An image-based scene representation and rendering framework incorporating multiple representations

Ingo Bauermann and Eckehard Steinbach

Institute of Communication Networks
Media Technology Group
Technische Universität München
Email: {bauermann,steinbach}@ei.tum.de

Abstract

A variety of image-based scene representations like light fields, concentric mosaics, panoramas, and omnidirectional video have been proposed in the past years. These image-based scene representations provide photorealistic interactive user navigation in a 3D scene. As the trade-off between acquisition complexity, freedom of movement and rendering quality differs for the diverse techniques, the most efficient scene representation and rendering technique should be selected with respect to scene content and complexity.

Splitting a scene into partial representations which are adapted to local requirements is proposed in this paper. Besides meaningful restrictions to user movement, the transition between different image-based scene representations is addressed to provide an efficient image based walkthrough for large and complex scenes. We identify rendering parameters to achieve a seamless transition between different representations and present results for stitching together concentric mosaics, omnidirectional video and light fields.

Keywords: Image-based rendering, concentric mosaic, omnidirectional video, interactive walkthrough.

1 Introduction

Interactive navigation through virtual environments is a popular application on the Internet. Interesting objects, places, buildings or apartments can be explored in photorealistic quality over wired and even wireless channels. Worldwide there are more than 100 million browsers in use which are capable of rendering interactive environments.

Traditionally an interactive walkthrough application renders the desired views of a scene using detailed 3D models. The 3D scene is constructed by defining geometry and texture properties for all objects in the scene. For large and complex scenes this modeling technique is tedious and time consuming. The automatic generation of 3D models of real scenes is an active area of research. Many real world objects like clouds, fur, hair, etc. either require a tremendous modeling complexity or can not be modeled realistically at all.

To simplify the modeling process and to make rendering complexity independent of scene content, image-based scene representations have been proposed that capture a real scene with one or more calibrated digital cameras and store the scene as a set of intensity samples. In its most general form image-based rendering techniques are based on the plenoptic function [1], a 7D function describing a 'set of rays passing through any point in space':

$$P(x, y, z, \phi, \varphi, \lambda, t)$$

Given the parameterization of this plenoptic function for a specific scene, the intensity value for every light ray is registered by its viewpoint, direction, wavelength and time. Generating views is just a matter of composing appropriate intensity values [4]. Most image-based rendering techniques assume a static scene and discrete color values (RGB) which leads to the removal of two degrees of freedom and results into a 5D plenoptic function. A reduction of the number of degrees of freedom has an impact on acquisition complexity, virtual movement constraints, memory consumption, rendering quality, and rendering complexity.

As an example, a locally constrained movement of the virtual camera during interactive scene navigation allows us to pick an image-based scene representation with lower acquisition complexity and less memory usage. Looking at a natural scene it turns out that the degrees of freedom of interactive user movement can be constrained without affecting the perception of immersive navigation through a dedicated area if local scene content and complexity are taken into account. While navigating through a narrow hallway or up and down a staircase translational user movement orthogonal to the main direction might not be required. In such cases the complexity of the corresponding representation can be reduced significantly. Figure 1 shows the map of an artificial scene with local movement requirements. Translational navigation is only required within the black circles and along the path connecting the two rooms. All viewpoints should support a panoramic view.

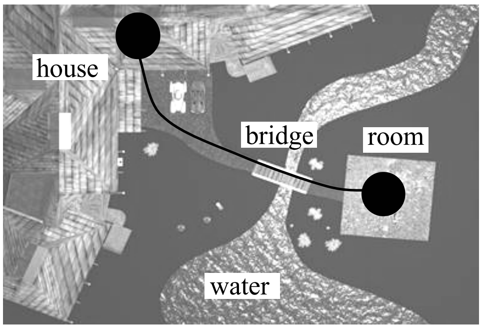


Figure 1: Map of a scene that is modeled using two image-based representations (black areas) with different degrees of freedom for user navigation.

An efficient way of representing this scene would be to use a concentric mosaic within each of the rooms and an omnidirectional video along the path from one room to the other. An important question arising in this context is how to transition from one representation to another.

To achieve a smooth transition between image-based representations many parameters have to be considered. Besides different extrinsic and intrinsic camera parameters at the stitching point, distortion due to representation dependent approximations and simplifications has to be compensated for.

In this paper we propose an image-based scene representation and rendering framework that incorporates multiple representations which are adapted to local scene content regarding acquisition complexity, freedom of user movement and memory consumption. The transition between the representations is smoothed by adapting camera parameters near the stitching point and a simple morphing technique is used to reduce the remaining distortion.

The remainder of this paper is structured as follows. Section two shortly describes related work on modeling and combining scene representations. In the third section the representations which are used in this work are described. An overview over the parameters used for registering scene representations is given in the fourth part. Results for a transition between a concentric mosaic [14] and omnidirectional video [13] along with jointly rendering of concentric mosaics and a light field are presented in the fifth chapter. This work closes with a conclusion on stitching image-based representations and notes on future work.

2 Related Work

A first approach using multiple representations for interactive navigation was introduced by Lippman. The Movie Map [3] uses many images and image sequences taken from different viewpoints as the scene representation. The user can switch between different viewpoints to choose a path to navigate on. Switching between panoramic images taken from discrete viewpoints is proposed in [2]. Smooth translational movement is not supported.

Defining a 5D parameterization of the plenoptic function McMillan and Bishop [4] use only two cylindrical projections of a scene. An efficient image warping technique is proposed to provide a smooth transition between the two panoramas and to compose nearby views. Other examples using view interpolation and image warping techniques [5, 10] can be found in [5] and [9]. To extend the local movement constraints 3D reconstruction algorithms [4, 5, 6, 9] are used to approximate scene structure. The rendering complexity, however, is very high due to the difficult correspondence problem.

The Light field [7] and Lumigraph [8] representations use a four dimensional parameterization of the plenoptic function consisting of a large database of light rays. The movement is constrained to a box either for the scene or the user. The stitching problem has been addressed by Hanrahan and Levoy [11].

Plenoptic stitching is described in [12]. Multiple omnidirectional videos [13] are captured on several paths within a scene and free user navigation with smooth transitions between different paths is achieved. Acquisition complexity is high as a motorized cart and active camera calibration are used. The transition between several videos yields very good results.

3 Scene representation

The image-based representations used in this work are described in this section. Properties regarding complexity and rendering quality are mentioned. A review of image-based scene representations is given in [15].

Omnidirectional Video

Omnidirectional video enables surround viewing of a scene with the navigation constrained on a fixed path. A mirror system or fisheye lens provides a 360° horizontal field of view. Views are easily generated by warping the panoramic image into the correct perspective according to a given viewing direction. The acquisition of omnidirectional video is easy as a standard digital video camera can be used.

Compared to other image based acquisition methods the spatial resolution is not constant due to the acquisition geometry and low as a full panorama has to be captured with one shot. Considering walking over a narrow path or through a hallway this representation provides the best trade-off between memory consumption and freedom of user movement.

Concentric Mosaics

Concentric Mosaics [14] capture a 3D plenoptic function by constraining camera motion to a circle. An outward looking standard camera mounted on a camera crane is used to provide a free translational and rotational user movement within a planar circle. Novel views are easily generated interpolating columns of reference images assuming a constant depth.

Generally the spatial resolution is high while the amount of data is large (e.g.: 2GB of raw data using 1800 high resolution reference images). Due to the need of random access to many reference images to render one view compression is difficult. By using manifold hopping [16] the memory consumption can be reduced with the cost of higher rendering complexity.

The main problem with concentric mosaics is that parallax is only rendered correctly in the horizontal direction. Vertical distortion is always present while navigating. A smooth transition to other representations even to nearby captured concentric mosaics is not possible without depth information.

Navigating over places and in rooms this representation provides full planar user navigation.

Light Fields

The Light field used in this work consists of only a planar circular scan of an artificial object with blue background. Capturing this kind of light field can be done using an inward looking camera mounted on a rotating camera crane. The blue channel is used as the alpha channel to blend directly into a novel view. Normally light fields provide correct vertically and horizontally parallax. Using a planar scanned light field only horizontal parallax can be rendered correctly.

For small objects and all inward looking scenarios this representations provides full planar navigation if the distance to the object is large. However, the rendering complexity and memory consumption are high.

4 Registration

In this section a scene organization tool is presented. Parameters are identified for stitching together different image-based representations. Concepts for parameter adaptation near the stitching point are discussed.

4.1 Scene organization

To provide simple acquisition without elaborate and costly automatic calibration and registration we assume that partial image representations are captured using a single standard camera and a camera crane. Rendering parameters are obtained using a software tool (see Figure 2) which generates a scene organization object containing all relevant positioning and adaptation data. This scene organization object references partial representations. The renderer is part of the representation and is adapted to the specific plenoptic sampling structure.

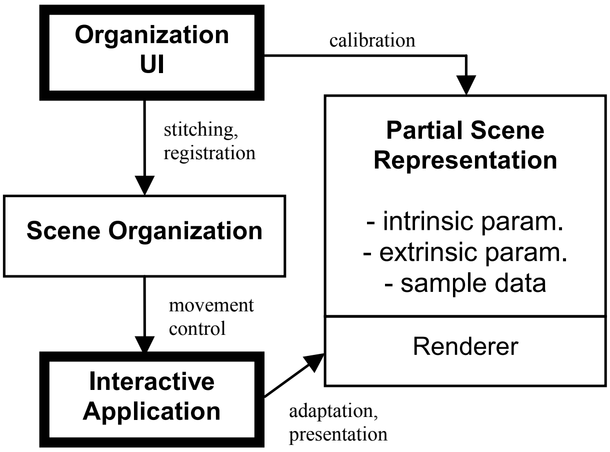


Figure 2: Structure of the scene organization tool. Calibration is done using a user interface which also generates the scene organization object. An interactive application can access the data of the scene organization and invokes the renderer of a partial scene representation to generate interactive walkthroughs.

4.2 Parameters

The common parameters that have to be registered for every representation and adapted during the stitching process are identified in the following section.

Extrinsic Parameters

The absolute position (x, y, z) of the representation in common world coordinates has to be determined. The navigation area can either be a point (single panorama), path (omnidirectional video), a planar area (concentric mosaic), or 3D space (light field). The parameterization of the navigation area uses geometric primitives in this work. At any point within the area, rotational and translational movement constraints are specified.

Intrinsic parameters

The horizontal and vertical field of view (θ, ϑ) (focal length and geometry of the image plane) of the virtual camera may be also constrained. For walkthrough applications a limitation of the vertical field of view is reasonable. The image resolution

$$\psi = \frac{\text{reference samples}}{\text{output pixel}}$$

can change significantly depending on the current representation, user position and other intrinsic parameters. This may lead to movement or zooming constraints. Distortion due to insufficient camera calibration like radial distortion or pixel skew is not compensated for in our work at this time.

Additional Parameters

Lightning has to be considered while rendering a representation into the view of another. In the case of simultaneously rendering a light field and a concentric mosaic this becomes obvious.

Non static scenes require registration in time. Though this case is not considered in this work, discrete changes, e.g., in the lightning by interactively switching off the light in a room, can be associated with interactive virtual buttons or dynamic events. Other examples would be additional representations for different seasons, weather, etc.

For concentric mosaics the constant depth assumption has to be determined. This parameter has an impact on the visibility of vertical distortion and aliasing artifacts.

4.3 Manual registration

A scene organization tool provides the possibility to manually place scene representations into a 2D map representing the navigation plane. Figure 3 shows two windows of the user interface.

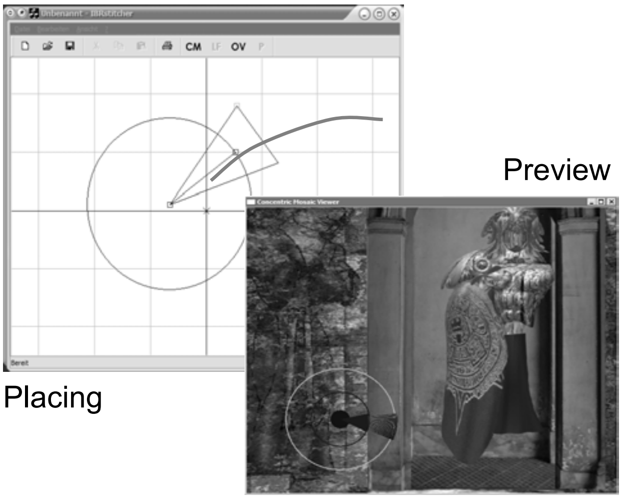


Figure 3: User interface of the scene organization tool. Placing window and preview window.

Transition from concentric mosaics to omnidirectional video

For concentric mosaics the freedom of movement is constrained to a circle with radius r [14] and can automatically be calculated given the radius R of the camera path during acquisition and the field of view (FOV_{ac}) of the capture device:

$$r = \sin\left(\frac{FOV_{ac}}{2}\right) \cdot R$$

Outside the circle with radius *r* horizontal rotation is limited with respect to the distance of the viewpoint (u,v) to the middle of the concentric mosaic (x_C, y_C) and the field of view of the virtual camera $FOV_{virtual}$ as shown in Figure 4:

$$\phi < 2 \cdot \operatorname{asin}\left(\frac{r}{\sqrt{(u-x_C)^2+(v-y_C)^2}}\right) - FOV_{virtual}$$

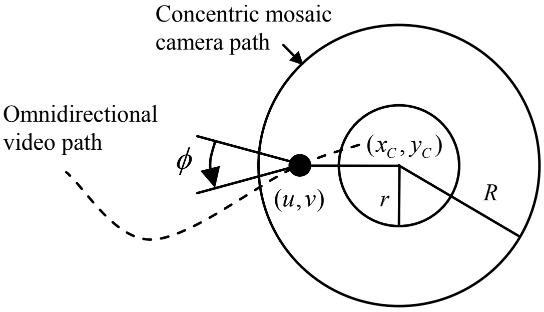


Figure 4: Navigation area and horizontal rotation constraint ϕ to user navigation at position (u,v) in concentric mosaics.

As vertical parallax can only be rendered correctly with the viewpoint located on the camera path the stitching point to other representations should be chosen here. The rotational freedom at these points is $\phi = 0$ for $FOV_{ac} = FOV_{virtual}$. Only one reference image is available.

For omnidirectional video, parallax is rendered correctly and a large field of view can be reconstructed at any place along the captured path.

The stitching point selected in this work is a point with an outward looking virtual camera positioned at (u,v) between the two circles with radius *r* and *R* as shown in Figure 4. The field of view can be adapted controlled by an optional window in the scene organization tool showing the difference image between both views and should be as large as possible without introducing too much distortion. For vertical displacement during capturing, the elevation of the anchor point of the representations can be manually adapted.

Concentric mosaics and light fields

To insert objects into the image-based scene like products and furniture an alpha blending technique is used. An inward looking concentric mosaic is used as a simple light field representation. Blue is used as the background color during capturing and as alpha channel.

Correctly rendering concentric mosaics and light fields jointly is a challenging task due to the horizontal only rendering of parallax and the constant depth assumption using concentric mosaics. Also occlusion has to be considered. The light field used in this work does not provide vertical parallax and therefore has to be placed far from possible viewpoints. Figure 5 shows the vertical distortion problem while stitching together concentric mosaics and light fields.

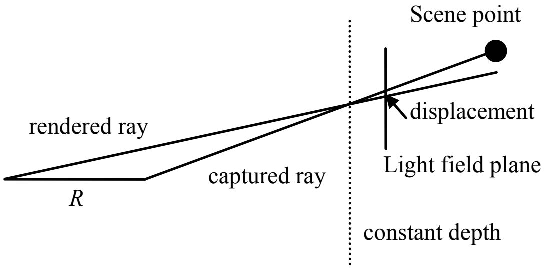


Figure 5: The captured ray seen from the camera path is rendered with a displacement on the image

plane and light field plane due to inaccurate depth assumptions.

Rendering a light field into a concentric mosaic can only be performed correctly with depth information.

4.4 Smoothed transition between different representations

At the stitching point between concentric mosaics and omnidirectional video the views generated generally show distortion. Mainly parallax, the different resolution, and inaccurate manual adaptation cause differences between the same views generated from different representations at this position. To smooth the transition a simple morphing algorithm is used in this work.

The images are partitioned into non-overlapping blocks of 8x8 pixels and block matching is performed between the two images. The resulting displacements are interpreted as motion vectors and are used to predict the concentric mosaic view from the video and vice versa. Using a sequence of 8 frames when moving to the stitching point, one view is morphed into the other by motion-compensated interpolation. Mismatches during the correspondence search have a visible effect that can be improved by filtering the motion vector field.

4.5 Automatic parameter adaptation

At the stitching point camera parameters are manually fixed for the concentric mosaic and omnidirectional video.

Parameters like position, viewing direction, field of view are forced to match the fixed values at the stitching point dependent on the distance of the current viewpoint from the stitching point. Figure 6 shows a capture area for the transition between concentric mosaics and omnidirectional video. In this work viewing parameters are modified towards the fixed viewing parameters at the stitching point. Different adaptation profiles are shown in figure 7 with different curves. Parameteradaptation while navigating using the omnidirectional video is calculated using the current frame number.

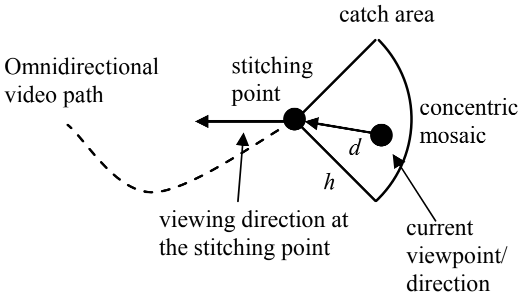

catch area
Omnidirectional video path
stitching point
concentric mosaic
d
h
viewing direction at the stitching point
current viewpoint/ direction

Figure 6: Parameter adaptation near the stitching point. Within a catch area the viewing direction and position is forced to the state at the stitching point if the user moves towards the transition point.

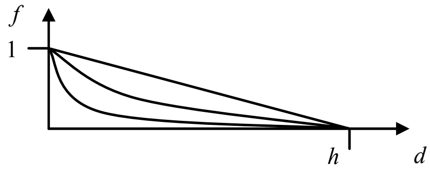


Figure 7: Example curves used for parameter adaptation near the stitching point. Position, viewing direction and field of view are adapted. At the stitching point a defined state is reached.

5 Results

In this chapter results are presented for stitching together concentric mosaics and omnidirectional video besides joint rendering of concentric mosaics and an alpha blended light field.

Concentric Mosaics and Omnidirectional Video

We rendered a concentric mosaic and omnidirectional video of a virtual scene. Intrinsic camera parameters except for the field of view of the capture camera are not known. Extrinsic camera parameters are known only for the concentric mosaic. A top view of the scene is shown in Figure 1. Figure 9a and 9b show the views at an arbitrarily chosen stitching point as illustrated in Figure 8. The result of manually correcting vertical position and field of view are shown in Figure 9c and 9d. The view using the omnidirectional video predicted from the concentric mosaic is shown in Figure 10a. The difference images at the stitching point before and after manual correction and after morphing are shown in Figure 10b,c and 10d. Four frames of the transition sequence are shown in Figure 11. Note the distortion due to a block mismatch near the bridge.

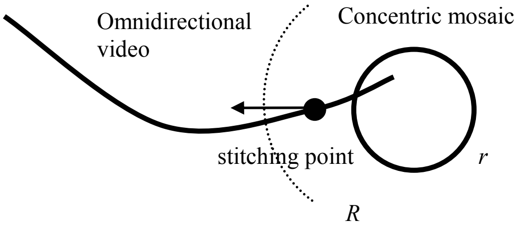


Figure 8: Stitching point and viewing direction for the transition between the concentric mosaic and the omnidirectional video.

Concentric Mosaics and Light fields

We rendered a simple light field using a concentric camera path and a blue background. Blue is chosen to be the alpha channel for blending the light field into the view of a concentric mosaic. Accurate placing is not possible due to vertical distortion and the constant depth assumption. However, horizontal parallax and perspective projection is rendered correctly. Figure 12 shows three views for the jointly rendered scene.

6 Conclusion and future work

The image-based scene representation and rendering system proposed in this paper uses different image based rendering techniques to provide a efficient and interactive walkthrough in a real environment. Concentric mosaics, omnidirectional video and light fields are used to represent a large scene. These representations can be acquired using a single standard camera and the rendering complexity is small. The choice of the appropriate partial scene representation is taken with respect to local scene content and complexity.

Simplifications and assumptions in the parameterization of different partial scene representations make it difficult to achieve a seamless transition between two representations.

An automatic camera parameter adaptation during the walkthrough is used to provide a good rendering quality at any viewpoint. The adaptation also restricts user movement at the transition point to ensure a transition with minimum distortion. Residual errors are alleviated using a simple morphing technique that is based on block matching and motion compensation.

Future work will include real representations and geometric objects to provide augmented reality scenes.

3D scene reconstruction algorithms will be used to automatically register different representations.

References

[1] S. E. H. Adelson, and J. R. Bergen, “The plenoptic function and the elements of early vision,” *Computational Models of Visual Processing, Chapter 1,* Edited by Michael Landy and J. Anthony Movshon. The MIT Press, Cambridge, Mass. 1991.

[2] S. E. Chen, “QuickTime VR – An Image Based Approach to Virtual Environment Navigation,” *Computer Graphics: Proc. of SIGGRAPH 95,* pp. 29-38, August 1995.

[3] A. Lippman, “Movie Maps: An Application of the Optical videodisk to Computer Graphics,” *Proc. Of SIGGRAPH 80,* 1980.

[4] L. McMillan and G.Bishop, “Plenoptic Modeling: An Image-Based Rendering System,” *Computer Graphics: Proc. of SIGGRAPH 95,* pp. 39-46, August 1995.

[5] S. E. Chen and L. Williams, “View interpolation for image synthesis,” In *Proc. SIGGRAPH 93*, pp. 279–288, 1993.

[6] S. Laveau and O. Faugeras, “3-D scene representation as a collection of images,” In *Proc. International Conference on Pattern Recognition*, pp. 689–691, 1994.

[7] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proc. SIGGRAPH ’96*, pp. 31–42, 1996.

[8] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Proc. SIGGRAPH’96*, pp. 43–54, 1996.

[9] Kang S.B. and Szeliski R., “3D Scene Data Recovery Using Omnidirectional Baseline Stereo,” *IEEE Computer Vision and Pattern Recognition (CVPR 96), pp.* 364-370, 1996.

[10] S. M. Seitz and C. R. Dyer, “View Morphing: Uniquely Predicting Scene Appearance from Basis Images,” *Proc. Image Understanding Workshop*, 1997, 881-887.

[11] http://graphics.stanford.edu/projects/lightfield/

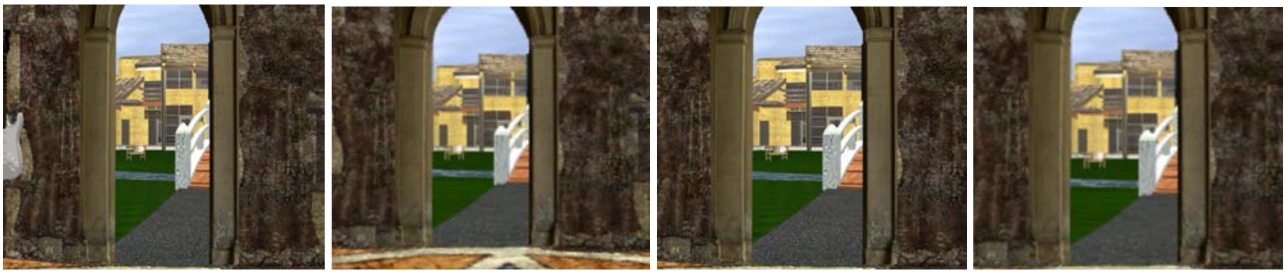
[12] Daniel G. Aliaga, Ingrid Carlbom, "Plenoptic Stitching: A Scalable Method for Reconstructing 3D Interactive Walkthroughs," Proceedings of ACM SIGGRAPH, pp. 443-450, 2001.

[13] Christopher Geyer and Kostas Daniilidis, “Omnidirectional Video,” *The Visual Computer*, accepted, 2002.

[14] H. Shum and L. He. “Rendering with concentric mosaics,” *Computer Graphics (SIGGRAPH'99)*, pp. 299-306, Aug. 1999.

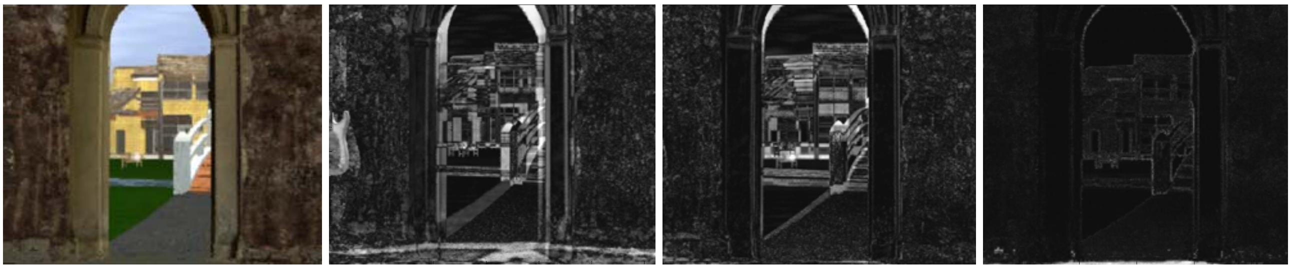
[15] Shum, H-Y. and Kang, S.B., “A Review of Image-based Rendering Techniques,” IEEE/SPIE Visual Communications and Image Processing, pp. 2-13, 2000.

[16] H. Shum, Lifeng Wang, Jin-Xiang Chai, and Xin Tong, “Rendering with Manifold Hopping,” International Journal of Computer Vision(IJCV), 50(2), pp. 185-201, November 2002.



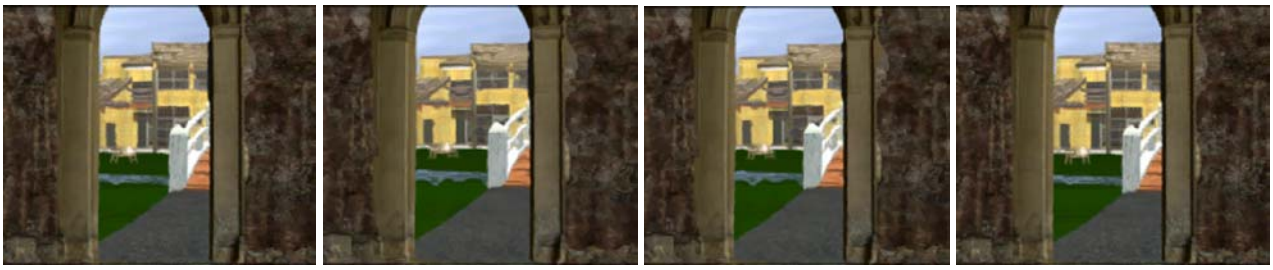
(a) (b) (c) (d)

Figure 9: Generated views at the stitching point using (a) the concentric mosaic and (b) the omnidirectional video with different camera parameters. Generated views at the stitching point after manual correction for (c) the concentric mosaic and (d) the omnidirectional video.



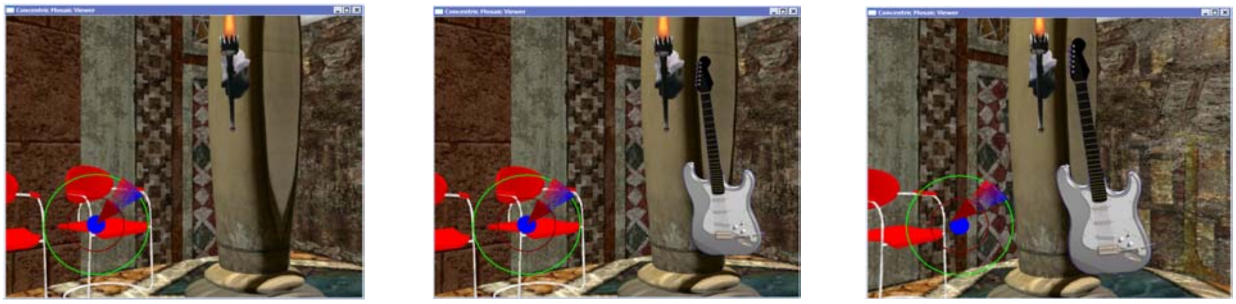
(a) (b) (c) (d)

Figure 10: (a) View of the omnidirectional video predicted from the view of the concentric mosaic at the stitching point. Difference images before (b) and after (c) manually correcting camera parameters at the stitching point. (d) Difference image after automatic smoothing of the transition using block matching.



(a) (b) (c) (d)

Figure 11: (a)-(d) Four frames of the transition sequence.



(a) (b) (c)

Figure 12: (a) View rendered using the concentric mosaic. (b) The light field is rendered into the concentric mosaic placed near the constant depth. (c) Another view of the jointly rendered scene after modification of the viewpoint and viewing direction.