

SPACE-VARIANT IMAGE CODING FOR STEREOSCOPIC MEDIA

Arzu Çöltekin

Geographic Information Visualization and Analysis, Department of Geography,
University of Zurich, Switzerland

ABSTRACT

This paper presents a brief overview of space variant image coding for stereoscopic media and reports on findings from a study using foveation for stereoscopic imaging. Foveation is a perceptually motivated approach to image coding based on the structure of human fovea and it is well studied in image and video processing domains. However it is less exploited for three-dimensional (3D) space even though it is potentially well suited also for 3D, *e.g.* for level of detail management in gaze contingent stereoscopic displays. In this paper we present results from a stereoscopic foveation implementation to test this argument. A brief discussion on computational as well as human factors for successful management and presentation of stereoscopic media is also provided based on current literature.

Index Terms— Stereoscopic vision, stereoscopic visualization, stereoscopic media, foveation, gaze-contingent displays

1. INTRODUCTION

Stereoscopic media has been used across domains, some of them mainstream some of them highly specialized, *e.g.* entertainment industry (stereoscopic films and animations), immersive visualizations in virtual environments (games, interdisciplinary research or medical applications) and fields dealing with three dimensional modeling (photogrammetry). The research presented in this paper has originally been motivated to provide more efficient data management and visualization for photogrammetry (for stereoscopic geo-data), however presented concepts, methods and results are potentially relevant to any stereoscopic media.

With a rekindled interest in 3D cinema and with the developments in digital devices including 3D-enabled television and mobile devices, efforts to make stereoscopic media more accessible -and stereoscopic content more comfortable to view- are potentially relevant to a wider audience today than before. One such effort, which is in this paper's scope, is to render displays in a space-variant manner similar to human visual perception to make them more efficient and potentially add cognitive benefits [8][10][14]. This approach can be viewed as a level of detail (LOD) management method for gaze-contingent displays

[2][3][5]. The results from an implementation demonstrates that by rendering the LOD as it is relevant to the viewer, we obtain *perceptually lossless* stereoscopic compression, help minimize resource requirements and avoid wasting valuable communication bandwidth.

2. SPACE VARIANT NATURE OF HUMAN VISION

Human visual system (HVS) has a space variant nature, *i.e.* the spatial resolution is not uniformly perceived by human eyes. This is readily verified when we look at the distribution of photoreceptors over human *fovea* (Figure 1). Fovea has two types of photoreceptors, *cones* and *rods*. Cones are very dense in the center, where there are almost no rods and rods are very dense in the periphery where there are very few cones [4].

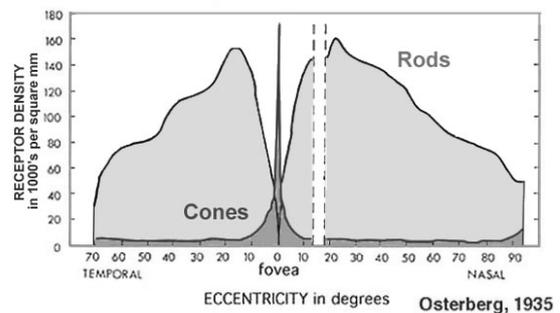


Figure 1. Photoreceptor distribution in fovea [11]. The dotted line to the right of the center represents the blind spot.

Cones capture high resolution and color information while rods record lower resolution and monochromatic information [1]. Thus as we progress towards the periphery from the central foveal area, the quality of vision will decline both in terms of resolution and color.

It is also apparent that our vision is not spatially uniform in 3D space. Once the eye accommodates on an object, the spatial resolution is lower for objects that are closer or further away from the accommodated point. Depth of field (DOF) is essentially a monocular concept, *i.e.* each eye has a "focusing" mechanism based on its lens system that allows certain amount of visual information to be coded in higher resolution than other areas. While DOF is monocular, for finer discrimination we use the input from

two eyes, and binocular (stereoscopic) depth perception is called *stereopsis*. Stereopsis is a feature in primates who have their eyes in front of their heads. It relies on a slightly shifted left and right input in two eyes and the binocular disparity that occurs on retina because of this slight shift. The brain fuses these two images into one, and geometric relationship between retinal disparity and distance to the perceived object allows the brain to process a more precise 3D perception up to a certain distance (exact distance of relevance is debated, commonly reported as 25-30 meters [3]). To manage the level of detail across the visual field, HVS typically integrates the information recorded by the photoreceptors in the fovea and the lens system in each eye with stereopsis.

3. FOVEATION FOR STEREOSCOPIC MEDIA

Foveation, as a term in image and video processing research, is used to refer to an approach where a display is rendered in a non-uniform fashion resembling the structure of the fovea. Doing this has several proven and implied advantages. Computationally, foveation provides a significant compression, and therefore offers a better resource and bandwidth management [6][7][8]. In an ideal case where we can assess the viewer's point of interest successfully (e.g. via eye-tracking), the resulting non-uniform (foveated) image should correspond to what the eye would naturally perceive. Hence the compression is referred to as *perceptually lossless*, i.e. even though computationally speaking we dramatically reduce information in the periphery, this should be unnoticeable for the viewer [7][9][10][13].

Foveation has been utilized in two-dimensional image and video processing applications by a variety of researchers and media producers, however less exploited for three-dimensional space. It is possible to extend the approach to stereoscopic images, videos and other media based on the stereoscopic acuity with potentially added benefits both in human and machine performance.

3.1. An Example Implementation of Foveation for Stereoscopic Imagery

In this section we present an implementation that allows us to test the argument that foveation for stereoscopic visualization can offer computational benefits. The implementation is named *Foveaglyph*, based on words fovea and anaglyph. It takes a stereoscopic image pair as input, performs image matching to obtain a disparity map, creates an image pyramid, calculates 3D information and provides a foveated anaglyph image as an output based on a point of interest (POI) that is interactively selected by the user (Figure 2).

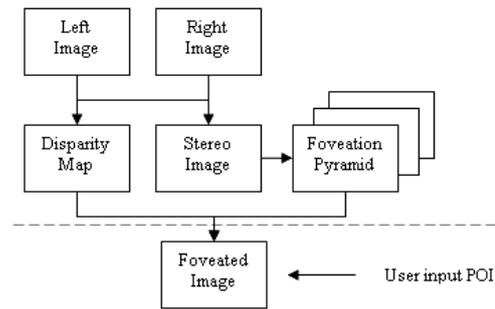


Figure 2. Foveaglyph's flow of processes.

3.1.1 Model

For testing and demonstration purposes and for computational simplicity we use an approximate model (Figure 3). For each pixel in the foveated image, the pixel's LOD is determined based on its distance from the POI. The LOD will decrease as the distance from POI increases. Once the user specifies the point of interest, the distance d is calculated between the POI and each pixel visited [2][3]

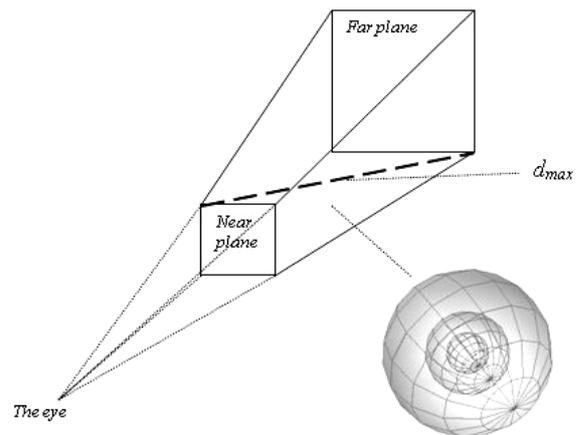


Figure 3. A simplified representation to multi-resolution 3D space.

By default, *Foveaglyph* uses a step function, which switches the LOD at a threshold D . This threshold is determined based on the maximum possible distance in the image setup, and the number of levels available in the image pyramid. If formulated, D is as follows:

$$D = \frac{d_{max}}{L}$$

Where d_{max} is maximum possible distance in the working space and L is the number of available levels in the image pyramid. In this model (***Figure 3) near plane and far plane are determined by the maximum and minimum disparity values. Varying resolution volume rings (concentric spheres) are constructed inside this pyramid which is

the bounding box. In partitioning the depth space, the disparity values are used as indicators of changing depth. But the fact that the disparity of the closer objects are bigger, defines finer depth intervals for near viewing space. The z value is calculated based on disparity information and is used in combination with image x, y coordinates as part of the equation to calculate the distance in pixels. A lod , when printed in lowercase italic letters, of the (0 to L-1) levels in the image pyramid (Figure 4), is the index number of each level (i.e. 0th level is the best resolution, and would be expressed as $lod = 0$). For each pixel it is calculated using the following function:

$$lod = \frac{dL}{d_{max}}$$

This is a step function (e.g. lod is an integer) where d is the distance between the POI and the pixel to be determined, lod is the pixel's LOD (resolution). This causes a linear decrease of quality in steps towards the periphery. The scale factor determining the downscaling ratio used between subsequent images in the image pyramid determines the reduction of quality in each step.

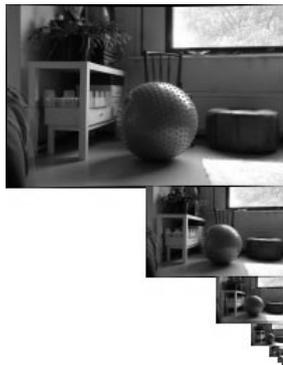


Figure 4: Image pyramid for eight levels of detail.

3.1.3 Results

Foveaglyph allows for 2D foveation as well as 3D for purposes of comparison (Figure 5). The compression provided by *Foveaglyph* is not at a fixed rate. This is because of several variables starting with the user-driven parameters in the program, such as how many levels of detail is desired, what the scale ratio should be and what kind of LOD function should be used (currently only distance LOD available). Also the location of the POI will change the amount of compression, i.e. if the POI is located towards the periphery rather than the center, the highest resolution areas occupy less of the image and the resulting

image has fewer pixels. Hence the resulting image size will be smaller.

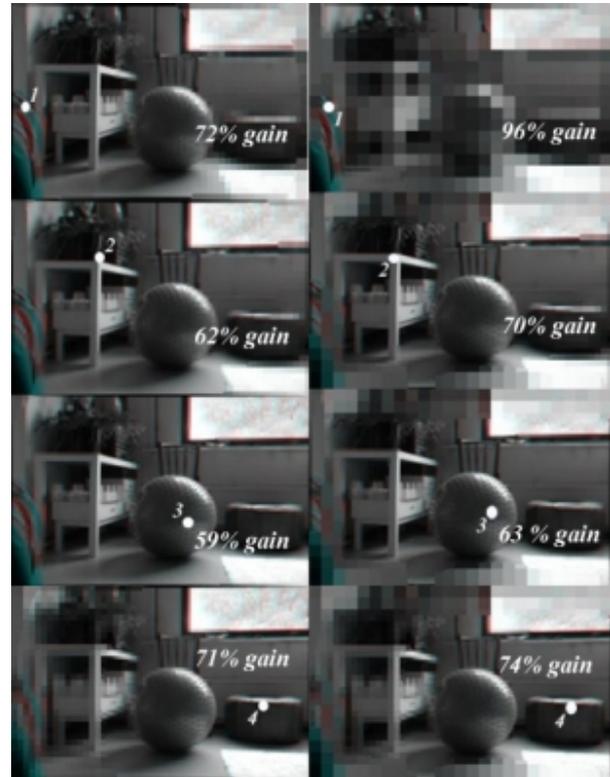


Figure 5. Compression results with four different points of interest, labeled with numbers.

The compression gain has been calculated based *effective pixel count*, which is a measure for determining which pixel comes from which level, and the total number of each. It should be also noted that the results in the above image (Figure 5) are particularly exaggerated to demonstrate the foveation effect (e.g. one can see the pixels getting coarser for the closer objects in 3D). It is easily possible to apply a smoothing filter for the resultant images to appear visually more pleasing.

4. CONCLUSIONS AND DISCUSSION

We argued and demonstrated that stereoscopic foveation offers computational advantages. Several questions remain at this point – some of which we will try and elaborate briefly. One question is whether this process is suitable for real time applications or not. If we can assume that the image matching and image pyramid are processed off-line, as it is in our implementation, foveation itself offers plausible times for real time processing (0.3 seconds for 781x512 image size, and 4.6 seconds for a 3137x2084 image). These will be even faster if the processes are handled in the graphics processing unit (GPU) instead of the CPU [5].

A second point to consider is that gaze-contingency paradigm (where this approach would be most useful) typically requires that we can determine where a person looks. Therefore, limits the potential to a single user scenario. This is likely to be acceptable for many desktop applications, and personal head-mounted viewing systems. It is possible to allow the viewer to use a pointer (instead of tracking the eyes). Parkhurst & Neibur [10] report that it is not really necessary to have an eye tracking system. However, this is still complicated in multi-viewer scenarios and a great benefit of doing this sort of work would be for large screens and in this case we need to consider multiple viewers at the same time. One approach for this could be foveating the scene based on saliency of objects [8].

While we demonstrated that there is computational gain, another motivator for this work has been that this solution is *perceptually lossless*. There are some user studies which have findings that support this [9][13], however more user studies are necessary to establish proper behavioral evaluation. Another potential benefit of simulating the depth of field along with foveation is that it may be helpful in situations where virtual simulator sickness occurs [8][14]. This claim is also based on literature and has not been sufficiently tested.

Human visual system has a level of detail management that is seamless and fast. Understanding mechanisms involved in HVS that is to do with spatial and temporal resolution can lead to computational models of vision and visualization for the benefit of an interdisciplinary community and audience.

11. REFERENCES

- [1] Ahnelt, P. K., Kolb, H. and Pflug, R., 1987. Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. *J. Comp. Neurol.*, 255: 18-34.
- [2] Coltekin A., Haggrén, H., 2006. Stereo foveation. *The Photogrammetric Journal of Finland*, 20 (1): 45-54.
- [3] Coltekin A. 2006. *Foveation for 3D Visualization and Stereo Imaging*. Ph.D. thesis, Helsinki University of Technology, Helsinki, Finland. ISBN 951-22-8017-5, URL: <http://lib.tkk.fi/Diss/2006/isbn9512280175/>.
- [4] Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E. and Kalina, R. E., 1987. Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 236: 579-582.
- [5] Duchowski, A. T. and Çoltekin, A., 2007. Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging. *ACM Trans. Multimedia Computing and Communication Applications*, 3(4):1-18. DOI= 10.1145/1314303.1314309.
- [6] Geisler, W.S., J.S. Perry, 1999. Variable-resolution displays for visual communication and simulation. *The Society for Information Display*, 30:420-423.
- [7] Kortum, P., Geisler, W. 1996. Implementation of a foveated image coding system for image bandwidth reduction. In: *Human Vision and Electronic Imaging, SPIE Proceedings*, 2657:350-360.
- [8] Linde, van der, I., 2003. *Space-variant Perceptual Image Compression for Gaze-contingent Stereoscopic Displays*, PhD Thesis, Faculty of Science and Technology, Department of Computing, APU. Jan 2003
- [9] Parkhurst, D., Niebur, E., 2004. A feasibility test for perceptually adaptive level of detail rendering on desktop systems" In: *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* Los Angeles, California, Pages: 49 - 56 ISBN: 1-58113-914-4.
- [10] Parkhurst, D. J. and Neibur, E. 2002. Variable resolution displays: A theoretical, practical and behavioral evaluation. *Human Factors* 44:611--629.
- [11] Osterberg, G., 1935. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica*, 6:1-103.
- [12] Wheatstone, C., 1838. On some remarkable, and hitherto unobserved, phenomena of binocular vision -- Part the first. *Philosophical Transactions of the Royal Society of London*, 1838: 371-94.
- [13] Watson, B., Walker, N., Hodges, L.F., 1996. Effectiveness of spatial level of detail degradation in the periphery of head-mounted displays. In: *Conference Companion on Human Factors in Computing Systems*. ACM.
- [14] Ware, C. 2004. *Information Visualization - Perception for Design. Second edition*. Elsevier Inc., Morgan Kaufmann Publishers. ISBN 1-55860-819-2.