MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

8—29

A Method for Reconstructing Structure from Omnidirectional View Sequence without Feature Matching

Daniel Moldovan *
Wakayama University

Takahiro Miyashita †
ATR

Hiroshi Ishiguro ‡
Osaka University

Abstract

This paper describes a new method for directly reconstructing the 2D structure of an environment from image sequences taken by an omnidirectional camera moving on a straight line. By exploiting the characteristics of the omnidirectional images generated in an immersive way on both sides of the camera line, we synthesize a 3-D visual representation of the environment. Slices of this volume directly encode changes due to motion of the camera. The 2D structure, parallel to the ground plane, will emerge by analyzing the 3-D visual representation, without using the feature matching process needed in multiple camera stereo.

We present the theory behind this technique, describe two initial experimentation for indoor and outdoor environments, and discuss our preliminary results.

1 Introduction

From the beginning of computer vision research, there have been many works aimed at recovering the three-dimensional information from two-dimensional images. The problem of recovering depth (the three-dimensional information) from a set of images is essentially the correspondence problem. Finding potential correspondences usually involves matching some image property in one or more images. Feature matching method proposed so far, such as template matching is not stable especially for long base lines between cameras. This paper is presenting an original stereo method that succeeded in synthesizing the 2D structure of an environment from a sequence of omnidirectional (OD) images recorded on a rectilinear path. Our method does not require matching process but it rather needs to generate many OD images (called *virtual OD images* throughout this paper).

This paper is organized as follows. Section 2 details the method for synthesizing the 2D structure from virtual OD images. Section 3 presents the results of the experiments using indoor and outdoor scenes. Section 4 concludes the paper.

1.1 Technical Overview

Similar approaches to ours are methods for approximately realizing plenoptic functions [1, 10], such as lumigraph [3] and light field rendering [11]. J.P Mellor et al. [12], analyzed an epipolar image in order to accumulate evidence about the depth at each image pixel. Though, in their approach they were focused only on the detection of the structure of the isolated objects while in our method we analyzed the environment thoroughly in an immersive way. More over, they employed a 3D arrangement of many cameras while we used a video sequence recorded when a single OD camera was moving along a straight path.

In order to retrieve 3D information from environment, Kawasaki et al. [8] proposed a hybrid method of the epipolar-plane image (EPI) analysis of OD images and the model-based analysis. One limitation of this method is that it cann't be applied in real-time applications because an apriori model of the environment is needed. In contrast with their approach, our method directly synthesize the 2D structure without any apriori information.

By being able to create any view from any position to any direction on the ground, Takahashi's work [13] is the most closely related with ours. However, their work is to reconstruct normal views with a limited visual field. In our work, we have improved this idea in order to find directly the structure of the environment from many virtual OD images.

2 Structure from Virtual OD images

2.1 Synthesize of Virtual OD images

In order to increase the number of cameras, we are using images recorded from an OD view sequence of a camera that is moving along a straight path, parallel with the ground. From each of the recorded images we extract the color information found in the circle that is coplanar with the focal point of the hyperboloidal mirror (Figure 1). The purpose of this procedure is to acquire the information related to the part of the environment located at the same height from the ground as the OD camera. Virtual OD images are generated coplanar with the original OD images, by finding and collecting one pixel at a time from each of the extracted circles. Therefore, the shape of a virtual OD image becomes an arc of an imaginary circle (Figure 2). In the process, if the corresponding rays of an original circle and an imaginary circle respectively are collinear and if they have the same length and direction (towards the same side of the robot path), the visual information corresponding to the pixels located at the extremities of the rays is the same.

*Address: Sakaedani 930, Wakayama, 640-8510, Japan. E-mail: c12s002@center.wakayama-u.ac.jp

†Address: Hikaridai, 2-2-2, Seikacho, Sorakugun, Kyoto, 619-0288, Japan. E-mail: miyasita@atr.co.jp

‡Address: Yamadaoka 2-1, Suita, Osaka, 565-0871, Japan. E-mail: ishiguro@ams.eng.osaka-u.ac.jp

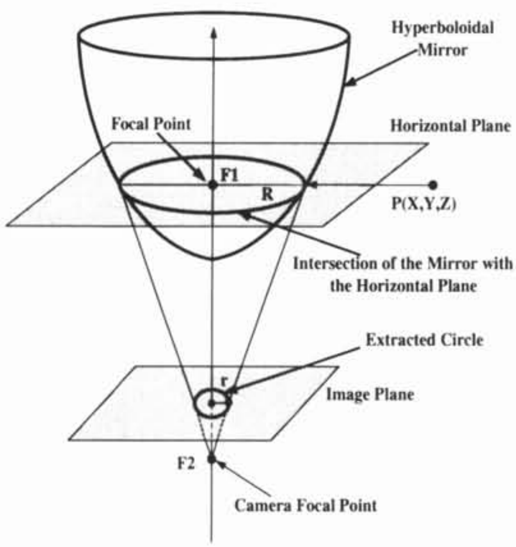

Figure 1: Extracting the circle coplanar with the focal point of the hyperboloidal mirror

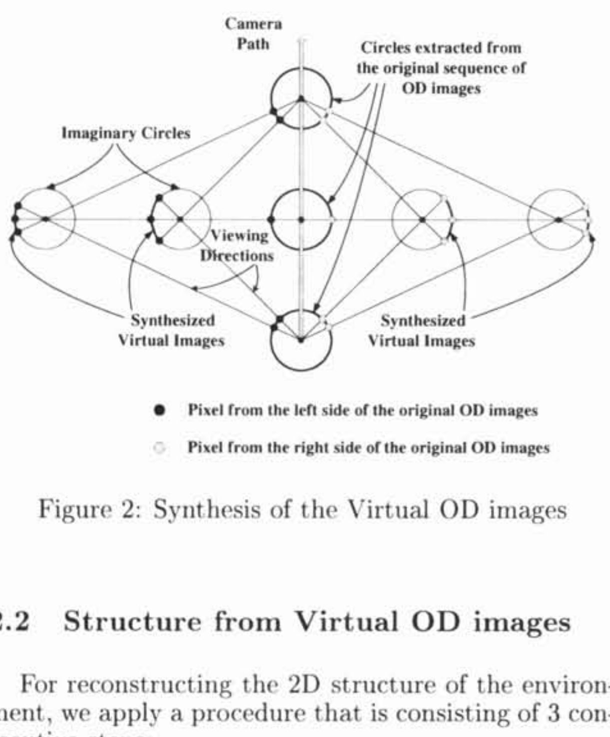

Figure 2: Synthesis of the Virtual OD images

2.2 Structure from Virtual OD images

For reconstructing the 2D structure of the environment, we apply a procedure that is consisting of 3 consecutive steps:

1. From the sequence of extracted circles we are generating virtual OD images on one side of the robot path as shown in Figure 3. Next is a process of rendering a number of virtual lines by importing one pixel at a time from each synthesized virtual OD image. Virtual lines, corresponding to a certain angle of the pixels with the line perpendicular to the camera path, are gathered into a virtual line image. For one side of the camera path we generate 180 virtual line images. For each virtual line image, the process is building a Boolean matrix that will record the data related to the presence of the color information in the newly generated image. That is, each location in the matrix will have assigned a TRUE or a FALSE value.

2. By arranging of all virtual line images one on top of the other we can build the 3D Visual Representation Volume (VRV) with I, X, Y axes representing the vertical section, distance from the camera path and the camera movement, respectively (see Figure 4). The projection of this volume on the X-Y plane will give us the 2D structure of the environment. The projection is seen as a process of detecting the common areas where pixels exhibit similar RGB values.

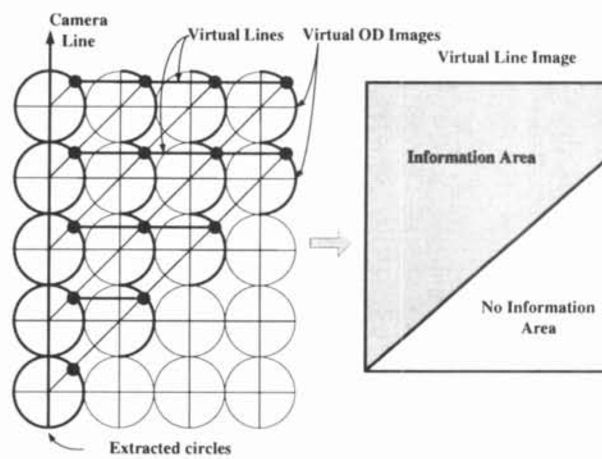

Figure 3: Synthesis of the Virtual line image

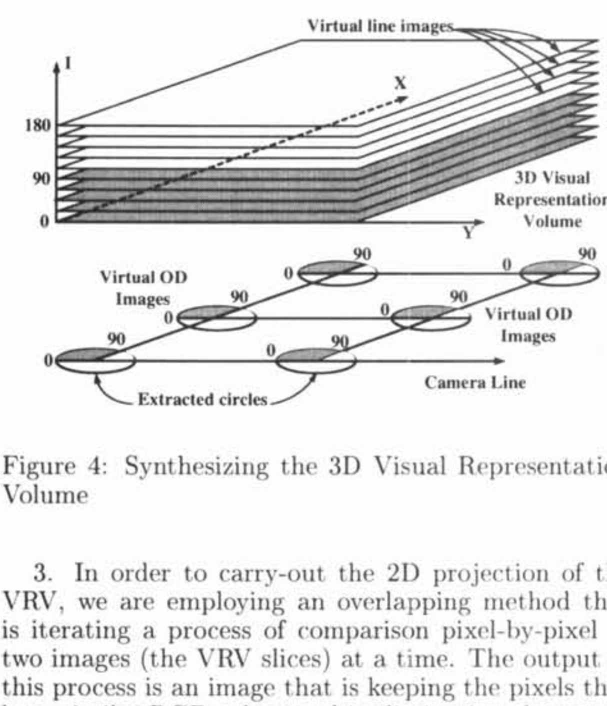

Figure 4: Synthesizing the 3D Visual Representation Volume

3. In order to carry-out the 2D projection of the VRV, we are employing an overlapping method that is iterating a process of comparison pixel-by-pixel of two images (the VRV slices) at a time. The output of this process is an image that is keeping the pixels that have similar RGB values and is eliminating the rest of them. After getting the result of comparison between two initial images, the overlapping process continues with the comparison of the result with the next slice from the VRV. The process ends when there is no more slice to compare with.

3 Experimentation results

3.1 Indoor environment

The indoor experiment has been done using a static environment from our laboratory (Figure 5). Original

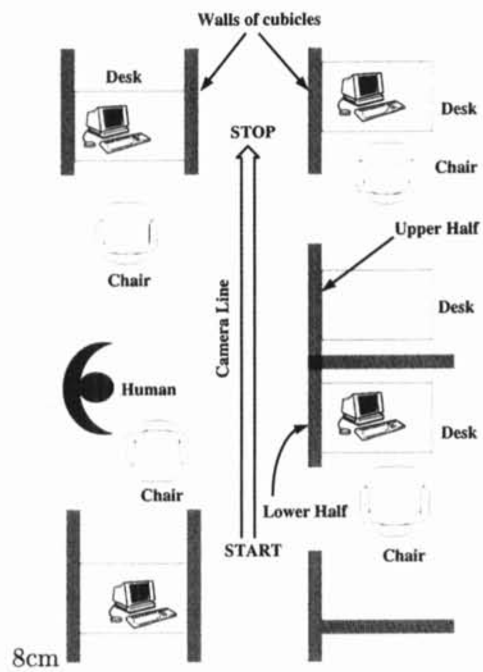

Figure 5: Indoor environment

images were recorded with a frequency of one image / 0.30 cm. The resulted structure on the left side of the path is shown in Figure 6. The meanings of the objects encircled and numbered is: (1) the corner of the wall; (2) the chair; (3) the standing person and the chair; (4) the corner of the wall. By following the same procedure we have determined the structure on the right side of the path (Figure 7). The meanings of the encircled areas are as follows: (1) represents half of the upper wall; (2) represents the lower part; and (3) is the edge of the desk. The two distinct walls that were detected

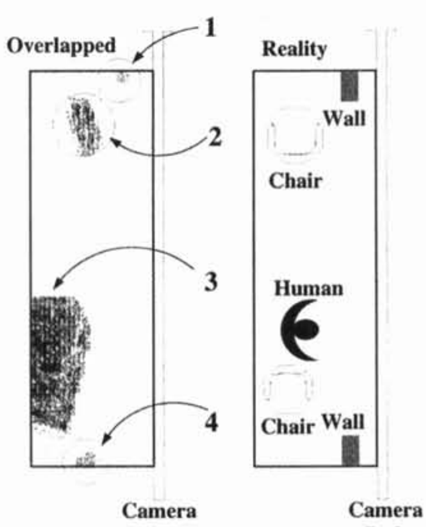

Figure 6: Left Side Structure

correspond in fact to the reality (in the original images they are separated by an edge of a different color). We have to mention that even with a low accuracy, the results proved the right location of the objects in the surrounding environment.

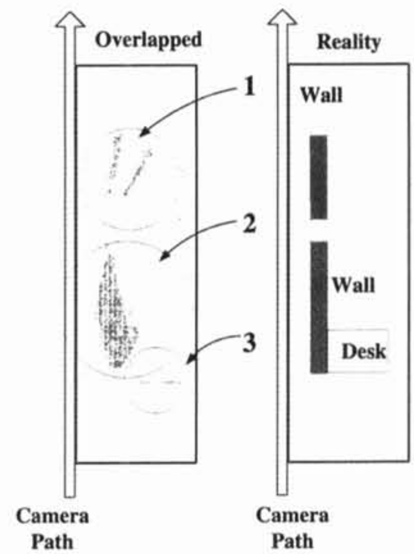

Figure 7: Right Side Structure

3.2 Outdoor environment

Next are the results for an outdoor scene represented by six poles that are located in our university campus. In order to eliminate the cases of occluded views we took multiple paths around the poles. These paths are obtained by using the T-Net, one of our previous work [4] that allows careful approximation of local areas with straight lines and a precise control of the camera motion while actively tracking a pair of feature points. The camera (mounted on top of a electrical wheelchair) ran between intermediate feature points with a constant speed while recording images with a frequency of one image / 1.5cm (Figure 8).

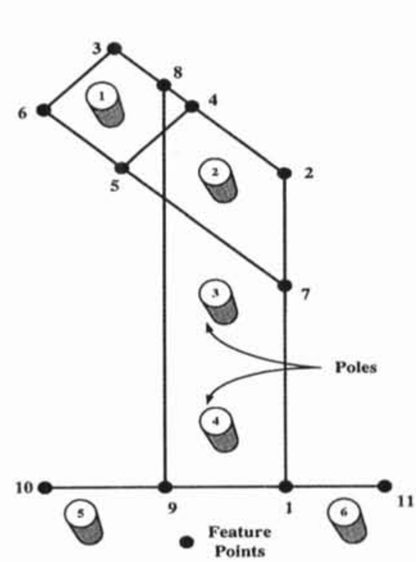

Figure 8: Outdoor environment

The 2D global map of the environment (Figure 9) was built by combining intermediate 2D maps resulted from analyzing the OD images along the following paths (see Figure 8): 1-2, 2-3, 4-5, 6-7, 8-9, 10-11. These intermediate paths correspond to skeletons of the local areas. Because poles 5 and 6 were analyzed based on a single camera path, their corresponding 2D structure does not have the same solid consistency as the rest of the detected poles.

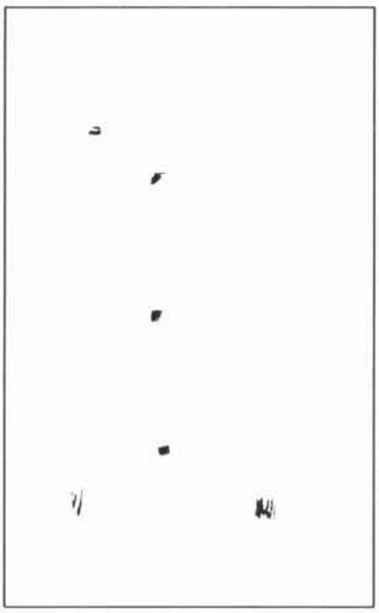

Figure 9: 2D Structure for the outdoor environment

As proved by the results, one application for our method might be the recovery of the environment's coarse structure. Comparing with the indoor environments, where the flat floor constraint is satisfied for the entire camera path, the results for outdoor scenes are more liable of being affected by errors due to camera's tilt variations. However, in the case of structures of big sizes (like buildings or cars) located in the near vicinity of the camera's path, small variations of camera's tilt are not significant for the overall result.

4 Conclusion

This paper described a new method for reconstructing the 2D structure of the surrounding environment from image sequences taken by an OD camera. We generate virtual images on both sides of the path and exploit the changes that appear in each virtual image that encounter with an object. The originality brought by this method is that is uses only the rich visual information in order to solve the correspondence problem. The strength of our method is that it requires a single omni-camera and the processing is done in real time being well suited for real-world applications. A weak point is represented by the low accuracy in detecting the objects shape. Future work will focus on overcoming this limitation and on extending our method to 3D structure.

References

[1] E.H. Adelson and E.H. Bergen, "The plenoptic function and the elements of early vision", MIT Press, 1991.

[2] Robert C. Bolles and H. Harlyn Baker, "Epipolar-Plane Image Analysis: A Technique for Analyzing Motion Sequences", pp. 26-36 Readings in Computer Vision, Ed by Martin A Fishler and Oscar Firschein.

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph", Proc. SIGGRAPH, pp. 43-54, 1996.

[4] Hiroshi Ishiguro, Takahiro Miyashita and Saburo Tsuji, "T-Net for navigating a vision-guided robot in a real world", IEEE International Conference on Robotics and Automation, 1995.

[5] Hiroshi Ishiguro, Takeshi Maeda, Takahiro Miyashita, and Saburo Tsuji, "A strategy for acquiring an environmental model with panoramic sensing by a mobile robot", IEEE International Conference on Robotics and Automation, pp. 724-729, 1994.

[6] Hiroshi Ishiguro, Kenji Ueda and Saburo Tsuji, "Omnidirectional visual information for navigating a mobile robot", IEEE International Conference On Robotics and Automation, pp. 799-804, 1993.

[7] Hiroshi Ishiguro, Masashi Yamamoto, and Saburo Tsuji, "Omni-directional stereo", IEEE Trans. Pattern Anal. Machine Intell., 14(2):257-262, February 1992.

[8] Hiroshi Kawasaki, Katsushi Ikeuchi, and Masao Sakauchi, "Spatio-Temporal Analysis of Omni Image", Computer Vision and Pattern Recognition, Vol. 2, pp. 577-584 (June 2000).

[9] Hiroshi Kawasaki, Katsushi Ikeuchi, Masao Sakauchi, "Light Field Rendering for Large-Scale Scenes", Computer Vision and Pattern Recognition, Vol. 2, pp. 64-71, Hawaii, USA (Dec. 2001).

[10] M. Landy and J.A. Movshon eds, "Computation models of visual processing", MIT Press, 1991.

[11] M. Levoy and P. Hanrahan, "Light field rendering", Proc. SIGGRAPH, pp. 31-42, 1996.

[12] J.P. Mellor et al. "Dense Depth Maps from Epipolar Images", M.I.T Artificial Intelligence Laboratory, Nov. 1996.

[13] Takuji Takahashi, Hiroshi Kawasaki, Katsushi Ikeuchi, and Masao Sakauchi, "Arbitrary View Position and Direction Rendering for Large-Scale Scenes", Computer Vision and Pattern Recognition, pp. 296-303, (June 2000).