Imaging Geometry for Concentric Mosaics.

G. C. Watson, N. Vaughan and M. W. Wright

Edinburgh Virtual Environment Centre, The University of Edinburgh, Scotland, UK

Abstract

Image-based models of environments can be captured with the use of a rotating video camera. In this paper we consider the nature of the display to develop a simple quantitative measure of the quality of ray sampling which we use to compare different imaging geometries for concentric mosaics. We propose a novel arrangement for the rotating camera, which we show to produce a better sampling of rays for viewing on displays where the display surface is fixed in space and the view position is varied.

We show how this improved capture geometry can be employed in a head tracked display to produce a display that resembles a virtual window through which a captured environment can be viewed.

1. Introduction.

The traditional rendering approach to computer graphics starts by constructing a complete geometric and photometric model. Recently, image-based modelling and rendering techniques have been proposed by Faugeras[3] and others to recover geometric and photometric models from a collection of photographs.

Constructing such models is, in general, time-consuming, often difficult or sometimes impossible for real environments. Having such models makes it possible to add other graphical elements, such as lights and shadows later. On the other hand, if we only want to re-render images at a collection of viewpoints, all we need is a plenoptic function[1], which describes the irradiance perceived from all the observer's viewpoints.

Much image-based rendering work has been based on the plenoptic function. The original 7D plenoptic function was defined as the intensity of light rays passing through the camera centre at every location, at every possible viewing angle, for every wavelength and at any time. The Lumigraph[5] and Lightfield[6] systems presented a clever 4D parameterisation of the plenoptic function if the scene (or conversely the camera view) can be constrained to a bounding box. If the viewpoint is fixed and only the viewing directions and camera zoom can be altered, the plenoptic function simply becomes a 2D panorama.

The dimensionality of each of the various plenoptic functions depends on the space of viewpoints (or images) it represents. To represent all possible viewpoints without artefacts, we need to have a complete plenoptic function, which is 5D ignoring time and wavelength. As long as we stay outside the convex hull of an object, we reduce the dimensionality of the plenoptic function from 5 to 4 and if we stay at a given point the dimensionality further reduces to only 2 (e.g., a panorama).

2. Concentric Mosaics.

Shum and He[4] presented a novel *3D* plenoptic function that they termed *concentric mosaic*s. By constraining camera motion to planar concentric circles, they created image-based objects by composing slit images taken at different locations along each circle. Concentric mosaics index all input image rays naturally in 3 parameters: radius, rotation angle and vertical elevation. Unlike panoramas in which the viewpoint is fixed, concentric mosaics allow the user to move freely in a circular region and observe significant parallax and lighting changes without geometric or photometric scene models.

Figure 1 shows two rays at the extreme ends of the field-of-view of a rotating camera. The radius of the concentric mosaic indexing these rays is r, the two points on the mosaic where these rays are actually indexed are shown as P and P'. Columns of rays from closer to the centre of the camera's field of view are indexed into concentric mosaics with a radius less than r.

A view can be reconstructed for any viewpoint enclosed within the largest concentric mosaic – of radius *r* in figure 1. Refer to Shum and He[4] for a full rendering equation.

Rendering with concentric mosaics causes vertical distortions in the rendered images, because off-the-plane rays cannot be synthesized correctly without knowing depth information. Shum and He[4] introduce a simple depth correction by assuming all points in the scene are at a constant distance from the viewer.

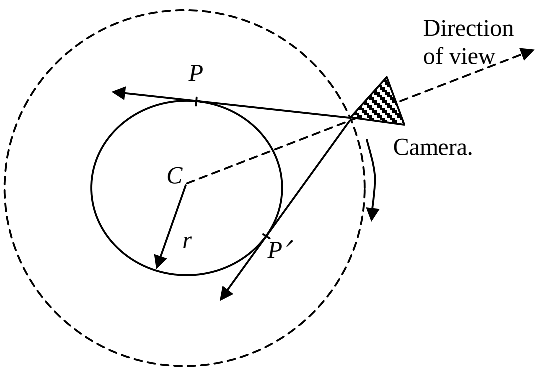

Figure 1. *Vertical view of rotating outward facing camera rig for capture of concentric mosaics.*

By considering the viewing geometry, this paper presents a simple analysis of errors introduced into the rendered image by this vertical approximation and proposes an alternative capture geometry that can be shown to perform better for situations where the display surface is fixed in space. These errors in the reconstructed images still occur even when accurate depth information is available to correct the vertical distortions. This is due to occlusions and lighting effects – our analysis still applies although the perceived improvement will be much less.

3. Capture of concentric mosaics.

Shum and He[4] observe that a concentric mosaic can be created from an outward pointing camera rotated about a point behind the optical centre of the lens system. This configuration is shown in figure 1.

We propose an alternative configuration for this capture rig by placing the camera pointing towards the centre of rotation rather than away from it, this we refer to as the *inward* pointing case (figure 2). This configuration clearly involves a more carefully designed rotation rig than the outward pointing case to prevent parts of the camera rig appearing in the cameras field-of-view.

4. An analysis of error.

In any graphics system it is difficult to produce an quantitative expression for error since it is the perceived quality of the final rendering that we are trying to evaluate. However in the simple case of reconstructing a novel ray from a sample set we can derive an expression relating to the quality of the sampling with respect to the desired ray.

A concentric mosaic consists of a collection of rays sampled from regularly spaced locations around a circular path. In order to reconstruct a novel ray, its irradiance needs to be deduced from the sampled set. This process involves interpolation from the values of the nearest rays in the sample set. Since we can parameterise a ray in terms of its direction and a point through which the ray passes, we can derive an expression for the quality of a derived novel ray in terms of the angular difference between the novel ray and the sampled ray, and the tangential distance between the two rays.

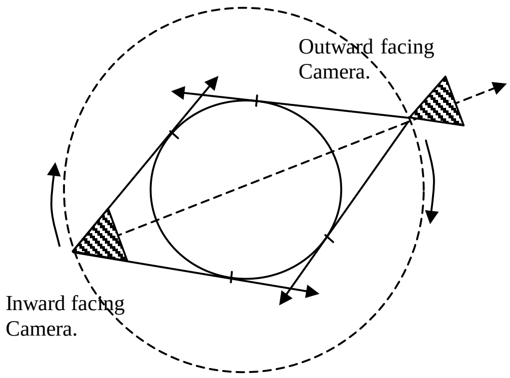

Figure 2. *The alternative inward facing camera position.*

In the absence of geometric information about the scene, Shum and He[4] approximate the depth of the scene to be constant; their choice of sample ray to reconstruct the off-plane image rays is determined by intersecting the desired ray with a surface at the approximate scene distance and finding the nearest sample ray that intersects that point. The quality of the final rendered images in this case clearly depends upon the depth complexity of the scene and the accuracy of the approximation.

For clarity of our analysis we don't approximate the depth of the scene to be a constant. Our choice of sampled ray to use for reconstruction is simply the ray closest in direction to the required ray. Suitability of this ray for reconstruction depends on the difference between the two rays in space; since the sample ray is parallel to the required ray we define our error metric to be the tangential distance between the required ray and the sample ray – referred to as *e* in figure 3.

We further assume direction errors due to finite resolution of the cameras to be small – i.e. a camera captures all rays intersecting its optical centre that are within its field of view.

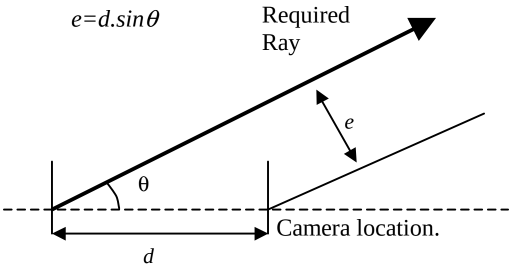

Figure 3. *A vertical section showing an arbitrary off-plane ray.*

From figure 3, for 0< θ < 90° it can be seen that the error *e* is directly proportional to the distance between the position of the virtual camera and the nearest sample camera position – labelled *d* in figure3; we have shown a well-known and perhaps rather obvious *maxim* of image-based modelling and rendering to be true for our assumptions, namely the best reconstructed views come from locations closest to the original camera positions!

For our proposed capture rig geometries, we now have the means to compare the *inward* and *outward* pointing cases for different displays.

Figure 4 shows a viewing situation where the viewer is looking at a fixed display surface, such as on a monitor screen. In our application the image is keystone corrected to remove distortion effects of viewing the screen off-axis, creating the impression of looking through a window rather than at a screen. The off-axis viewing geometry further allows the generation of correct stereo views.

Two camera locations are shown, the outward pointing case denoted *C* and the inward looking case denoted *C′*. The inward pointing camera can be seen to be closer to the view position for this configuration and hence is a superior sampling position for the camera.

In effect the display surface is fixed in space and the view position is moving around the fixed display - the direction of view is always towards the centre of the display.

To ensure all rays intersecting the display have been captured, the display surface has to be contained inside the diameter of the largest concentric mosaic. From figure 4 it can be seen that the most efficient alignment between the display and the real centre of rotation of the capture rig is for the centre of the display to be placed at the centre of rotation, and hence at the centre of the concentric mosaic.

Since the display is positioned at the centre of rotation, the view direction is always pointing *inwards* towards the centre and hence the inward pointing camera will always produce the better sampling.

If the display surface is able to move, such as in the case of a head-mounted display, the view direction will be *outward* from the centre as the viewer turns their head. This results from the centre of rotation of the head being located *behind* the eyes. Thus when the head rotates the eyes translate in a similar manner to the outward pointing camera; the outward pointing capture geometry is clearly more appropriate for the head-mounted display device.

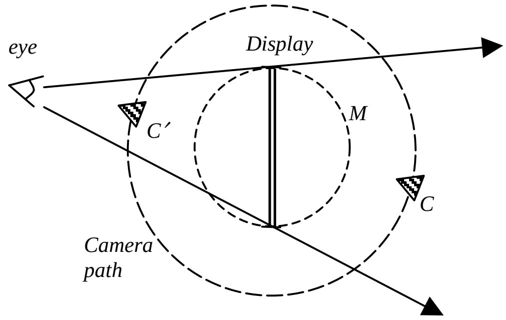

Figure 4. *Typical viewing geometry for viewing a fixed display device such as a monitor.*

5. An application.

Using the capture geometry thus described, we have implemented a novel tele-presence application allowing a head-tracked user to move freely in a horizontal half-plane bounded by the plane of a monitor. When viewing a geometrically correct virtual environment in stereo through the monitor, the experience is more akin to looking through a window than at a display device. The head-tracked user has the ability to move their head horizontally to any position in front of the display. They have no vertical freedom of movement, but in practice viewers rarely find this constraining and still have a strong sense of 3D perception, as was demonstrated some years ago by Benton[2] with experiments using holograms.

6. Conclusions.

Due to the diverse nature of scenes and displays, it is very difficult to produce a quantitative measure for image quality in an image-based capture and rendering system.

In this paper we have made a number of simplifying assumptions in order to make possible the derivation of such a quantitative measure. This measure has shown us the importance of knowing the likely viewing and display geometry.

We have introduced a novel capture rig geometry for concentric mosaics which we have shown to be superior in cases where the display surface is fixed and the viewer is free to move.

7. Further Work.

Shum and Szeliski[7] extract depth information from a concentric mosaic using a stereo matching algorithm. Further investigation will be required to establish whether the inward pointing camera geometry offers any advantages for stereo matching.

Acknowledgments.

This work forms part of a Master of Science dissertation in Informatics at the University of Edinburgh. The authors would like to acknowledge the assistance of Sharp Laboratories of Europe for the loan of equipment for this project, and the UK Engineering and Physical Science Research Council for funding support.

References

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

2. S. A. Benton. Survey of holographic stereograms. In *Proc. SPIE Vol. 367 Int. Soc. Eng*., pages 15–19, 1983.

3. O. D. Faugeras, Laveau S., Robert L., Csurka G., and Zeller C. 3-D reconstruction of urban scenes from sequences of images. *Computer Vision and Image Understanding*, 69(3):292–309, March 1998.

4. H-Y Shum and L-W He Rendering with Concentric Mosaics. In *Computer Graphics Proceedings, Annual Conference Series*, pages 299-306, Proc. SIGGRAPH'99 (Los Angeles), August 1999. ACM SIGGRAPH.

5. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Computer Graphics Proceedings, Annual Conference Series*, pages 43–54, Proc. SIGGRAPH'96 (New Orleans), August 1996. ACM SIGGRAPH.

6. M. Levoy and P. Hanrahan. Light field rendering. In *Computer Graphics Proceedings, Annual Conference Series*, pages 31–42, Proc. SIGGRAPH'96 (New Orleans), August 1996. ACM SIGGRAPH.

7. H-Y Shum and R. Szeliski. Stereo Reconstruction from Multiperspective Panoramas. In *Seventh International Conference on Computer Vision (ICCV'99*), pages 14-21, Kerkyra, Greece, September 1999.