

## Omnidirectional Image-based Rendering

Hynek Bakstein and Tomáš Pajdla

Center for Machine Perception  
Dept. of Cybernetics  
Faculty of Electrical Engineering  
Czech Technical University in Prague  
bakstein@cmp.felk.cvut.cz, pajdla@cmp.felk.cvut.cz

### Abstract

The paper summarizes omnidirectional image-based rendering techniques giving a user a complete freedom of motion in a virtual reality. Image-based representation provides a photorealistic scene representation. We discuss two VR scenarios, the “window into reality” and the “door of perception”. We overview the complete process from the acquisition to the rendering focusing on the advantages of omnidirectional images for the user experience. We also discuss the acquisition setups and practical issues such as the equipment used for real setups, presented for example at the CeBit 2005 fair.

### 1 Introduction

Image-based rendering (IBR) is an alternative way of creating a virtual environment. As opposed to computer graphics with a 3D model-based rendering, IBR does not require any scene model, or only a rough scene geometry description (for example an ellipsoid for a room) is sufficient. The main idea behind the IBR is an image synthesis, when a set of images is captured at some viewpoints and a much bigger number of novel viewpoints and viewing directions is synthesized. Moreover, the photorealistic output captures the real lighting conditions and scene properties such as reflections, transparency, occlusions, see Figure 1.

This paper summarizes our previous work in IBR done in the scope of an EU project BeNoGo [3]. The main purpose of this paper is to overview all necessary steps leading to a working IBR setup and present the results of the project. We will cover the required theory as well as the practical issues and give references to previous work for more details. The paper is organized as follows. Section 2 introduces IBR, Section 3 contains overview of X-slits, the IBR method we used, Section 4 presents example applications, and Section 5 gives notes on practical realization of the IBR with common hardware.

### 2 IBR methods

Following the classification from [11], one can divide the IBR methods with respect to the amount of the geometrical information required. One extreme are methods requir-



Figure 1: Comparison between a 3D model (top) and an IBR representation (bottom) of the same scene, the Aalborg tower.

ing a rough 3D model of the scene, for example [9], having the advantage that only a few images are required for the scene representation, the other extreme are methods which do not make use of any scene geometry, such as [6, 7], but require a huge number of images. The best trade-off is offered by methods which use a crude approximation of the scene to lower the number of input images. We have chosen X-slits [13] which are in general an example of a method which does not use any geometrical information, but the number of images can be greatly reduced by simple constant depth assumption without degrading the output image quality [2].



**Figure 2:** A consumer level compact omnidirectional camera with a fisheye lens.

### 3 X-slits IBR principles

The main idea behind X-slits IBR is a non-central projection where the light rays meet two “slits” instead of a single point. One of the slits coincides with the path of the controlled camera motion on a parametrizable path, such as a line or a circle. Provided with the controlled motion, the implementation gets much simpler, but this condition can be relaxed [10]. We assume that the camera moves on a controlled moving platform on a circle or a line and captures images in periodic time intervals. The other slit corresponds to the position of a user of an IBR system and is considered to be perpendicular to the plane in which the user moves. Since the camera moves in the same plane, only occlusion in the horizontal direction can be modeled, but this is sufficient for most cases.

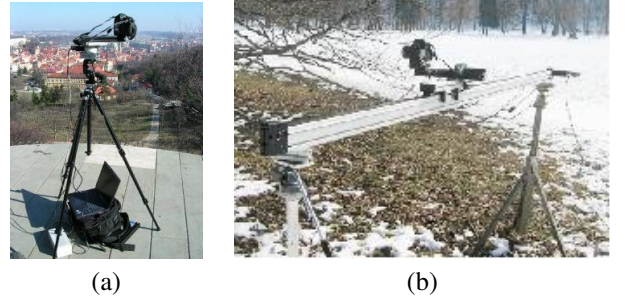
The X-slits IBR pipeline is as follows. First we acquire our images on a parameterizable path and compose them into the image volume. Image processing is required for omnidirectional images, but even narrow field of view images often need some radial distortion correction. We use fisheye lenses because they

- have field of view of  $180^\circ$  (off-the-shelf lenses, cheap and available),
- can be modeled by a central projection,
- and no part of the image is obscured by the camera.

Images from the fisheye lens have to be converted into volume images using off-line [1] or on-line [8] calibration method. An example of compact consumer level omnidirectional camera utilizing a fisheye lens can be seen in Figure 2. The acquisition setups can be seen in Figure 3.

The images in the volume are organized so that the  $z$ -axis is parameterized by the same parameter as the acquisition path and each image has columns and rows marked by angles. Then, we cut the volume by a slice depending on the camera motion (circular or linear) and the user position, that is the position of the vertical slit  $V$ . We have to determine which columns  $\beta$  in the volume images  $I_\gamma$  correspond to respective columns  $\alpha$  in the IBR image

$$\beta = f_\beta(\alpha, V), \quad (1)$$



**Figure 3:** Camera setups for (a) circular motion and (b) linear path.

$$\gamma = f_\gamma(\alpha, V). \quad (2)$$

Note that the IBR image has the same vertical field of view as the input images and the horizontal field of view is  $360^\circ$  for the circular motion and the same as the of the input images in the linear motion. We advocate using omnidirectional input images to give the user complete freedom in the viewing direction. Mind that the camera moves in the plane without rotation, narrow field of view would limit the experience in the virtual world. Proper image processing allows transformation of omnidirectional images into images which can be treated in the same way as narrow field of view ones [1]. We will concretize the functions  $f_\beta$  and  $f_\gamma$  in the following sections together with specific contributions of omnidirectional input images.

#### 3.1 Circular acquisition path

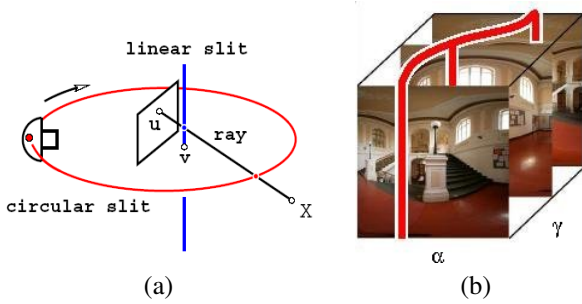
Our first acquisition system utilized a camera rotating on a motorized turntable looking outwards, see Figure 4(a). The IBR image corresponds to a non-linear cut through the volume, as it is depicted in Figure 4(b). This image is a panoramic image of the environment, only a small part is presented to a user who can move freely in the disc created by the camera acquisition path and look around in arbitrary direction. To maintain high visual fidelity of the IBR images, increasing the disc radius must result in a higher number of images. At the end of the BeNoGo project, we used a 120cm disc and 720 images. The functions  $f_\beta$  and  $f_\gamma$  become

$$\beta = \arcsin\left(\frac{R}{r} \sin(\alpha)\right), \quad (3)$$

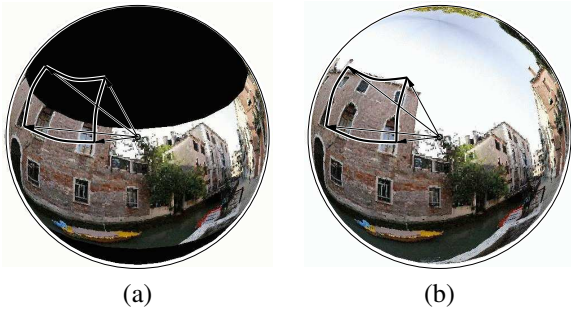
$$\gamma = \beta - \alpha. \quad (4)$$

The vertical slit  $V$  is determined by the ratio  $\frac{R}{r}$ , where  $r$  is the radius of the circular acquisition path and  $R$  is the distance of  $V$  from the center of rotation.

Omnidirectional images in this case allow the user to use the complete disc as the *region of exploration* (REX) and also allow arbitrary vertical viewing direction. Narrow field of view camera limits the viewing directions, see Figure 5(a), compare with Figure 5(b). As we mentioned before, the IBR images is a panoramic image covering  $360^\circ$  in the horizontal direction and omnidirectional input images add  $180^\circ$  in the vertical direction. Only a small part of this panoramic image is presented to the user as it is depicted in Figure 6. The size of the actual image depends on the field of view of the device used for visualization. Of course, a

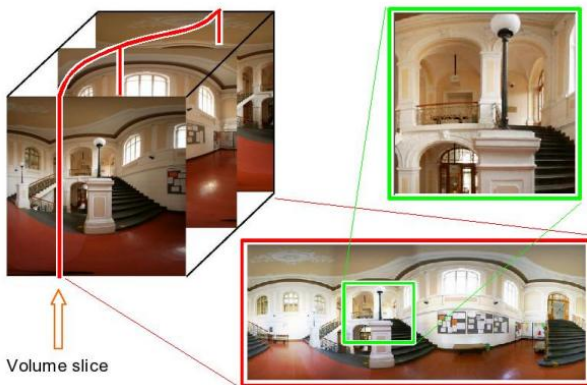


**Figure 4:** (a) Principle of X-slits rendering with circular camera path. (b) Each X-slits image corresponds to a non-linear cut through the image volume.



**Figure 5:** (a) Narrow field of view of the input images limits the viewing direction of the user. (b) Omnidirectional input images give freedom in the viewing direction.

mapping into the image coordinates from the spherical representation of the panorama is required.



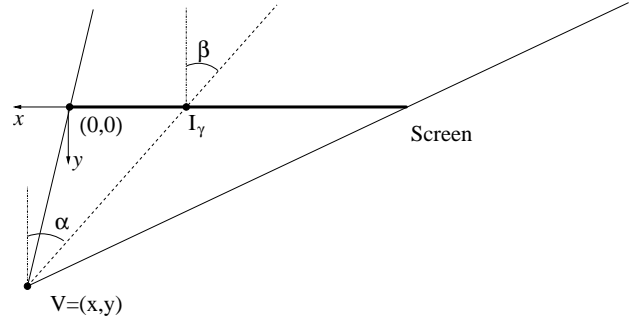
**Figure 6:** The panoramic image covers  $360^\circ \times 180^\circ$ , only a small part is presented to the user.

### 3.2 Linear acquisition path

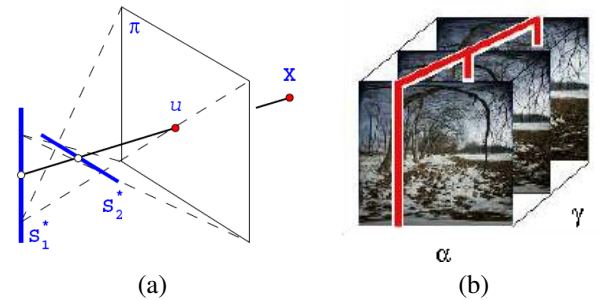
The second acquisition setup developed during the BeNoGo project consisted of a linear motion platform which can be extended up to 6 meters. The camera motion is again motorized, providing a constant speed. Each linear cut through the image volume corresponds to a X-slits image and its slope is determined by the user position with respect to the acquisition path [10]. The functions  $f_\beta$  and  $f_\gamma$  become

$$\beta = \alpha, \quad (5)$$

$$\gamma = y \arctan(\beta) + x. \quad (6)$$



**Figure 7:** X-slits rendering with linear acquisition path. The vertical slit is at position  $(x, y)$  and the IBR column  $\alpha$  is the column  $\beta$  from the image  $I_\gamma$ . See text for details.



**Figure 8:** (a) Principle of X-slits rendering with linear camera path. (b) Each X-slits image corresponds to a linear cut through the image volume.

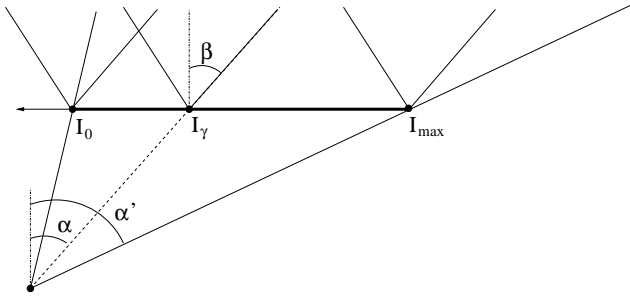
The vertical slit  $V$  is identified by its position  $(x, y)$  with respect to the acquisition path, the situation is illustrated in Figure 7. We can see that the above equation is linear in the position of the virtual camera and that for each virtual camera, we get a linear slice of the input volume, since the columns in the volume images as well in the virtual images can be organized by  $\tan \beta$ , see Figure 8.

In theory, one should acquire as many images as there are columns in the output image on an acquisition path with the same length as the screen used for visualization. While the latter condition can be easily fulfilled, storing up to a 1000 images in the memory is still quite a problem. But the number of images can again be reduced using the same arguing as for the circular motion [2].

Omnidirectional input images give the user unlimited field of view, as it is depicted in Figure 9. Limited field of view  $\beta$  means that there will not be light rays (columns from volume images) for IBR columns with  $\alpha > \beta$ . The motion of the user will be limited, while omnidirectional volume images allow the user to move arbitrarily in front of the screen.

Unlike in the case of circular acquisition path, the cut through the image volume leads to an image which is presented to the user directly after vertical scaling of the columns simulating the forward/backward motion, see Figure 10. Clipping by the display resolution is discussed later in this paper, but it also works on the columns only.





**Figure 9:** Limited field of view  $\beta$  does not allow to render full IBR image at all desired positions. All IBR columns between  $\alpha$  and  $\alpha'$  require columns from volume images that are outside their field of view.



**Figure 10:** The whole IBR images is presented to the user in case of the linear motion. Additional scaling works on the columns only.

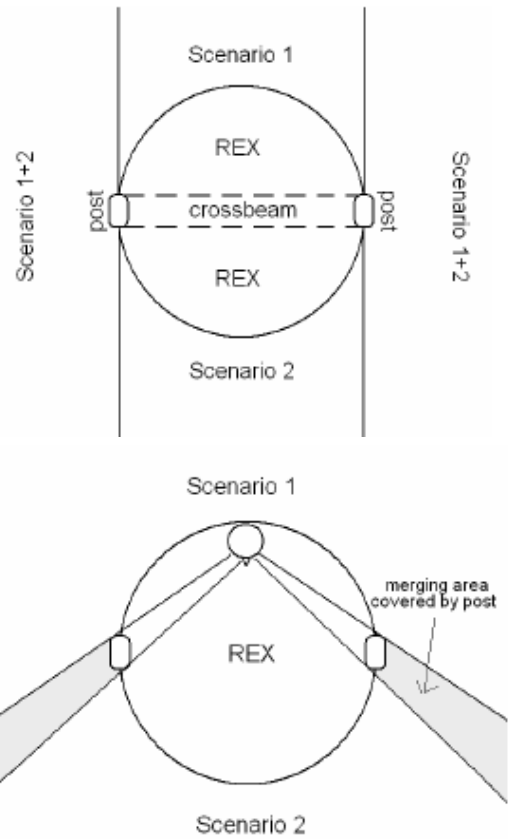
## 4 Example applications

We describe two applications developed as demonstrators for the BeNoGo project. The first one, developed by our partner in the project [12], requires head mounted display but shows that a very impressive scenario can be realized with IBR at no additional cost, just by combining two datasets. The second demonstrator does not require any specialized hardware, only a computer screen or projector with a webcam is sufficient for a realistic illusion of a virtual window.

### 4.1 Door of perception

The door of perception utilizes two or even more datasets captured on a circular path. Continuous segments of these datasets are connected together into a single set, a proper augmentation has to be used to cover the “seams”. The name “door of perception” comes from a demonstrator of the BeNoGo project shown at the CeBit fair, where both a real and a virtual doorframe create a separator between two IBR worlds. By looking and walking through the door, the user can move from one world to another in a natural, captivating, way. Top part of Figure 11 illustrates the principle of the door of perception. By moving in the REX, the user can move from Scenario 1 to Scenario 2. Recall that the IBR images are in fact mosaic images covering the whole  $360^\circ$  in the horizontal direction. This means that some part of the IBR images is composed only from images of Scenario 1 while some other part is composed only from images of Scenario 2. However, parts of the mosaic image can be composed from images of either Scenario 1 or 2, depending on the position of the viewer in the REX.

The composition of two datasets does not only create an interesting illusion but can also be used to get rid of “uninteresting” parts of the respective datasets and to encourage the



**Figure 11:** Principle of the door of perception scenario, see text for a description. Image from [12].

user to move around in the virtual environment. The datasets do not have to be split at an exact ratio, but the above mentioned demonstrator used 50/50 division for simplicity and we will assume it in the following paragraphs.

We mentioned above that a proper augmentation has to cover the seams, where the datasets are connected. A natural object is the doorframe, which can be also physically augmented enhancing the feeling of presence in the virtual world. The physical augmentation can be seen in Figure 12. More formally, the augmentation has to fulfill the following criteria

1. The augmentation has to be placed at the edge of the REX so that the user cannot walk past it.
2. The width of the augmentation has to be larger or equal to eye separation distance.
3. The augmentation must extend in the vertical direction towards infinity.

The augmented doorframe acts as a real doorframe, so that the user can see the second world through the door and outside the door, the user sees the world in which he is standing. This effect is illustrated in the bottom part of Figure 11 and is realized by the IBR method without any additional processing or computation but for a traditional 3D model based rendering, this would not be a trivial task.



Figure 12: Door of perception in action at CeBit 2005, from [12].

## 4.2 Window into reality

When the camera moves on a linear path, X-slits IBR can be used to simulate a “window” like experience. The user can move arbitrarily in front of a “window” represented by a computer or projection screen. The position of the viewer corresponds to the position of the vertical slit  $S_1$ , see Figure 8(a). The camera path corresponds to the horizontal slit  $S_2$  and coincides with the screen.

The user can actually move also up and down, the screen representing the window provides correct “clipping”, so that new parts of the image are uncovered by the vertical motion of the user, see Figure 13. These new parts are not generated by the IBR itself, instead they are represented by resampled columns of the volume images, because the camera moved only horizontally.

To simulate forward/backward motion, we have to scale the columns by a uniform scaling factor [13]. The rescaled image is clipped by the resolution of the visualization device. The scaling factor depends on the scene depth, see Figure 14.



Figure 13: Example frames from Window into reality demonstration at CeBit 2005.

The window REX gives convincing feeling of looking through a real window. A working setup was demonstrated at the CeBit 2005 exhibition. We plan to use the same principle of the window REX in an interactive art installation

created by a computer art group.

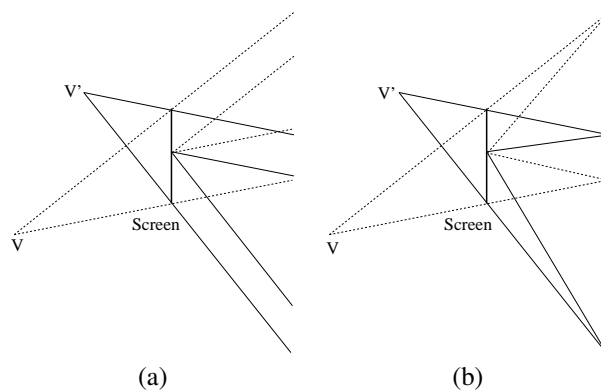


Figure 14: Clipping of the vertical field of view by the screen assuming (a) the scene in the infinity and (b) close objects.

## 5 Practical realization

We have used several technologies throughout the duration of the BeNoGo project. Data acquired for the disc REX were visualized using a head-mounted display (HMD) or in the CAVE [5]. The tracking was magnetic (Polhemus Fastrak) or ultrasonic (Intersense IS-600) combined with inertial sensor (Intersense InertiaCube2) for orientation. The same tracking was used for the window REX. But the tracking devices were available only through our partner in the project. Therefore, we also implemented an optical tracker using OpenCV [4] which allowed to run a light-weight version of the demonstration SW on a notebook. To cope with the memory requirements, a client server architecture was developed during the project, allowing distribution of volume image over nodes in a cluster of computers. However, the system demonstrated at CeBit was running on a single PC with 2GB of memory and the window into reality was presented on a notebook with only 512MB of memory.

## 6 Conclusion

We have presented a complete overview of steps required for practical implementation of an IBR system for photorealistic representation of real complex scenes in a computer. The steps covered acquisition systems, the necessary theory, example applications, and practical issues of realization using a commonly available hardware instead of specialized devices such as HMD and CAVE. We described in detail two scenarios with different acquisition setups and we demonstrated advantages of omnidirectional input images for both of them.

## Acknowledgement

The authors were supported by the FP5 EU under project IST-2001-39184, by STINT under project Dur IG2003-2 062 and by the Ministry of Education, Youth and Sports of the Czech Republic under the grant no. MSM6840770013.

## References

- [1] H. Bakstein and T. Pajdla. Rendering novel views from a set of omnidirectional mosaic images. In *Proceedings of Omnivis 2003: Workshop on Omnidirectional Vision and Camera Networks*, page cdrom only, Los Alamitos, CA, June 2003. IEEE Press.
- [2] H. Bakstein and T. Pajdla. Visual fidelity of image based rendering. In D. Skočaj, editor, *Proceedings of the Computer Vision Winter Workshop 2004 (CVWW'04)*, pages 139–148, Ljubljana, Slovenia, Feb. 2004. Slovenian Pattern Recognition Society.
- [3] BeNoGo. <http://www.benogo.dk>, 2002.  
<http://www.benogo.dk>.
- [4] Intel Corp. Opencv, open source computer vision library, 2001.  
<http://www.intel.com/research/mrl/research/opencv/>.
- [5] C. Cruz-Neira, D.J. Sandin, T.A. DeFanti, R. Kenyon, and J.C. Hart. The cave, audio visual experience automatic virtual environment. *Communications of the ACM*, pages 64–72, June 1992.
- [6] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Computer Graphics Proceedings (SIGGRAPH'96)*, New Orleans, pages 43–54, Aug. 1996.
- [7] M. Levoy and P. Hanrahan. Light field rendering. In *Computer Graphics Proceedings (SIGGRAPH'96)*, New Orleans, pages 31–42, Aug. 1996.
- [8] B. Mičušík and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. In D. Martin, editor, *CVPR 2003 : Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 485–490, Los Alamitos, USA, June 2003. IEEE Computer Society.
- [9] P. Rademacher and G. Bishop. Multiple-center-of-projection images. In *Computer Graphics Proceedings, Annual Conference Series SIGGRAPH'98 (Orlando)*, pages 199–206, July 1998.
- [10] A. Rav-Acha, Y. Shor, and S. Peleg. Mosaicing with parallax using time warping. In *Second IEEE Workshop on Image and Video Registration (IVR'04)*, 2004.
- [11] H.-Y. Shum and S. B. Kang. A review of image-based rendering techniques. In *IEEE/SPIE Visual Communications and Image Processing (VCIP) 2000*, pages 2–13, June 2000.
- [12] R. Stenholt, R. Laursen, and C. B. Madsen. Extending ve platform functionality. Technical Report D5.3, CVMT, Aalborg University, 2005.
- [13] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: The crossed-slits projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(6):741–754, June 2003.