Low-complexity Image-based 3D Gaming

Ingo Bauermann and Eckehard Steinbach

Institute of Communication Networks, Media Technology Group
Technische Universität München
Munich, Germany
{ingo.bauermann,eckehard.steinbach}@tum.de

Abstract

In this paper we present a low-complexity capturing and real-time rendering technique for interactive and photorealistic immersive gaming applications using 3D anaglyph vision. In contrast to conventional geometry-based rendering engines, this approach uses image-based techniques to guarantee high quality graphics that can be achieved at low computational costs. In combination with explicit depth information, interactive and augmented environments can be represented. The rendering process is simplified by approximating perspective projection to allow high quality graphics along with an interactive game-play even on mobile devices. Results show that depth perception is not affected by the introduced distortions.

1 Introduction

Interactive gaming applications traditionally use computer graphics approaches based on geometric modeling to create virtual worlds the user can navigate in and objects he or she has to interact with. Textures are mapped on surfaces and view dependent warping is used to generate novel views. Additionally, surface properties and special rendering techniques can be defined to achieve lighting effects like shadows and reflections. These modeling steps are tedious and time consuming tasks and realistic models can only be rendered in real time when optimized graphics hardware is available. Generally speaking, increasing realism in geometry-based gaming applications is accompanied by increasing computational complexity.

Since the introduction of the plenoptic function [1] in 1991, more and more research interest is focused on using photographs of the real world as a reference for view generation, e.g., [4], [5], [6], [7]. The plenoptic function describes a scene completely including lighting and surface properties like refraction and reflection. For every possible viewing position in space (x, y, z), for every possible viewing direction (φ, ϕ), and for every wavelength λ one intensity sample is taken. If changes of the structure of the scene or the lighting over the time t are considered the plenoptic function becomes a seven dimensional description of a scene:

$$P_7 = P(x, y, z, \varphi, \phi, \lambda, t)$$

As it is not possible in practice to sample the plenoptic function in all seven dimensions, simplifications have to be made. For this, some degrees of freedom are often not considered, e.g., discrete wavelength sampling, a static scene and constant intensity along a light ray are assumed for light fields and the lumigraph [5],[6]. Furthermore, for concentric mosaics [7] the y dimension is not considered which constrains user navigation to a fixed height. Such image-based scene representations can be acquired using standard video equipment. Arbitrary views of the real scene are rendered by interpolating parts of the captured reference images. In image-based rendering the rendering complexity does not depend on the complexity of the scene. Figure 1 shows the relation between rendering complexity in geometry-based scene representations and image-based scene representations as a function of the scene complexity.

A drawback of scene independent rendering complexity for image-based representations is the huge amount of data that has to be stored containing the reference images. Modified standard compression schemes can be used to handle image-based scene representations of larger areas properly and at the same time provide the desired random access to image data, for an overview see [8].

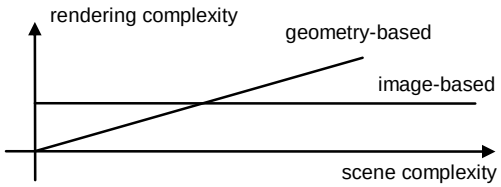

Figure 1. Rendering complexity as a function of the scene complexity for image-based and geometry-based scene representations.

A widespread image-based scene representation with relatively low memory consumption is a panoramic image. For scene representations consisting of panoramic images no smooth translational movement is allowed. View generation is possible for free rotation around a vertical axis. The plenoptic function is sampled in two dimensions:

$$P_2 = P(\varphi, \phi)$$

Panoramic images covering 360 degrees of horizontal field of view can be created using panoramic cameras (see e.g. [10]) or by stitching together several regular images taken from the same viewpoint. Interactive systems that use panoramic video are already in use [4]. Such systems allow translational navigation along the path the video was captured on. Because of the simple acquisition and low memory consumption, panoramic imaging is chosen for scene representation in this work.

In order to allow the user to interact with the environment and objects, depth information has to be available to handle occlusions and collisions correctly. For geometry-based models, depth information is inherent. Image-based scene representations, however, do not need explicit depth information for photorealistic view generation. Depth information has to be extracted from the data set or explicitly captured, calibrated, and stored for a scene representation. Exact geometry reconstruction from real images is a challenging task. Even for densely sampled image based representations, a general solution to the correspondence problem has not been found yet. Explicitly capturing depth using range scanning devices is a promising alternative.

The geometric information captured can be used for several rendering tasks in image based rendering. It can be used for generating stereo views from a single panorama as well as for lighting effects and occlusion handling in augmented reality systems. Integrating image-based scene representations and geometric approaches for interactive gaming systems is the aim of this work.

The remainder of this paper is structured as follows. In Chapter 2, the scene acquisition consisting of the capturing of a single viewpoint panorama and explicit depth information is described. Chapter 3 discusses the generation of stereoscopic panoramic images to render anaglyph views. In Chapter 4 simplifications for the rendering process and the distortions introduced are investigated. Augmentation with geometric models is described in Chapter 5. In Chapter 6 experimental results for the proposed low complexity interactive gaming application are presented. Chapter 7 gives conclusions and remarks for future work.

2 Scene Acquisition

The acquisition system used in this work consists of a digital still image camera SONY DCS-S85, a tripod and a laser range finder from SICK [11]. First, a panoramic image is composed from a series of regular images. The images are taken approximately from the same viewpoint and for a set of viewing angles covering the full 360 degrees horizontal field of view. The determination of intrinsic camera parameters is performed using the algorithm described in [12]. A part of the resulting panorama of a lecture room after stitching and blending is illustrated in Figure 2.

Figure 2. Part of the panoramic image of the captured scene.

Spatially coarse depth information is obtained by mounting the laser range finder on the tripod at approximately the same viewpoint as the still image camera. The construction of the depth panorama at color panorama resolution is done by a simple color segmentation algorithm performed on the color information to preserve accurate correspondences especially at depth discontinuities. This procedure is semi-automatic. Depth errors due to a weak calibration at object boundaries are manually corrected by adjusting color and depth edges. One part of the resulting depth panorama is shown in Figure 3.

Figure 3. Part of the depth panorama of the captured scene.

3 Stereo Panorama Composition

For 3D stereoscopic rendering, two multi viewpoint panoramas are precalculated from the single viewpoint panorama for the right and the left eye, respectively. Depth information z is obtained from the depth panorama. Figure 4 shows the construction of the stereo panoramas.

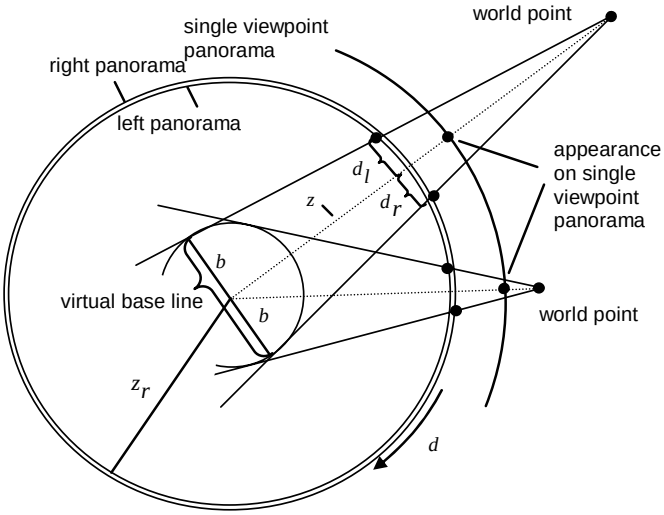

Figure 4. Construction of the stereo panorama. The disparity between the left and right panorama is calculated from the depth information z.

For every pixel in the single viewpoint panorama, a corresponding depth value z and a fixed relative depth z_r are used to calculate the displacement in the left and right multi viewpoint panorama dependent on the virtual baseline $2b$:

$$d_r = \frac{b(z - z_r)}{z} = -d_l \qquad (1)$$

This warping step is performed using bilinear interpolation. Occlusion artifacts at depth discontinuities caused by translating the single viewpoint to multiple viewpoints are illustrated in Figure 5. Interestingly, these artifacts do not appear as annoying during stereoscopic rendering as they do for monocular rendering. Reducing these artifacts could be achieved using, e.g., the approach proposed in [13].

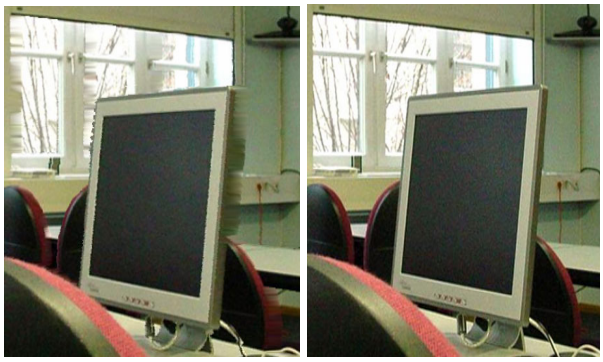

Figure 5. Artifacts at depth discontinuities in the right panorama caused by occluded regions (left) and the same cutout from the single viewpoint panorama (right).

The resulting scene representation is similar to the one described in [3]. Figure 6 shows one perspective view onto the stereo panorama as a red-cyan anaglyph stereo image. The relative depth z_r is chosen to match the mean depth of the scene. High quality versions of the color images can be viewed from [2].

Figure 6. Perspective stereoscopic view rendered from a single viewpoint panorama with depth information (anaglyph).

4 Interactive Rendering

To reduce the rendering complexity for the interactive gaming application, only one panorama is precalculated from the single viewpoint panorama and the corresponding depth information to achieve a photorealistic stereoscopic environment. The red channel on this single panorama is used to provide the left view. The blue and green channels

hold the right view and additionally give an impression of color for rendering with the red-cyan anaglyph technique. A fourth channel stores the depth value for every pixel for augmentation with geometry based objects as described in the next chapter.

A further reduction of the rendering complexity is achieved by replacing the view dependent perspective mapping from the panoramic image to the planar image plane by a simple orthographic projection of the panoramic image unfolded to a planar image. This can equivalently be represented by warping the image plane to a cylinder. Together with the disparity from equation (1), the rendering geometry for a pinhole model and our approach are illustrated in Figure 7.

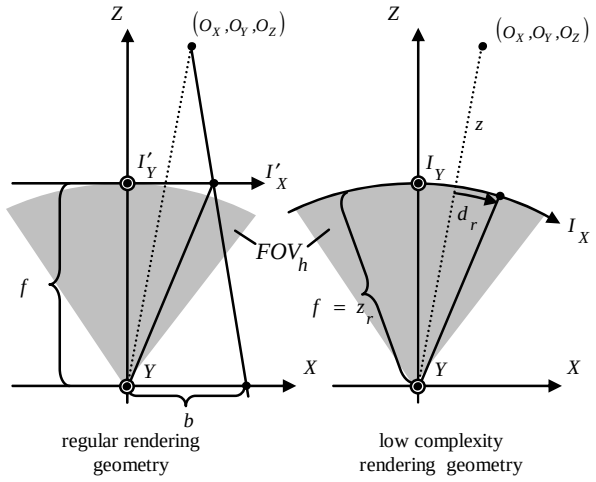

Figure 7. Top view of the projection of a world point *O* onto the right eye image plane for a pinhole model (left) and the projection proposed in this work using the disparity from equation (1) (right).

For error analysis the complete mapping from world coordinates *(X,Y,Z)* to the right eye image plane is investigated. For the pinhole model this mapping becomes:

$$I'_X = b + f \cdot \frac{(O_X - b)}{O_Z}$$

$$I'_Y = f \cdot \frac{O_Y}{O_Z}$$

Similarly, the simplified mapping proposed in this work can be expressed as follows:

$$I_X = d_r + f \cdot \operatorname{atan}\left(\frac{O_X}{O_Z}\right)$$

$$I_Y = f \cdot \frac{O_Y}{\sqrt{{O_X}^2 + {O_Z}^2}}$$

In contrast to I'_Y for a rotation of the coordinate system, which corresponds to a change of the virtual viewing direction, I_Y is independent from this viewing direction. The maximum angular error $\Delta\varphi_h$ and $\Delta\varphi_v$ between the pinhole model and our simplified approach in horizontal and in vertical direction for a rendered right eye view is illustrated in Figure 8 and Figure 9. The reference depth and focal length is $f = z_r = 1$, the virtual baseline is $2b = 0.66$ and the horizontal and vertical field of view of the virtual camera is $FOV_h = 50°$ and $FOV_v = 40°$, respectively.

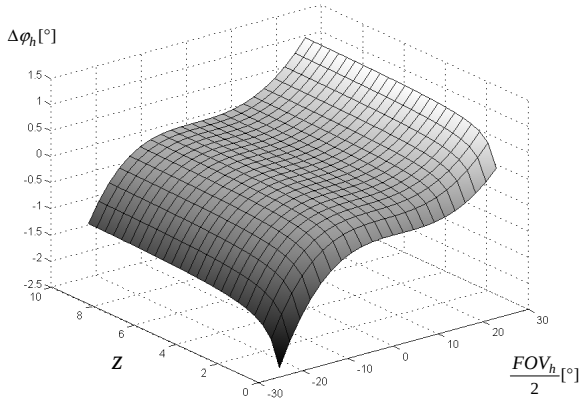

Figure 8. Maximum angular horizontal error $\Delta\varphi_h$ for rendering a virtual view from a stereo panorama using our approach compared to a pinhole camera model. By constraining the horizontal field of view, the horizontal error can be kept small for a reasonably large field of view.

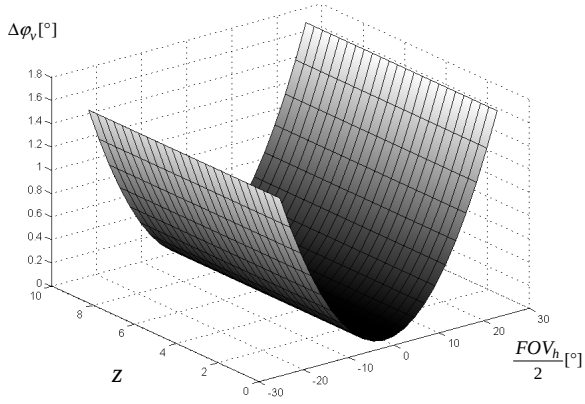

Figure 9. Maximum of the absolute value of the vertical angular error $\Delta\varphi_v$ for rendering a stereo panorama using our approach. By constraining the vertical or the horizontal field of view, the distortions caused by not using a perspective projection can be minimized.

These large angular errors might not seem acceptable, however, depth perception is not affected, as only a small error is introduced for local relative disparity (see [3] for a detailed description). Figure 10 shows two right eye views from a single viewpoint panorama rendered using a pinhole camera model and the simplified model. Note that the simplified mapping is not straight line preserving as a perspective mapping is.

The main advantage of this technique is that, in the case of low complexity rendering, a rotation around the viewpoint results in a simple copy operation of pixel data of a rectangular cut-out of the panoramic image to the application window (see Figure 11). No warping due to the perspective mapping from a cylindrical to a planar surface is needed.

Figure 10. Rendered right eye views from the single panorama with depth information for the pinhole model (left) and the simplified model (right). Note the curved rear edge of the desk in the front for the simplified model.

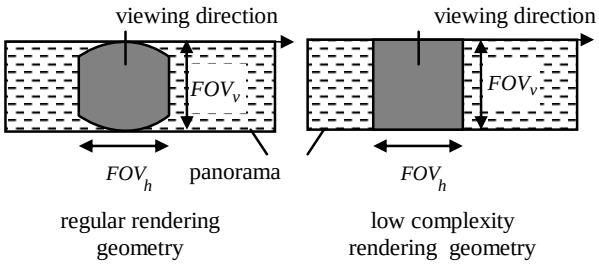

Figure 11. Pixel data used to render a virtual view (gray) from the panoramic image (dashed) for the pinhole model (left) and the proposed scheme (right).

5 Scene Augmentation

The rendering technique described in Chapter 4 provides an interactive image-based scene representation which allows a free choice of the viewing direction around a vertical axis. The scene can be viewed in 3D using red-cyan anaglyph glasses. The depth information stored with the stereo panorama additionally allows for inserting geometry based objects and even other image based representations. Occlusions and collisions can be rendered correctly.

Except for precalculating a stereo panorama, the scene depth is used to add a third dimension to the game play. In our gaming application a target is rendered into the scene and the user has to aim by placing a cross-hair correctly in horizontal and vertical direction. Additionally, the depth of the cross-hair has to be adjusted to hit the moving target. A predicted position of the target has to be estimated as the bullet shot into the scene moves on a parabolic trajectory and takes a while to hit the 3D-position of the cross-hair. A stereoscopic view with augmented images is illustrated in Figure 12. Color versions can be obtained from [2].

The images of the inserted objects, the cross-hair, an animated bird, a snowball, a cannon and a score counter, are rendered in the left and right view using the disparity measured using (1) according to their position in space. The size is also adjusted according to the depth of the object.

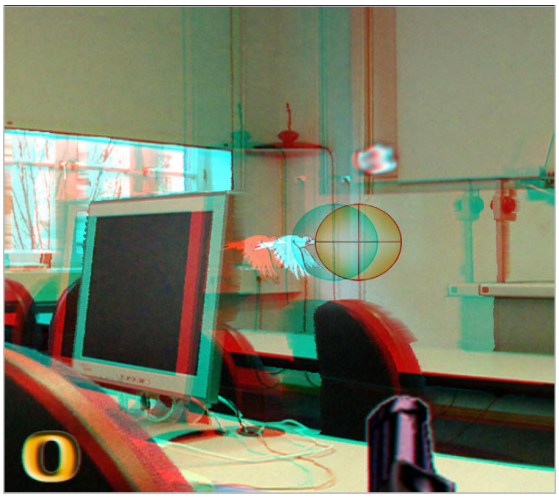

Figure 12. A stereoscopic view of the interactive 3D game. The crosshair has to be adjusted horizontally, vertically, and in depth. A snowball is shot from the cannon following a parabolic trajectory into the scene hitting the 3D position of the cross-hair.

A Z-buffer is used to render occlusions correctly by comparing the objects Z-values with the scene depth pixel by pixel. Alpha blending is achieved by setting the depth value of transparent pixels of the augmented objects to infinity. The collision of

the snowball with the target is detected by comparing their position in space.

6 Results

The rendering process of the proposed stereoscopic 3D game is very fast as only pixel data has to be copied for the image based environment. Interpolation is only needed to adjust the size of the inserted objects. This also can be simplified by prestoring image data for various object sizes. The angular error introduced by the simplified rendering process is up to 3° for common virtual camera parameters. Though this error is large, depth perception is not affected, as disparity is rendered only with a small error. Color videos of the running application for perspective mapping and our simplified approach as well as color versions of all images can be viewed at [2].
Figure 13 shows the occlusion handling of the gaming application. The cross-hair is partly placed behind the monitor.

Figure 13. Inserted cross-hair into the image-based scene representation. The monitor occludes the cross-hair partly.

The relative depth can be adjusted to move the whole scene forward and backward. Figure 14 shows the gaming application with some inserted objects appearing in front of the image plane.

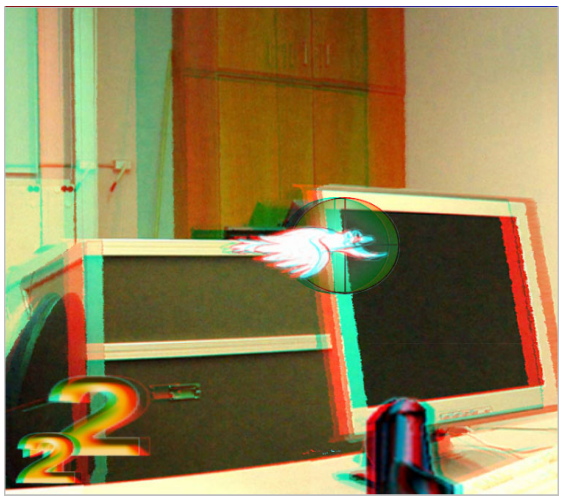

Figure 14. Scene rendered with a relative depth chosen as the mean depth of the scene. Foreground objects appear to be in front of the image plane.

Figure 15 shows the result of the collision detection. The snowball hits the projector and is stopped.

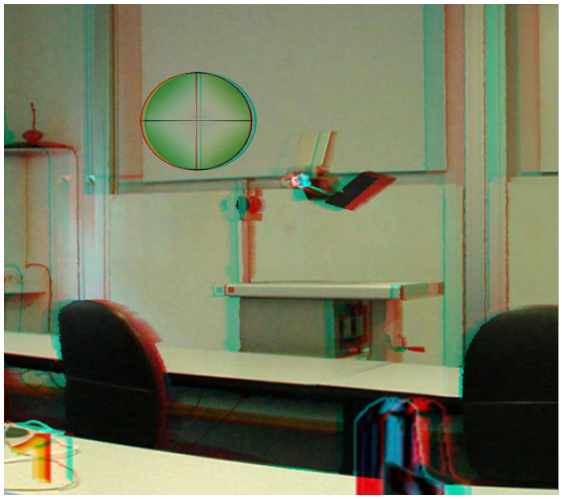

Figure 15. Collision of the snowball with the projector. The trajectory gets blocked by obstacles in the scene.

Figure 16 shows a view of the scene where distortions due to the simplified rendering process can be perceived clearly. Depth perception, however, is not affected.

Figure 16. Distortion caused by the simplifications of the rendering process. The horizontal lines in the upper part of the image appear curved. Depth perception is not affected.

7 Conclusion and Future Work

In this paper we present an image-based gaming application which can be viewed in stereo. Virtual objects are inserted into the scene and a 3D interaction is allowed to modify the objects in the scene. The complexity is very low as only copy operations are needed to render views onto the scene. The distortions introduced are large but acceptable. Depth perception is not affected. The low complexity rendering process makes our technique suitable for photorealistic 3-D gaming applications on mobile devices with low computational resources. The anaglyph representation for stereo images in this paper could be replaced by autostereoscopic views. Autostereoscopic displays for mobile devices are currently under development and will soon hit the market.
Future work will include reconstruction of occluded regions during the creation of the stereo panorama as well as automatic calibration of the camera and laser scanner. Translational movement of the capturing camera could significantly reduce the distortions without hurting complexity.

References

[1] E. H. Adelson and J. Bergen, “The plenoptic function and the elements of early vision,” *Computational Models of Visual Processing*, pages 3–20. MIT Press, Cambridge, MA, 1991.

[2] http://www.lkn.ei.tum.de/lkn/mitarbeiter/ingob/IBR/SP.htm

[3] S. Peleg and M. Ben-Ezra, ”Stereo panorama with a single camera,” *Proc. Computer Vision and Pattern Recognition Conf.*, 1999.

[4] S. E. Chen, “QuickTime VR – An Image Based Approach to Virtual Environment Navigation,” *Proc. of SIGGRAPH ‘95*, pp. 29-38, Aug. 1995.

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

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

[7] H. Shum and L. He., “Rendering with concentric mosaics,” *Proc. of SIGGRAPH’99*, pp. 299-306, Aug. 1999.

[8] C. Zhang and T. Chen, "A Survey on Image-Based Rendering - Representation, Sampling and Compression," *Carnegie Mellon Technical Report: AMP03-03*.

[9] S. Baker and S. K. Nayar, “A theory of catadioptric image formation,” *IEEE ICCV’97*, pp. 35–42, Jan. 1998.

[10] Shree K. Nayar, "Omnidirectional Vision," *Proc. of Eight International Symposium on Robotics Research (ISRR)*, Shonan, Japan, Oct. 1997.

[11] http://www.sick.de/de/products/categories/auto/lasermeasurementsystemsindoor/lms_291_s05/de.html

[12] http://www.vision.caltech.edu/bouguetj/calib_doc/

[13] Iddo Drori, Daniel Cohen-Or, and Hezy Yeshurun, “Fragment-Based Image Completion,” *Proc. of SIGGRAPH 2003*, pp. 303-312, ACM Press, New York, NY, 2003.