Town digitizing for building an image-based cyber space

Satoshi Koizumi[1], Guiming Dai[2] and Hiroshi Ishiguro[2]

[1] CREST, Japan Science and Technology Corporation
[2] Wakayama University
Sakaedani 930, Wakayama 640-8510, Japan

Abstract. This paper proposes a town digitizing for building an image-based cyber space. The cyber space consists of omnidirectional cameras and a large number of omnidirectional images connecting them. We discuss on details of the town digitizing technique and possibilities of the cyber space in this paper. The technique is devided into three steps: omnidirectional image acquisition, parameter estimation among omnidirectional images, and smooth interpolation among them. The omnidirectional images are taken using the powerwheel or the tripod with the omnidirectional camera. Then the parameters among omnidirectional images are estimated by Hough transform and template matching. Based on the estimated parameters, the system smoothly interpolates the omnidirectional images and generates continuous virtual views. That is, the system allows users to move in the virtual space like a previous 3-D graphics system. This paper also shows the validity of the town digitizing by building a model of Kyoto city.

1 Introduction

One of the important issue of a virtual space is the quality of the views. Better quality of the views brings better reality to the virtual space. Almost all of previous works build the virtual space based on 3-D geometrical models of the environment. Therefore, it takes costs for measuring the environment and building the 3-D models[2]. For example, it takes a couple of years to build a city model. In addition to the cost for building 3-D models, the previous works[1, 3, 4] need to render images to the 3-D models in order to obtain the photo-quality views. That is, they take three steps: measuring the environmental structure, acquiring 3-D CAD models and rendering photos taken in the real world.

On the other hand, this paper proposes a new image-based method to build a virtual space. Our technique, called town digitizing, use an omnidirectional camera developed by [7]. By taking omnidirectional images in the real world with the camera, we get visual information that is sufficient for building the virtual space. The omnidirectional images are smoothly interpolated and generate view sequences for exploring in the virtual space. Figure 1 shows a comparison between the 3-D model-based method and the omnidirectional image-based method.

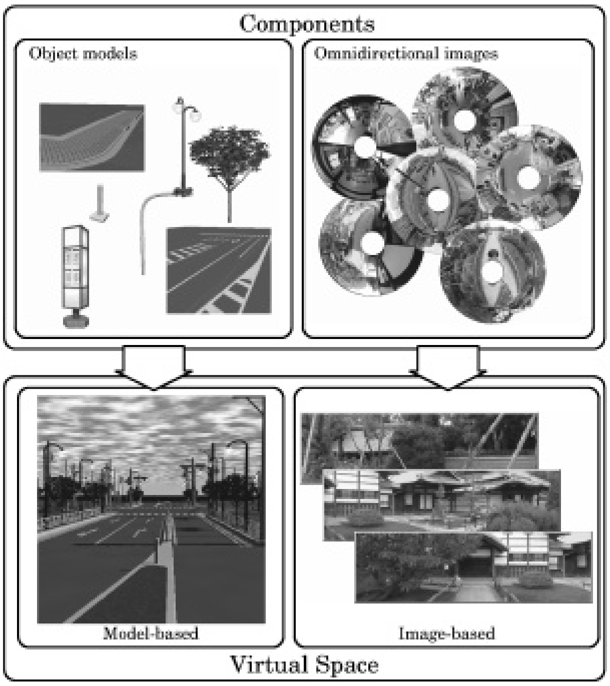


Fig. 1. 3-D model-based V. S. and Image-based V. S.

A demerit of the image-based method is the high memory cost. However, recent progress of computer hardware provides us sufficient capacity in the hard drive and onboard memory. We consider the image-based method will be one of the solutions for building the virtual space.

The town digitizing technique allows us to build a photo-realistic city model quickly. It consists of the following three steps.

(1) Image data acquisition with an omnidirectional camera
(2) Parameter estimation among the omnidirectional images
(3) Virtual view generation by smoothly interpreting the images based on the estimated parameters.

Especially, the parameter estimation is important in this technique. This paper proposes automatic methods using visual properties of omnidirectional images. In addition to the town digitizing technique, we have built a Kyoto city model by the technique. This paper shows the impressive experimental results in Kyoto city.

2 Image-based cyber space

A town digitizing technology is an important fundamental one that influences the reality of a virtual space. The conventional method to build town models needs much computational power to measure the geometry. On the other hand, the digital image with high resolution can provide rich information which enables us to feel the reality of objects in the environment. Based on the digital images, we build a virtual space. Figure 2 shows the conceptual design of the virtual space. This virtual space has connections to the real world by the omnidirectional cameras. Our purpose is to build a platform that connects networked-virtual worlds and the real worlds. Therefore, we call it "cyber space platform".

The platform consists of three components. The first one is a digital omnidirectional image. The image is able to get the environmental information around of a place where the image is taken. In Figure 2, the omnidirectional image represents as a black sphere. The next component is a technique to connect two omnidirectional images that are taken in an interval. This technique smoothly interpolates them and generate white pipe between black spheres in Figure 2. Finally, the third component is an omnidirectional camera to acquire the real-time information. The camera links to the real world and it is represents as the blue sphere in Figure 2. We call the camera Town Camera.

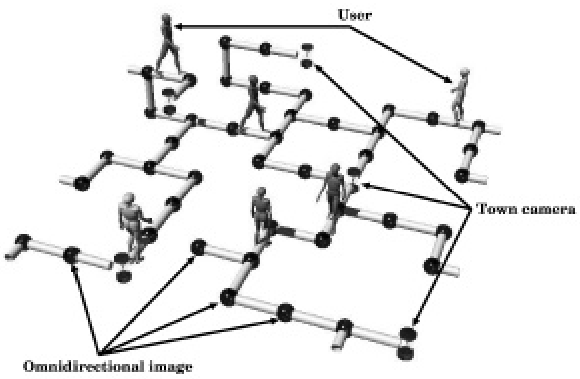


Fig. 2. Conceptual figure for the cyber space

3 Town digitizing technique

The town digitizing technique needs several software components. Figure 3 shows the configuration of a software system for the town digitizing. The system is divided into three parts: omnidirectional image acquisition, parameter estimation among omnidirectional images, and smooth interpolation among them.

An important thing when we take the image is to acquire them with the same quality. To maintain the same conditions for taking an omnidirectional image, we have developed two input devices, a powerwheel-type device and a tripod-type device. The next thing that we need to do is to interpolate among omnidirectional images. For this purpose, we estimate the parameters among them by using Hough Transform and template matching. By smoothly interpolating among them based on the parameters, users can move in the virtual space like a previous 3-D graphics system. This section describes the detail of the system.

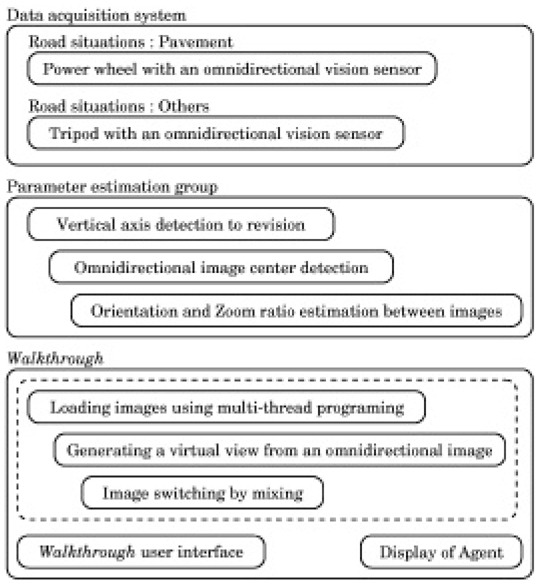


Fig. 3. Configuration of Town digitizing system

3.1 Image Data Acquisition

We have prepared two types of the device according to the situations as shown in Figure 4. We use the powerwheel on pavement roads and a tripod for others. The powerwheel has a 2-meter pole with an omnidirectional camera on the top. Figure 5 shows the attached omnidirectional camera. The user can take omnidirectional images with a cable release while sitting on the powerwheel. Further, by using a notebook computer, the user can always verify whether omnidirectional images keep the same quality. If it is difficult to use the powerwheel, the User uses the tripod.

Fig. 4. Image acquisition device of the powerwheel and the tripod type

Fig. 5. Attached omnidirectional camera

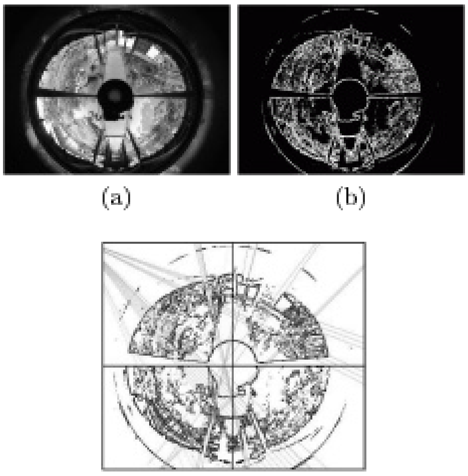
3.2 Parameter estimation

For building an image-based virtual space using omnidirectional images, the following six parameters are necessary.

- Parameters depending on an omnidirectional image.
  1. Tilting angle of the vertical axis of the omnidirectional camera.
  2. Position of the center point of the omnidirectional camera on the image.
- Parameters between two omnidirectional images.
  1. Direction between the observation points.
  2. Angle of elevation between the observation points.
  3. Zoom ratio which is in proportion to the distance.

The parameters depending on an omnidirectional image are necessary for generating virtual views from the omnidirectional image and for smoothly interpolating among them. The generated virtual view is distorted if these parameters are not precisely estimated. In the following sections, we introduce automatic methods for the parameter estimation based on visual properties of omnidirectional images.

Parameters depending on an omnidirectional image The algorithm[8, 9] transforming an omnidirectional image into a rectilinear images are sensitive to the tilt angle of the rotation axis of the omnidirectional camera and the position of the projection on the image plane. To decrease the distortion of the virtual view, the proposed system automatically estimates them.



(a) (b) (c)

Fig. 6. (a): Omnidirectional source image, (b): omnidirectional binary edge image, and (c): detected lines by using conventional Hough transform

The automatic method is divided to three steps: edge detection by Sobel filter, line-segment detection by Hough transform, and estimation of an intersection of the detected lines. First of all, we extract the omnidirectional edge-image by using Sobel filter. By discreminant analysis of the histogram of the edge image, the omnidirectional binary edge image is obtained. The source image and the binary image are shown in Figure 6-(a) and -(b), respectively. The reasons why we pay attention to the edge image are: vertical components of the object, like a pillar, around the omnidirectional camera represents the direction of the gravity in almost all environments, and their projection radiate from the center of the omnidirectional image. If the rotation axis is tilted, the center of those radial edges is shifted.

Here, we detect those radial edges as lines by conventional Hough transform. Figure 6-(c) shows the result of the detection for the binary edge image(Figure 6-(b)). We transform a pixel on the detected lines with Gaussian Distribution into the parameter space. By computing the summation of the pixel value, we obtain the intersection of them. That is, the position of the pixel that has the maximum pixel value represents the tilt angle of the rotation center.

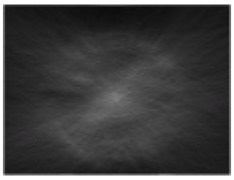


Fig. 7. Detected center point at the omnidirectional image by generalized Hough transform

Since the shape of the projections of the black needle and the hyperboloidal mirror of the omnidirectional camera is circle on the omnidirectional image, the center point at the omnidirectional image estimates by extracting their positions. Using generalized Hough transform for the omnidirectional binary edge image, we detect those circles. Figure 7 shows the parameter space of generalized Hough transform, where the parameters are the position of the center. In the image, the point of the pixel which has the maximum value is the position of the center point on the omnidirectional image. Based on these estimated parameters, the system revises the omnidirectional image by geometric operations. As the result, the system can generate virtual views without distortion from the omnidirectional image.

Parameters between two omnidirectional images Parameters between two omnidirectional images are also important in the smooth interpolation of them. There are three parameters: direction, angle of elevation, and zoom ratio. Figure 8 shows the geometrical relationship between two omnidirectional images. $X_iY_iZ_i$ and $X_jY_jZ_j$ -coordinates mean the coordinate in the omnidirectional camera at the each observation point(i-th place and j-th place) where a user takes i and j -th image. $Z_i(Z_j)$ is on the optical axis of the omnidirectional camera. θ_{ij} and ϕ_{ij} show the direction and the angle of elevation between the i-th point to the j-th point. Another parameter, the zoom ratio, depends on the view angle ψ_{ij}. Assuming that a distance from the focal point in the hyperboloidal mirror to the projection window and the pixel size of the projection window are constant, we can compute the zoom ratio R_{ij} as,

$$R_{ij} = \frac{1}{\tan \psi_{ij}} \tan \frac{\psi_0}{2} \tag{1}$$

where ψ_0 is a initial view angle, i.e. the zoom ratio is 1.

In Figure 8, there is only one straight line which links a focal point C_i at the i-th point to a focal point C_j at the j-th point. Both of omnidirectional images always take same scene which is on the straight line. Then the view projected by the window $A_i(B_i)$ and the view projected by the window $A_j(B_j)$ take the same scene roughly, where the angle between A_i and B_i is 180 degree. However, the size of the object in the view is different because of the distant observation points each other. By adjusting the zoom ratio, their size can be adjusted.

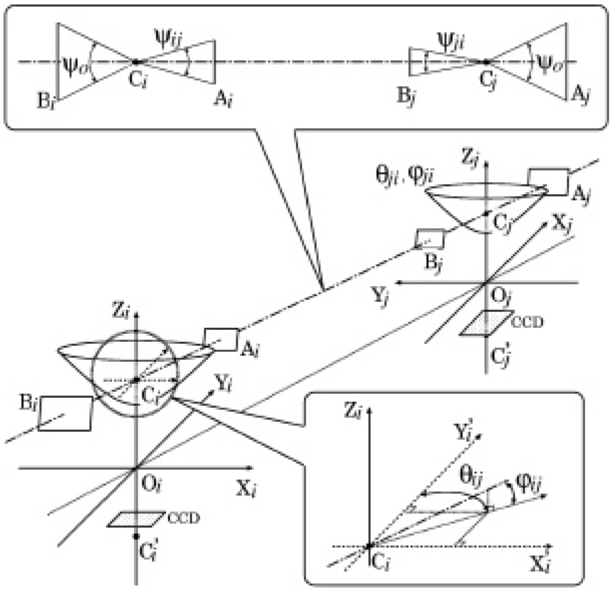


Fig. 8. Relationship between two omnidirectional images(vision sensors)

Based on the constraint, the proposed system computes cross-correlation between the projection window A_i and A_j and between B_i and B_j. By detecting the maximum value of the cross-correlation, the system can estimate the parameters.

Figure 9 shows the system window of the software for automatic parameter estimation. This software displays four views by window A_i, B_i, A_j, and B_j with the estimated parameters. We can check the result of the estimation from these views.

Fig. 9. System window of the software for parameters estimation

3.3 Smooth Interpolation

To allow users to move in the image-based virtual space, the system displays virtual view sequences interpolated smoothly. We have developed a viewer system that users can operate simply and understand their positions in the virtual space. The viewer system consists of three routines (shown in Figure 3). The first routine generates virtual views. The second one is the user interface to explore the virtual space. The third one displays the agent as the user to show the user's position in the virtual space.

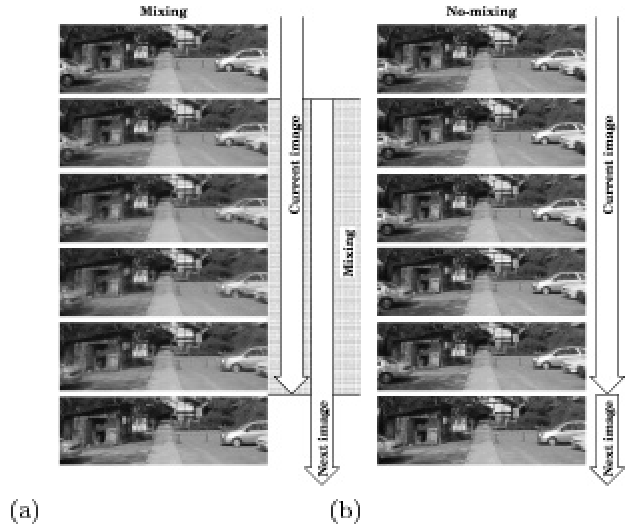


Fig. 10. Virtual view sequence with the mixing process and without the mixing process

The first routine is divided into three processes: the image loading process by using multi-thread programming, the virtual view generation process, and the mixing process to switch images. The resolution of an omnidirectional image is low around the image center. Therefore, a high-resolution camera is needed. However, according to the resolution of image, the data size becomes larger and the cost for loading is higher. The multi-thread programming solves this problem.

The second process generates virtual view sequences by increasing the zoom ratio. By increasing the zoom ratio, the system can allow users to move in the virtual space like a previous 3-D graphics system. Figure 10-(a) shows a generated virtual view sequence, as an example. We can find that it is sufficient for realizing the walkthrough. However, there is a gap between the enlarged view and the view which generates from the next omnidirectinal image. This problem cannot be solved by the simple switching. To decrease the gap, the last process in the first routine mixes two virtual views. Figure 10 shows the comparison of the simple switching and the mixing. By using this mixing process, we can decrease the gap as shown in the figure.

In order to realize the exploration in the virtual space by the users, we have divided all of omnidirectional images into sets of four/five neighbor images. The set represents a short route. At the end of the short route, it is connected to other routes and the users can choose a direction for the exploration. As an user interface, the second routine has a dialog-based controller. The users can choose the direction of the movement in the eight directions by using this controller. The walkthrough window is shown in Figure 11. The controller is at the upper-right in Figure 11.

The system displays a birdview map in the lower region of the system window. The last routine displays an icon as a user on the map. The system can compute the user's position from the utilized omnidirectional image and the zoom ratio. On the map, the direction of the view represents as an arrow.

Fig. 11. System window of Walkthrough

4 Virtual city models

To verify the town digitizing techiniques, we have built virtual spaces of Kyoto city, which is a historical city in Japan. As experimental environments, we have selected two areas in Kyoto, Shijou area and Arashiyama area. Shijou-town is a downtown where the traffic is always heavy. Arashi-yama area which is a scenic spot is a complex environment that includes many natural objects like trees. Figure 12 and Figure 14 show all of routes where we have taken the omnidirectional images in Shijou-town area and Arashi-yama area, respectively. Table 1 lists the number of image and the mileage in each area. An example of view sequence that is genrerated by the proposed technologies for each area is shown in Figure 13 and Figure 15, respectively.

Table 1. Number of images and milege

Area	# Images	Mileage
Shijou-town	8,611	16.21[km]
Arashi-yama	1,785	5.41[km]

Fig. 12. Routes on the map in Shijou-town area

In this experimentation, we have found the following problem. The problem is that a gap when switching images occurs in the situation where the surrounding objects are anti-symmetry along the route, for example, a sidewalk between a building and a large road. The mixing process decreases the gap, but it still remains. To solve this problem, we need to take omnidirectional images with a shorter interval. However, in other situations, we could confirmed that the mixing process shows an sufficient effect. In the case of the natural environment, it is possible to handle them. In the case of the previous 3-D graphics system, representation of the complicated surface is difficult. However, the proposed method works better since it is image-based method.

There are similar works to ours. In Movie-Map system[5], the user can walk along the virtual street based on images that have been captured before. However, the user can view the images only from the original viewpoint of the camera. Another similar system based on the image-based rendering is "QuickTime VR"[6]. In the system, users can look around a scene from fixed points. This system, however, does not allow users to walk around in the virtual space. The proposed method allows the user to view from the arbitary viewpoint of the camera and to walk along the street in the virtual space. That is, we can consider that the proposed method will be a general modeling method for the real world.

In the section 2, we have proposed a concept of cyber space. We believe the cyber space realized by the town digitizing technique and the town cameras will be a next-generation information infrastructure that connects between the computer network and the real world.

Fig. 13. Example of a virtual view sequence in Shijou-town area

References

1. R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, The Office of the Future: A Unified Approach to Image-Based Modelling and Spatially Immersive Displays, *SIGGRAPH '98*, 1998
2. H. Gouraud, Continuous shading of curved surfaces, *IEEE Trans. on Computers*, vol. 20, pp. 623-628, 1971

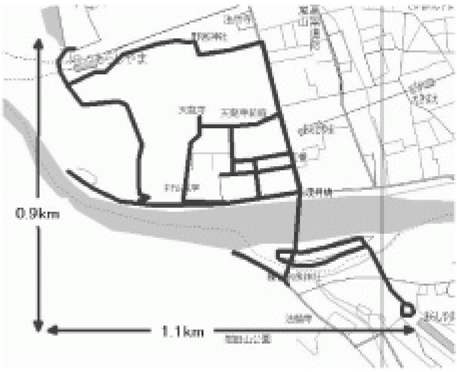


Fig. 14. Routes on the map in Arashi-yama area

3. S. B. Kang and R. Szeliski, 3-D scene data recovery using omnidirectional multibaseline stereo, *Proc. Computational Vision and Pattern Recognition Conf.*, pp. 364-370, 1996
4. T. Takahashi, H. Kawasaki, K. Ikeuchi, and M. Sakauchi, Arbitrary view position and direction rendering for large-scale scenes, *Proc. Computational Vision and Pattern Recognition Conf.*, pp. 296-303, 2000
5. A. Lippman, Movie-Maps. An application of the optical videodisc to computer graphics, *SIGGRAPH '90*, pp. 32-43, 1990
6. S. E. Chen, QuickTime VR - An image-based approach to virtual environment navigation, *SIGGRAPH '95*, pp. 29-38, 1995
7. H. Ishiguro, Development of low-cost and compact omnidirectional vision sensors and their applications, *Proc. Int. Conf. Information systems, analysis and synthesis*, pp. 433-439, 1998
8. K. Yamazawa, Y. Yagi, and M. Yachida, Omnidirectional imaging with hyperboloidal projection, *Proc. Int. Conf. Robots and Systems*, 1993
9. Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, Telepresence by real-time view-dependent image genreration from omnidirectional video strems, *Computer Vision and Image Understanding*, Vol. 71, No. 2, pp 154-165, 1998

Fig. 15. Example of a virtual view sequence in Arashi-yama area