Multiple Omni-Directional Vision Sensors (ODVS) based Visual Modeling Approach

Kim C. Ng*, Hiroshi Ishiguro**, Mohan M. Trivedi*

*Department of Electrical and Computer Engineering, University of California, San Diego, USA
**Department of Social Informatics, Kyoto University, Japan

Abstract

In this paper we present a multiple camera stereo method, namely *Smooth Interpolation*, for visual modeling of environment using omnidirectional vision sensors. The goal is to approximate the *plenoptic function* of a virtual walkthrough as closely as possible. Our technique can approximate the plenoptic function very closely with *spatio-temporal contents preserved*.

Keywords: multiple camera stereo, wide baseline stereo, range-space search, omni-directional vision sensors, ubiquitous vision, visual modeling, image-based rendering

1. INTRODUCTION

Visual modeling deals with development of a computer-based representation of the 3-D world at any desired vantage point. Automatic means for deriving such models have great demand and potential. However, developing such an efficient algorithm can be difficult, especially if the scene is large and dynamic objects prevail. Additional difficulties are introduced if concurrent, multiple different views are allowed. In this paper we discuss a set of methods to address such needs.

2. RELATED RESEARCH

There are two basic types of visual modeling methods utilizing multiple cameras/images that commonly exist in current literatures. One is an extension of multiple camera stereo developed in the computer vision community and the other method approximates the *plenoptic function* [1] with densely sampled raw images in computer graphics domain. The plenoptic function has been proposed as an ideal function representing complete visual information in a 3-D space.

Multi-camera stereo was applied to a virtual reality system called *virtual dome* [2]. Precisely calibrated 56 cameras observe targets inwards from the surrounding and reconstruct 3-D models in a small constraint area. [3] developed a 3-D modeling system called *multiple perspective interactive video*. Their multiple baseline stereo performs template matching on the image plane and finds corresponding pixels based on a volume occupancy array which covers the target area. Recently, [4] have improved their method and applied it to precise 3-D modeling for small objects with densely arranged 16 cameras.

* Email: kimng@ece.ucsd.edu; trivedi@ece.ucsd.edu
** Email: ishiguro@i.kyoto-u.ac.jp

The visual modeling with many densely sampled images does not require matching among them. However a very large number of cameras/images are needed. Methods, such as *lumigraph* [5] and *light-field rendering* [6] approximate the 5-D plenoptic function with a 4-D function. These methods are impractical for applications that cover large area, especially where dynamic objects dominate. Ideas of Boyd and Seitz compensate for the problem to some extent, however they are still inefficient for covering a wide area. Narayanan's approach using multiple baseline stereo is one of the practical solutions. However, their purpose is still to reconstruct *complete* 3-D models of small 3-D spaces, and then re-project pixels to a novel view. The method is both computationally and memory intensive. For applications that require wide scene coverage, more efficient methods are required.

3. OUR APPROACH

The approach developed in this paper is the hybrid of the two discussed methods based on unique properties of *omnidirectional images* (ODIs) generated by omni-directional vision sensors (ODVS). Our method does not extract 3D for the whole scene; as a matter of fact, we extract only a *single* 3D point lying along the center viewing direction of our desired virtual view plane. The virtual view plane at that 3D point is used to affine transform our selected raw image into the novel view. The selected raw image is chosen based on its correlation of distance and viewing direction to our desired virtual view.

In short, our solution to this problem is to directly acquire the plenoptic representation by using ODVS, which each sensor has a wide field of view. And interpolate between them (at where ODVS do not exist) with 3-D geometrical constraints given by our modified multiple camera stereo.

4. ODVS and ODI

For video acquisition, we use ODVS, which consist of a hyperbolic mirror mounted above the lens of a rectilinear camera. This configuration provides a 360-degree field of view [7]. **Figure 1** shows a compact ODVS. The height is about 3.5 inch including a camera unit that provides a NTSC video signal. **Figure 2** shows an ODI taken by a ODVS in an indoor environment.

Figure 1: ODVS

Figure 2: ODI

5. EXPERIMENTAL ENVIRONMENT

Figure 3 shows a hallway, which is our experimental environment. The hallway was chosen because it has the unusual "T" layout and its colors are very similar. This type of environment is usually difficult for traditional stereo to perform; however, our method has shown excellent results in generating high quality virtual views. The interval between the sensors is 2 feet. Our final goal is to synthesize a seamless image sequence with spatio-temporal context preserved along the indicated desired path from the sixteen ODIs.

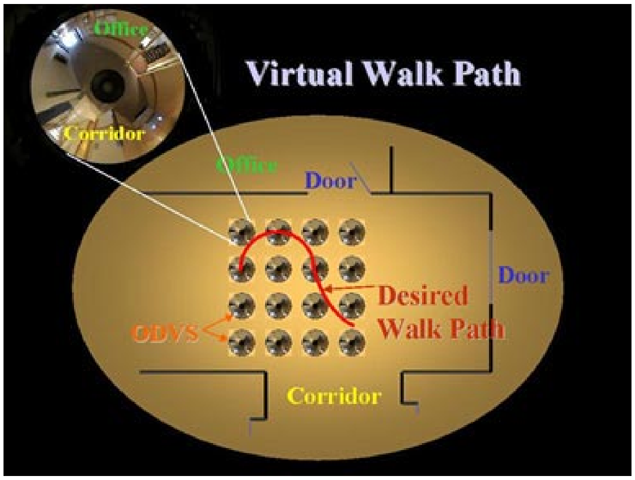

Figure 3: The experimental environment with the positions of ODI indicated and the desired walk path shown.

6. VIRTUAL PATH FOR VISUAL MODELING

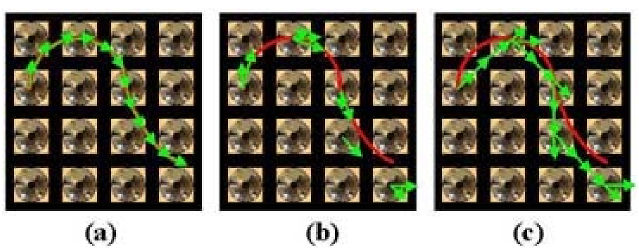

Figure 4: (a) Desired, smooth walk path for view synthesis. (b) Discrete paths with camera switching to approximate the desired path. (c) Discrete interpolation between camera pair with zooming stereo.

Figure 4 (a) shows the discrete paths (with arrowheads) which our method is used to approximate the smooth walk path.

Figure 4 (b) is one of the common image-based methods that approximate the desired path by jumping from one camera center to the next along the path (just like Apple's QuickTime VR). Multiple perspective views can be generated from a single ODI when appropriate 3D transformations are applied. The closet camera to the path is selected for view generation, and its viewing direction is aligned with the path motion. However, this type of virtual walkthrough easily loses the spatio-temporal continuity when jumping off one center to the next, unless the sensors /images are in very close range. The loss of spatio-temporal context causes visual confusion to the observer, and disorientation in the environment results.

Figure 4 (c) shows another possible approximation of the walkthrough by linearly interpolating the views between two cameras. The cameras are again selected based on their proximity and direction to the desired path. This discrete interpolation "smoothes out" the discontinuity seen in the camera switching method (**Figure 4** [b]). However, the effect is like "zooming" linearly to connect two cameras. In other words, no arbitrary walkthrough is possible.

For arbitrary, smooth walkthrough, we introduce the method---*Smooth Interpolation* by *multiple camera stereo*.

7. SMOOTH INTERPOLATION--- MULTIPLE CAMERA STEREO

An ideal interpolation is to synthesize images at arbitrary viewpoints without any constraint. Such smooth interpolation between ODIs needs to estimate the local environment structure and select the best ODVSs for the virtual view synthesis. We have employed a multiple camera stereo method and modified it for realizing robust range estimation.

The basic idea is similar to multiple baseline stereo [8] by exploiting the advantages of having redundant information, but we have modified it for general camera arrangements and arbitrary search path. Originality of our work is listed below:

- *Range-space search* method for finding corresponding points that accepts arbitrary camera positions and arbitrary search path.
- Camera selection based on robust statistics.
- Utilization of large templates for robust template matching.

Generally, cameras used for multiple camera stereo are arranged in a line; and the search for finding corresponding feature points is performed in the disparity/image space based on the image plane structure. However, if we suppose a general camera arrangement, we have difficulty in performing the search in the disparity space, especially if the cameras are not closely positioned and their field of view is large (such as of the ODI). In range-space search, positions of templates for finding corresponding points on camera images are given as projections of 3-D points along a virtual 3-D search line onto the camera images. **Figure 5** shows the range-space search performed between multiple ODVSs. Here, the search intervals

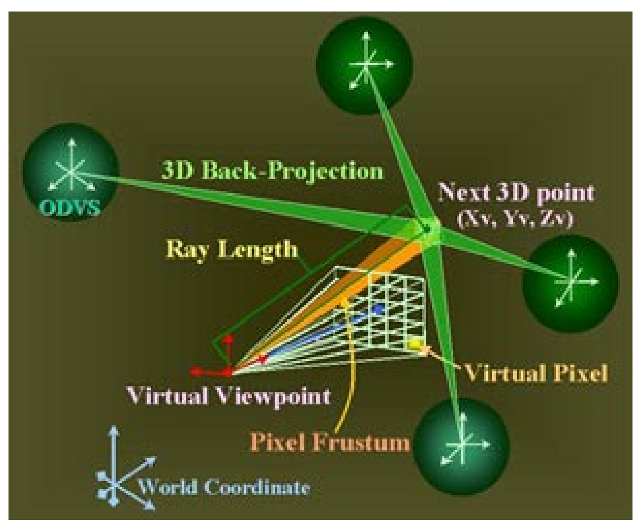

Figure 5: Multiple camera stereo with range-space search.

along the virtual 3-D line are determined based on pixel resolution of the nearest ODVS so as to preserve the maximum range resolution. Note that pixels in the ODI cover the 3-D space with different resolutions. This range-space search accommodates the window distortion, scaling, and foreshortening effects that larger baseline stereo inherits, which the simple rectangular template in the disparity space cannot trivially solve.

Based on the coarse range data, we have selected the best camera that simulates the virtual view at a virtual observation point. Here, the template used for the template matching is directly referred to as the virtual view. **Figure 6** shows the plenoptic representation based on the extracted coarse structure of the environment.

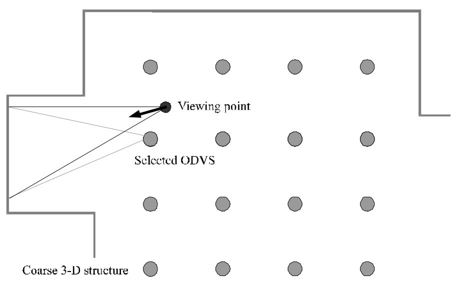

Figure 6: Plenoptic representation based on coarse structure of the environment.

For comparison we have generated the walkthrough image sequence using a standard digital camera in the real, experimental environment (**Figure 7)**. In this experimentation, the ODVS has CCD resolution of 410,000 pixels and an image capture board that can take an image of 640x480 pixels. The synthesized color images each has 179x126 pixels. **Figure 8** shows the virtual image sequence synthesized along the smooth path. A comparison of these image sequences show close resemblance between the "real" and "virtual" scenes.

8. CONCLUDING REMARKS

In summary, we are developing a promising method to provide arbitrary virtual viewing that is suitable in dynamic indoor as well as outdoor environments. As the results have shown, we can simulate the effect of seamless virtual walkthrough that closely approximates the desired walk path. This research will lead to the development of new generation of ubiquitous vision systems [9] for the applications such as virtual city walkthrough, remote surveillance [10], video conferencing system, and human-computer interaction.

Acknowledgements

We would like to thank the sponsors of this project: State of California, Sony Electronics Corporation, and Compaq Computer Corporation, through Digital Media Innovative (DiMI) program. Also, special thanks to Rick Capella, Sadahiro Iwamoto, Nils Lassiter, and other colleagues of CVRR lab.

References

[1] E. H. Adelson and E. H. Bergen. The Plenoptic Function and the Elements of Early Vision. Computation models of visual processing (M. Landy and J.A. Movshon eds.), MIT Press, 1991.

[2] P. J. Narayanan, P. W. Rander, and T. Kanade. Constructing Virtual Worlds using Dense Stereo. Proc. ICCV, Bombay, India, pp. 3-10, 1998.

[3] J. Boyd, E. Hunter, P. Kelly, L. Tai, C. Phillips, and R. Jain. MPI-Video Infrastructure for Dynamic Environments. Proc. Int. Conf. Multimedia Systems, 1998.

[4] S. M. Seitz and K. N. Kutulakos. Plenoptic Image Editing. Proc. ICCV, pp. 17-24, 1998.

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

[6] M. Levoy and P. Hanrahan. Light Field Rendering. Proc. SIGGRAPH, pp. 31-42, 1996.

[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] M. Okutomi and T. Kanade. A Multiple Baseline Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp.353—363, April 1993.

[9] K. C. Ng, M. M. Trivedi, and H. Ishiguro. 3D Ranging and Virtual View Generation using Omni-view Cameras. Proc. Multimedia Systems and Applications, SPIE Vol 3528, Boston, November 1998.

[10] K. C. Ng, H. Ishiguro, M. M. Trivedi, and T. Sogo. Monitoring Dynamically Changing Environments by Ubiquitous Vision System. IEEE Workshop on Visual Surveillance, Fort Collins, Colorado, June 1999.

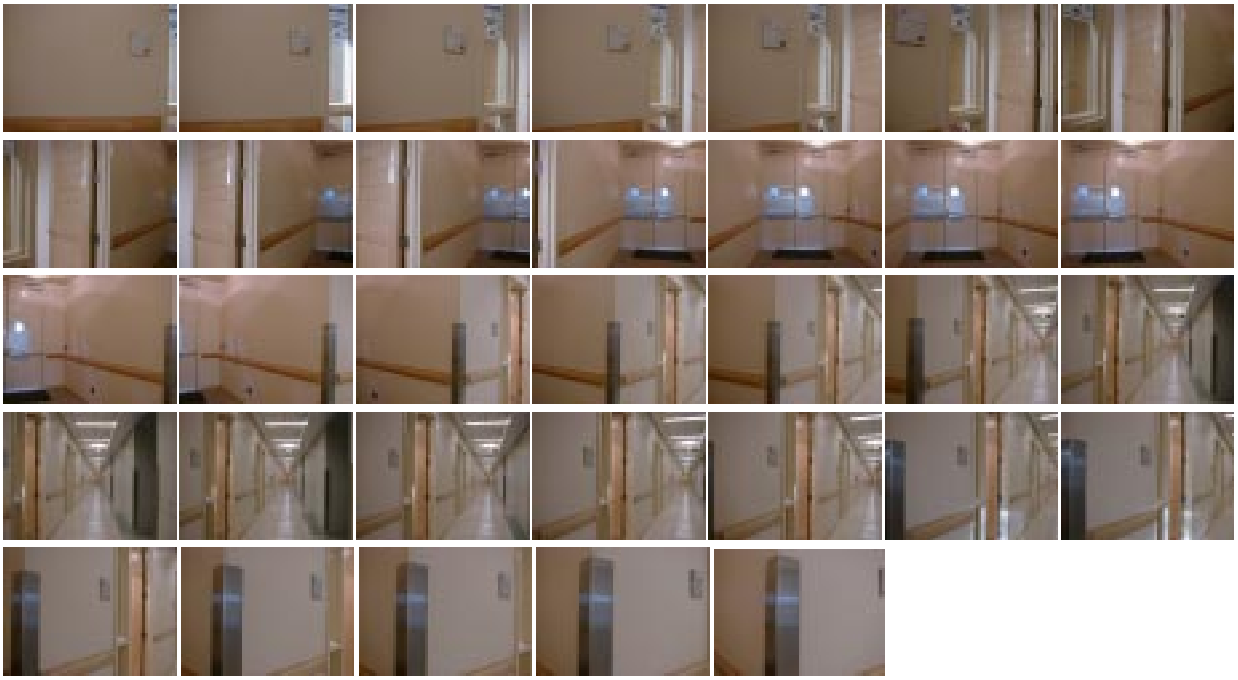

Figure 7: Image sequence taken by a standard camera along the desired virtual walk path.

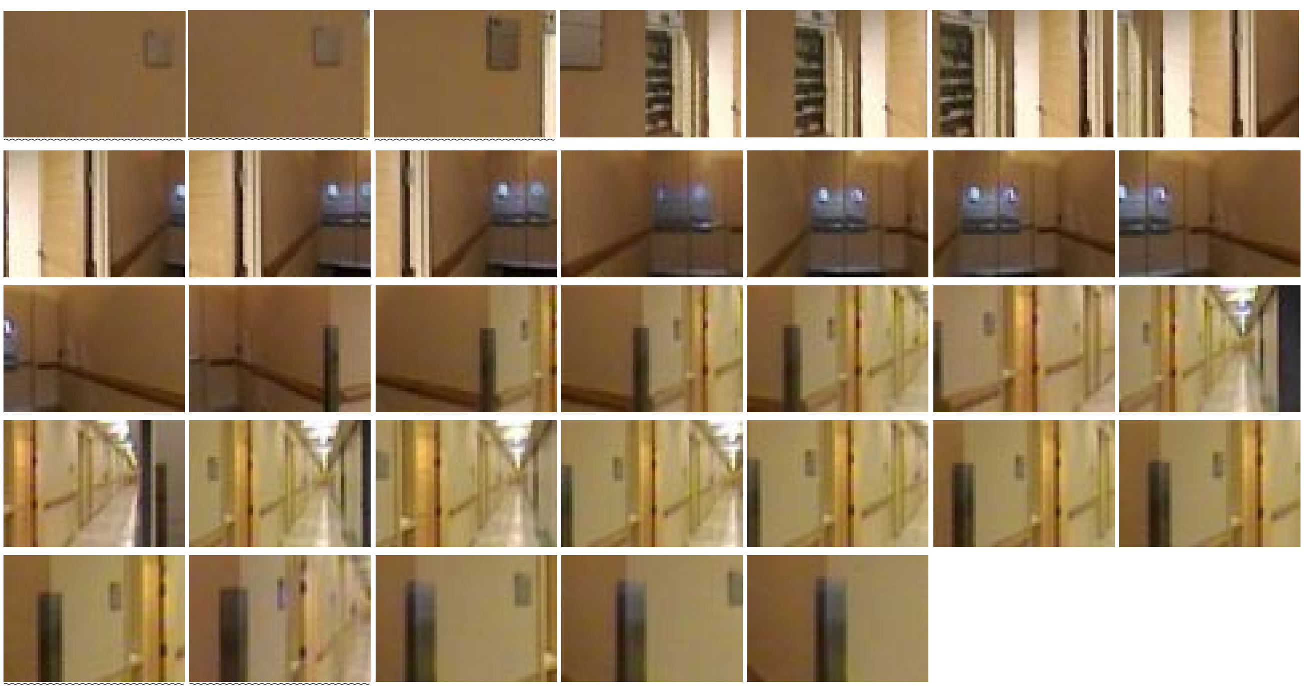

Figure 8: Synthesized image sequence by the *Smooth Interpolation* method using multiple omnidirectional images.