Egocentric vision in a 3D game using linear perspective and natural rendering

Robert Pepperell[1]
rpepperell@cardiffmet.ac.uk
Nicole Ruta[1]
niruta@cardiffmet.ac.uk
Alistair Burleigh[1]
aburleigh@cardiffmet.ac.uk
Heddwyn Loudon[1]
heddwynl@gmail.com

[1] Fovography Synthetic Vision Research Centre, School of Art and Design, Cardiff Metropolitan University

Abstract

Humans naturally see the world from an egocentric perspective. But important questions remain about how best to represent this perspective in image media. Most computer graphics engines use linear perspective (LP) to render 3D space. But previous studies suggest that images created using LP do not accurately represent the experience of being in front of a space, particularly for wide fields of view (FOV) where apparent distortions of object size and shape can occur. We report two experiments in which we compared different methods of representing the egocentric viewpoint in a real-time 3D gaming environment. One method was the standard LP projection used by 3D graphics engines, and the other used a novel projective geometry developed by the authors that is based on the natural structure of human vision, called Natural Rendering (NR). Our results showed that participants: 1) preferred to play the game in NR at a wider FOV setting (M=142°) compared to LP (M=115°) and 2) were 4.9 times more likely to estimate a target object's correct distance and 5.12 times more likely to give an accurate distance estimation if they were looking at NR renders compared to the LP ones. We conclude that NR approximates the perceived structure of the egocentric viewpoint more closely than LP, providing a more immersive and natural experience of virtual environments, and that it might have benefits for 3D computer graphics more generally.

Introduction

Humans naturally see the world from an egocentric perspective or first-person point of view. This perspective covers a wide field of view, over 180 degrees horizontally, and includes a large area of peripheral visual space. Most mechanically generated images (those made with computer graphics software or cameras, for example) are created using linear perspective geometry [19]. One limitation of linear perspective (LP) is its inability to represent very wide fields of view without apparent distortion. Therefore, the natural egocentric viewpoint, which often includes features of our own bodies within the wider visual periphery, is conventionally cropped from photographs, paintings, and cinematic images. This deprives our image media of a fundamental feature of our visual experience, namely our own embodied position in space, and one that acts as a reference point from which we make spatial judgements about our relationship to the world [1]. As the perceptual psychologist James Gibson noted, “in order to localize any object there must be a point of reference” [18]. Consequently, images created in LP that omit the full egocentric perspective tend to lack a natural, embodied sense of first-person point of view, and this can lead to poor judgements about relative sizes and distances in the image [22].

In this paper we will consider some of the limitations of LP and why it is unable to effectively represent the egocentric perspective. We will then report experiments in which we investigated how some of the limitations of LP can be overcome by using a novel form of 3D geometry modelled on the phenomenological structure of human vision that we call Natural Rendering (NR). In particular, we were interested in whether a more natural egocentric perspective generated in NR would allow users to better judge the distances between the virtual egocentric perspective and virtual objects in a 3D computer graphics scene. Finally, we will discuss the implications of our findings in the context of rendering techniques for computer graphics and the representation of egocentric perspective in digital media.

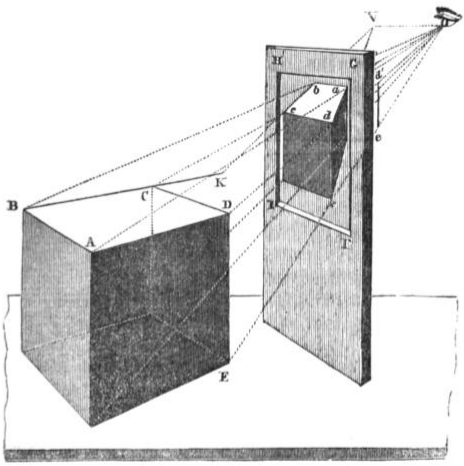

Figure 1. An illustration of the geometric principles underlying Alberti’s window, from Brook Taylor’s *Principles of Linear Perspective* (London, 1835). The viewer sees the 3D box as a 2D virtual object through a cropped window onto which the light rays from the box project through straight lines.

1. Linear Perspective and its Limitations

Linear perspective (LP) is the most commonly used projective geometry for mechanically rendering 3D space to 2D images. Many computer games, such as first person shooters, are played from an egocentric viewpoint and such views are represented using the inherent linear perspective geometry of real-time computer graphics engines such as Unity3D [17]. Central to the theory of LP is the notion of a fixed monocular point of view and of framing a rectangular portion of space, called 'Alberti's window' after the author of the first treatise on LP written by Leon Battista Alberti [2]. LP represents light rays coming from objects in the 3D space as straight lines passing and projecting to a flat picture plane as though one were looking through a virtual window into a space beyond (Figure 1).

LP has several limitations, which were known to the artists such as Leonardo da Vinci who first developed the technique [5, 20]. Despite the tradition of training artists in the methods of linear perspective in European art, these limitations meant that in practice they hardly ever applied its rules rigorously [4]. This is because artists are often primarily concerned with capturing and conveying the phenomenology of visual experience rather than strict optical accuracy, and this phenomenology is geometrically non-linear [21].

One of the major limitations of LP is its ineffectiveness for rendering wide FOVs (i.e. > 60 horizontal degrees) [4]. Wide FOV projections in LP suffer from apparent distortions of size and shape of objects, where marginal objects become stretched and central objects become minified. It is important to note that these are not distortions from linear perspective itself, as the images made with modern technology produce almost perfect rectilinear projections, but appear as distortions if the image is viewed from outside the correct centre of projection [13], which is where the disembodied eye is located in Figure 2.

The limitations of LP are also known to computer graphics engineers and users, particularly in the gaming community [6, 7, 8, 9] and as a consequence conventional 3D video games engines tend to restrict users to relatively narrow FOVs; games running on desktop PCs will typically have a FOV of approximately 90 horizontal degrees while games running on consoles typically have a FOV of 60 degrees. Representing the egocentric perspective of a character using such relatively narrow FOVs results in a view of the virtual world seen as if looking through a letterbox or scuba mask since much of the peripheral visual space is inevitably cropped. This means that key depth cues can be occluded, apparent size and distance relationships can be distorted, and camera movements can be more frequent and extreme as users move about to explore a space, which can result in motion sickness or nausea [10].

As a result of these limitations, users in virtual environments such as computer games with narrow FOVs can suffer in terms of playability and task performance. For example, it is more difficult for the players to resolve minified items of interest in the centre of the screen or to detect objects in a far periphery that is cropped. In sum, the limitations of LP contribute to an unnatural representation of the egocentric perspective and an impoverished experience.

2. Natural Perspective and Natural Rendering

The natural human visual system has many characteristics that are not replicated by standard computer graphics rendering techniques based on LP. For example, human vision is normally binocular, while screen-based renders are monocular. Human vision is dynamic; we constantly move our eyes to scan our surroundings and change the accommodation of the eyes as distances of viewed objects change. The linear perspectival viewpoint is generally fixed and lacking dynamic accommodation. We also rely extensively on peripheral vision to navigate the world, even if we are not always aware of it [11]. And the entirety of the binocular human visual field extends up to ~ 180 degrees of FOV, which is a span of visual space that is impossible to render using LP techniques [12].

It has been argued that humans see the world using Natural Perspective rather than linear perspective and there is a long history in art practice of attempts to represent this natural perspective in painting and drawing [3, 20]. Natural perspectives tend to capture the phenomenological structure of human vision, including the egocentric viewpoint, rather than accurately replicating the projection of light rays, as LP does.

The authors have developed a real-time computer graphics rendering technique based on natural perspective that reproduces key characteristics of human vision called Natural Rendering (NR). NR is a computational method of visually representing space in a non-linear way using a flexible projection pipeline [3]. The process dispenses with the standard linear perspective projection matrix approach employed in most 3D engines and instead applies a series of alternate geometric transformations that distribute the initial vertices of a 3D scene into a final viewing space in a way that corresponds more closely to the perceived structure of the human visual field, as measured using psychophysical methods.

Typically, this distribution results in an image in which the central region is significantly magnified and the peripheral regions are significantly minified as compared to a standard linear perspective projection. When compared to conventional wide-angle curvilinear rendering methods such as Fisheye effect or Panini post processing techniques the NR distribution exhibits less apparent curvature. The effect is more pronounced at wider fields of view, and the process is generally applied to render fields of view of > 120 degrees horizontally. This method offers the possibility of visualising wide FOV images that better represent the experience of being in front of a real scene [4]. NR has also the benefit of offering real-time calibration and personalization, allowing the user to directly manipulate 3D image space in new ways, and to select their preferred playing settings with more options than are offered by conventional linear rendering.

Previous research we have conducted showed that participants prefer wide-FOV images rendered in NR compared to LP and fisheye renders of the same space, and that NR images were judged as more ecologically valid and rated as having higher spatial validity than other standard image projections [3]. From this we concluded that NR images more closely match the egocentric view of the world than LP images or other standard non-linear projections.

In the present study, we wanted to investigate which FOV people preferred to use in a computer game when given the option to calibrate it in either LP or NR projections. We also hypothesized that they would be better at estimating target distances of objects in the NR projection compared to LP due to the more natural representation of the egocentric perspective obtained in NR. To test our hypotheses, we conducted two experiments using a customised version of a freely-available 3D game in Unity3D called Hammer 2 [14] to which we added an option to render the space using the NR projection technique.

Figure 2. An illustration of the natural rendering (NR) technique developed by the authors. The image on the left shows as wide-angle view (approximately 120 degrees) of the interior of a car rendered in linear perspective. Note the marginal distortions and the minification of the dashboard area. The image on the right shows the same view, but in natural rendering. Note the lack of marginal distortion and the enlarged dashboard area in the natural rendering version. Image credit: Shutterstock license.

3. Experiment 1: Identifying preferred FOV and Rendering

The aim of the first experiment was to identify the preferred FOV and preferred image rendering in the context of a 3D gaming environment [14]. The experimental setup of 21” computer screen displaying the game, and an interactive controller containing a dial allowing participants to directly manipulate the image geometry. The controller was setup such that it started at a value of 0.5, which corresponded to the default FOV setting of 90 degrees in LP. When turned anticlockwise to a value of 0, the image stayed in LP but the FOV could be increased up to 179 degrees; when turned clockwise to a value of 1, the image changed from LP to NR up to 180 degrees of FOV.

Participants were volunteers from student and staff population of Cardiff Metropolitan University (N=20; 78% females; average age = 23.3) and represented the demographics of that population. They were asked to play the game for a few minutes and adjust the dial, if they wanted to, in order to find the settings with which they were most comfortable playing with. Our results showed that 70% of the participants preferred to use NR over LP. The average horizontal FOV settings for NR was 142 degrees and 115 degrees of FOV for LP. These results suggest confirms our previous findings that people, when given the choice, prefer images with wide FOVs, and that NR is a better representation the egocentric perspective than LP.

Figure 3. A comparison between two different renderings of the same 3D computer graphic scene showing an egocentric perspective. The left image shows the scene rendered in linear

perspective, and at the 115 degrees field of view chosen by participants as the preferred setting. The image on the right shows the same scene, but in the natural rendering projection, and at the 142 degrees field of view users chose as the preferred setting in this projection.

4. Experiment 2: Estimating Distance of a Target Object

The purpose of the second experiment was to investigate if people are better at estimating the distance of a target object located in the central area of the scene while looking at NR compared to LP projections of the virtual space. To do this we conducted an online study using Prolific platform (https://prolific.ac/) to recruit participants and distribute our survey, which was developed in Qualtrics (https://www.qualtrics.com/uk/). Note that this was a different sample of participants from those in the previous experiment, and more demographically representative of the population at large.

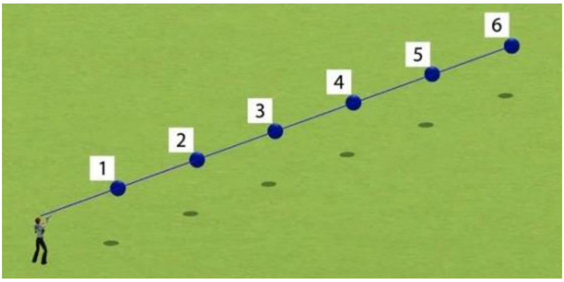

Figure 4. The image above shows the reference diagram participants saw before completing the training task.

Participants (N=52; 61.5% females; average age = 18) were asked to estimate the perceived distance of a target object, a blue ball, in a set of 72 still images captured from different locations in the virtual environment. In each image the target ball was presented in one of 6 fixed locations, each equally distant from the egocentric viewpoint of the avatar. Before the experiment participants were shown an image that illustrated the different possible positions of the target object in the virtual space (Figure 4). The stimulus set consisted of two images for each of six different locations in the virtual scene, giving a total of 12 scenes, each rendered at three different FOVs (100, 120 and 140 horizontal degrees) in both LP and NR projections. Participants also had to complete a training session, during which they estimated the distance of the target object presented in two selected locations (2 and 6), respectively at 100 and at 120 degrees of FOV, using both rendering techniques (see Figure 5). The images were presented in a random order in both the training and the experimental session.

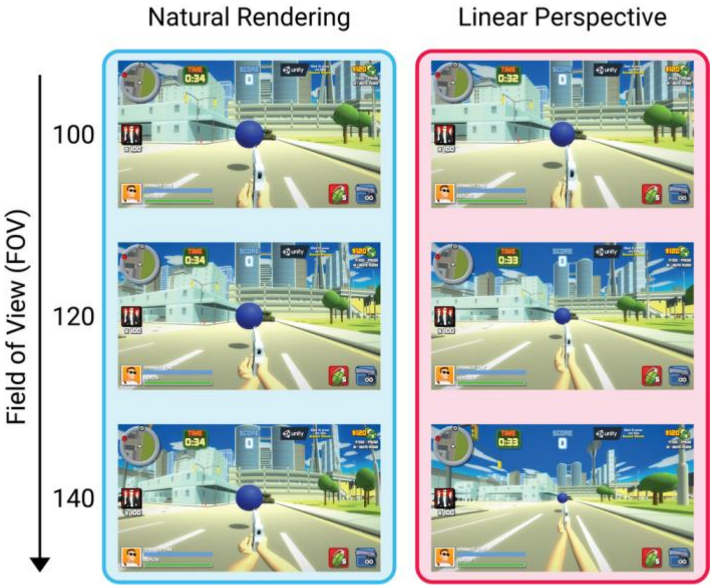

Figure 5. The image above illustrates both NR and LP images, for the three selected FOVs used in Experiment 2.

We used Rstudio [15] and lme4 package [16] to perform a linear mixed effects analysis to analyse participants' responses, with image rendering (NR vs LP) and FOV (100 vs 120 vs 140) as fixed effects. As nested random effects we had the intercepts for participants' identity and distance shown in the image (2 sets of 6 distances); as random effect we had image identity (N=72). Our results showed a main effect of image rendering, meaning that, overall, participants were 5.32 times more likely to estimate the correct distance and 4.07 times more likely to be accurate when looking at NR compared to LP images. Gender had no significant effect on the results.

Conclusions

The experiments reported here show that when given the option a majority of the people tested chose to play a computer game using NR rather than LP, and that they preferred to play the game in a very high FOV setting in NR. Moreover, we found that people are better estimating distances in the virtual space rendered in NR compared to LP. These results suggest that NR provides a better representation of the natural human visual field as experienced from the egocentric perspective.

Applying NR techniques more widely in 3D computer graphics might have significant benefits, beyond those specific to gaming. For example, computer generated architectural visualisations attempt to convey how a built space would appear to a user and these might be significantly improved by the use of a more valid egocentric perspective. Applications used for simulation, training or medical intervention might also be improved for the same reasons, particularly where they involve judgements about size and distance and interactions with virtual objects and spaces [23].

Our future research will further investigate the opportunities for enhanced calibration and personalization of image gometries in virtual environments in order to improve user experience. The long-term aim of NR is to develop real-time technology that matches as closely as possible the perceived structure of visual space, as seen from the natural egocentric perspective.

References

[1] Pepperell, R. (2015). Egocentric perspective: Depicting the body from its own point of view. *Leonardo*, 48(5), 424-429.

[2] Alberti L.B. (1435). *On Painting* (Translated by Cecil Grayson, 2004) Penguin Classics, London.

[3] Burleigh, A., Pepperell, R. and Ruta, N. (2018) 'Fovography: A naturalistic imaging media', In 3D Immersion (IC3D), 2017 International Conference on 3D Immersion (IC3D). Brussels, 11-12 December, IEEE.

[4] Pepperell, R., & Haertel, M. (2014). Do artists use linear perspective to depict visual space?. *Perception*, *43*(5), 395-416.

[5] Prienne, M.H. (1970). *Optics, Painting, and Photography*. Cambridge University Press.

[6] Taylor, L. (2002). *Video Games: Perspective, Point-of-view, and Immersion*. [ebook] Florida: University of Florida. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.197.2375 [Accessed 4 Oct. 2018].

[7] Valve Software. (2012). Field of View. [online] Available at: https://developer.valvesoftware.com/wiki/Field_of_View [Accessed 30 Jan. 2019]

[8] Temme, L. (2007). Linear perspective limitations on virtual reality and realistic displays. *Display Technologies and Applications for Defense, Security, and Avionics*. [online] Available at: https://www.researchgate.net/publication/253735242_Linear_perspective_limitations_on_virtual_reality_realistic_displays_-_art_no_65580K [Accessed 4 Oct. 2018].

[9] Smallman, H.S., St. John, M. (2005). Naïve Realism: Misplaced faith in realistic displays. *Ergonomics in Design*, Summer Issue: 6-13.

[10] Valve Software. (2012). Field of View. [online] Available at: https://developer.valvesoftware.com/wiki/Field_of_View [Accessed 30 Jan. 2019].

[11] McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision research*, *24*(1), 25-32.

[12] Howard, I., & Rogers, B. (1995). *Binocular Vision and Stereopsis*. Oxford, UK: Oxford University Press.

[13] Koenderink, J. (February, 2011). Vision as a user interface. Human vision and electronic imaging XVI. Vol. 7865, p. 786504. International Society for Optics and Photonics.

[14] Full Game Kit - Hammer 2. Available at: https://assetstore.unity.com/packages/templates/full-game-kit-hammer-2-43670 [Accessed Jan. 2019].

[15] RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

[16] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

[17] Poole, S. (2000). *Trigger happy*. New York: Arcade Pub.

[18] Gibson, J. J. (1950). *The Perception of the Visual World*. Cambridge, Mass.: The Riverside Press, p. 226.

[19] Kingslake. R, (1992). *Optics in Photography*. Bellingham: SPIE Press, 1992. [13] O. Kenneth, *Researches in Binocular Vision*, 2nd ed. New York: Hafner.

[20] Kemp, M. (1990). *The Science of Art: Optical Themes in Western Art from Brunelleschi to Seurat*. New Haven, USA: Yale University Press.

[21] Pepperell, R. (2017). Imaging Human Vision: An Artistic Perspective, in *Human Vision and Electronic Imaging 2017*, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Huib de Ridder, Proceedings of SPIE-IS&T Electronic Imaging.

[22] Ljung, K. (2017). *Effects of Field-of-View in First-Person Video Games: A Study on Camera Field-of-View in Relation to Game Design*. Undergraduate. Uppsala.

[23] Condino, S., Carbone, M., Piazza, R., Ferrari, M., & Ferrari, V. (2019). Perceptual Limits of Optical See-Through Visors for Augmented Reality Guidance of Manual Tasks. *IEEE Transactions on Biomedical Engineering*.