta)}+\frac{y^2}{h^2}=1 \tag{2}$$

(a) (b)

Figure 1: (a)Hyperboloidal omni spatio-temporal volume (b) Cross section of spatio-temporal volume of omni



(a) (b)

Figure 2: (a)model of paraboloidal mirror (b)locus of horizontal line on omni-camera

2.2.3. Examples

By cutting the STVO along a stream curve (eq.2), we can obtain cross sections, as shown in Fig.3. Here we denote those cross sections as a parabolic EPI.

In the STVO, a distant feature moves slowly, while a nearby feature moves rapidly. Due to the difference in the distance, a trace of an image feature, as shown in Fig. 3, has an S curve on a cross section. By applying a perspective transformation, we can successfully transform this S curve into a straight line as shown in Fig. 4 . This transformed EPI is equivalent to those defined in [2].

2.3. ModelEPI

Since the real-world video is taken from a vehicle moving along the street, we have to create the model EPI corresponding to this video data. The procedure for making model EPI is as follows.

1. Assume the path along which the vehicle travels.
2. Extract polygons of the buildings which lie within a certain distance from the path.
3. Create a 3D map using these extracted polygons.
4. Render the sequential images taken by the virtual camera which is installed in the virtual vehicle traveling along this assumed path with constant speed.
5. Create EPI from these sequential images.

Figure 3: omniEPI(i)

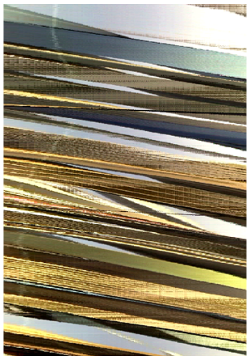


Figure 4: omniEPI(ii)

2.4. EPI - EPI matching

We define EPI as the common basis for matching. This EPI-based matching has some advantages compared with the usual EPI analysis to extract 3D features from EPIs. Namely, the route does not have to be exactly straight and the camera motion does not have to strictly keep the velocity constant, because this EPI-based matching compares regions generated from both real and simulated EPIs. This method does not need precise line detection because we are not going to extract parameters of lines; rather we will compare regions bounded by these lines for matching.

Thus, the 3D information retrieval problem results in a 2D matching problem. By assuming that the order of objects does not change, we can use the DP matching. Next we show the method for DP matching adapted for the 2D pattern.

2.4.1. 2D DP matching

We must carry out the matching between EPIs made from the omni-video camera and a model. EPI is a 2D image, so we have to employ the matching between 2D images. In this paper, we propose a 2D image matching method using the DP matching; the procedure for this method is as follows.

1. Make a rectangular parallel piped from two EPIs(Fig.5)
2. Cut this 3D polygon by a horizontal plane and make a 1D search plane
3. Carry out 1D DP matching (usual DP matching) for each 1D search plane(Fig.6)
4. Iterate **1** to **3** using the constraints so that the matching path makes a continuous plane

2.4.2. Result of 2D DP matching

To achieve high accuracy and robustness of the matching, we use three different patterns for matching, such as the edge of the building, the boundary of the building and the sky pattern. Fig.7 and 8 shows the result of DP matching. The former figure shows all the matching paths, and the latter figure shows only one selected path from 2D matching results.

When see the Fig.7, we can identify that all matching paths make a smooth curved surface because of the constraint that the all matching paths are continuous to any direction. In Fig.8, we can see that the matching result still has small errors, but nevertheless achieves sufficient accuracy for our proposed system.

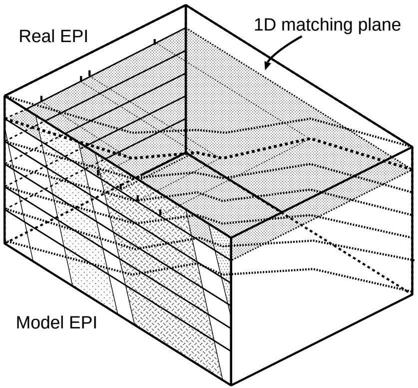


Figure 5: Concept of 2D DP matching

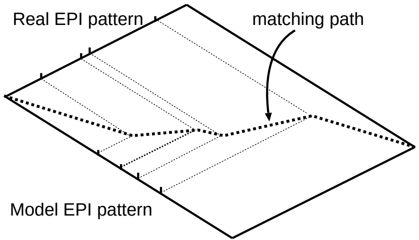


Figure 6: 1D search plane of 1D DP matching

3. OMNI PVI ANALYSIS

A vertical line in the 3D space is projected as a straight line on an omni image, along a radius direction. In this section, we describe how we will cut a STVO along a radius line. This operation is equivalent to the one to extract vertical slits of a usual perspective image and collect those vertical slits into an image. Thus, by applying this operation to the STVO, we can obtain a panoramic view image(PVI) defined by [7].

We can select any cutting line from infinite candidates of radius lines. All of these cutting lines make different PVIs. Usual PVIs, generated from perspective images, are restricted by the angle of the field of view of a camera. However, an omni camera has a 360 degree field of view. Thus, by using an STVO, we can generate any PVI toward any viewing direction. This is the great advantage of STVO in retrieval of texture images and map-making.

We will show three different PVIs given by three different radius lines (Fig.9, 10). We can see the difference in these images, depending on view directions.

4. 3D CITY MAP

We have performed some experiments to demonstrate the effectiveness of our method. We apply the model-based EPI method mentioned in Sec.**2.3** to omni video data and make 3D city map using the 3D information acquired by this method. The target omni video data we use in this experiment is taken from a vehicle which runs along a city street. And we use a 2D digital map as a model.

Outline of actual process is as follows.

1. Obtain 3D information from omni video data using model-based EPI analysis.
2. After restoring 3D information to video data, and simply cutting the PVI according to the acquired 3D information, we can obtain texture images of each building.

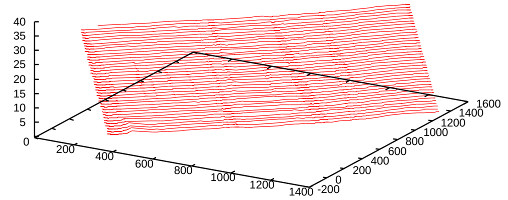


Figure 7: Result of the 2D DP matching

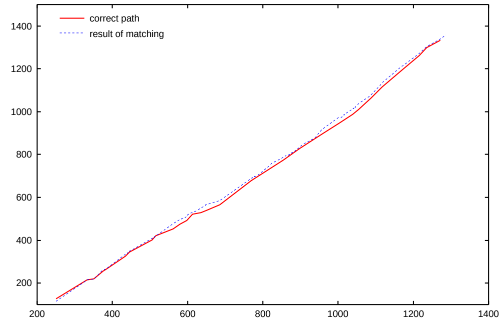


Figure 8: 1D search plane. The vertical axis represents the coordinate of EPI from the video data and the horizontal axis represents the coordinate of the map data. If the model data is precisely corresponding to the real world, this matching path describes a straight line from the origin of the coordinate axes to the end point. But there are some errors in the map data, and the EPI made from video also contains some distortion errors. Thus, the actual path does not describe a straight line; rather, it describes the line as a solid line in the figures. A broken line is the matching result obtained by our matching method.

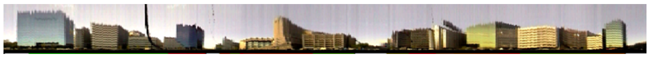


Figure 9: PVI plane 5

Figure 10: PVI plane 19

3. Make 3D city map from 2D digital map using openGL library and put texture images on these 3D polygons

We will show a prototype of the 3D city map construction system, which is successively developed by using our 3D information retrieval method. Fig.11 and 12 show the sample snapshots of a 3D city map. In Fig.11, we can see the whole textures of the buildings whose the top images are obtained, and in Fig.12, we can see that the texture of the side view is successfully mapped on the polygon of the buildings.

5. CONCLUSION

We developed an automatic 3D city construction system for mixed reality (MR) applications. To obtain whole texture images of the buildings, we adapted an omni video camera which has a wide field of view. And to obtain the side images of the buildings, we used 3D information acquired from omni video data.

To acquire 3D information robustly and accurately, we proposed an model-based EPI analysis which used 2D digital maps as a model and was achieved by matching between omni EPI and model EPI.

Figure 11: Virtual 3D map from video.

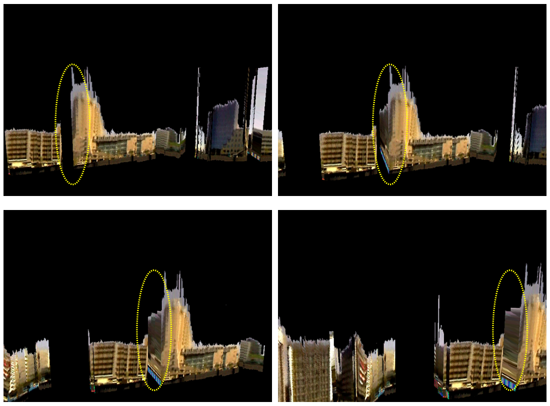


Figure 12: Snap shot of Virtual 3D map

To make omni EPI from omni video data, we made spatio-temporal volume from omni video data (STVO) and cut this volume at the stream line which is equivalent to the horizontal line in perspective image.

We also proposed an efficient matching algorithm between EPIs, that is, 2D DP matching to improve the matching result.

To demonstrate the effectiveness of our proposed method, we conducted several experiments using real-world omni video data. And the results of the experiments and snapshot images of the systems show that our proposed method and matching algorithm is successfully working.

6. REFERENCES

[1] The informedia digital video library project at carnegie mellon university. `http://www.informedia.cs.cmu.edu/`.

[2] R. Bolles, H. Baker, and D. Marimont. Epipolar plane image analysis:an approach to determining structure from motion. *Int.J.of Computer Vision*, 1:7–55, 1987.

[3] H. Kawasaki, T. Yatabe, K. Ikeuchi, and M. Sakauchi. Automatic modeling of a 3d city map from real-world video. In *ACM Multimedia99*, Oct. 1999.

[4] Y. Onoue, K. Yamasawa, H. Takemura, and N. Yokoya. Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding*, 71(2):154–165, Aug. 1998.

[5] N. K. Shree. Catadioptric omnidirectional video camera. In *Computer Vision and Pattern Recognition*, pages 482–488, June 1997.

[6] C. Thomasi and T. Kanade. Shape and motion from image stream under orthography: A factorization method. *Int.J.of Computer Vision*, 9:137–189, 1992.

[7] J. Y. Zheng and S. Tsuji. Panoramic representation of scenes for route understanding. In *International Conference on Pattern Recognition*, pages 161–167, June 1990.