Towards mixed reality applications on light-field displays

Citation
Balaz, A., Barsi, A., Kovacs, P. T., & Balogh, T. (2014). Towards mixed reality applications on light-field displays. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), July 2-4,2014, Budabest, Hungary* (pp. 1-4). Piscataway: Institute of Electrical and Electronics Engineers IEEE. https://doi.org/10.1109/3dtv.2014.6874717

Year
2014

Version
Peer reviewed version (post-print)

Link to publication
TUTCRIS Portal (http://www.tut.fi/tutcris)

Published in
3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), July 2-4,2014, Budabest, Hungary

DOI
10.1109/3dtv.2014.6874717

TOWARDS MIXED REALITY APPLICATIONS ON LIGHT-FIELD DISPLAYS

Ákos Balázs[1]*, Attila Barsi*[1]*, Péter Tamás Kovács*[1,2]*, Tibor Balogh*[1]

[1]Holografika Ltd, Baross u. 3. H-1192 Budapest, Hungary, [2]Department of Signal Processing, Tampere University of Technology, Finland

ABSTRACT

Mixed reality applications have been traditionally developed for stereoscopic or auto-stereoscopic displays scaling from AR glasses to Powerwalls and CAVEs. These approaches however, all suffer from the inherent shortcomings of the underlying technology such as limitations of stereoscopic rendering, tracking latency, low resolution and potentially uncomfortable wearable components. In this paper we present a novel rendering method that overcomes these limitations by augmenting the virtual images rendered on light-field displays. Our ultimate aim is to blur the line between real and virtual as much as possible and have a gadgetless, completely believable experience where the ambient environment is able to seamlessly augment the rendered image on the screen. Our approach makes it possible for the ambient environment to augment the virtual content rendered on the light-field display in real-time and in a view-dependent way which was not possible with earlier techniques.

Index Terms — light-field display, virtual environments, mixed reality, augmented virtuality, augmented reality

1. INTRODUCTION

Traditionally mixed reality applications were developed for various stereoscopic or auto-stereoscopic displays. These displays greatly vary in size and capabilities from head-mounted displays (HMDs) [1] to Powerwalls and CAVEs [5]. One type of HMDs is augmented reality glasses such as Google Glass [21]. These however, usually don't support stereoscopic rendering and are more geared towards presenting extra information and notifications unobtrusively to the user. Consequently such devices have very limited rendering capabilities (they typically use mobile CPUs and GPUs with limited amounts of memory) and low resolution which makes them unsuitable for those mixed reality applications which require high-end visualization. More traditional HMDs are geared towards head-mounted, immersive virtual reality. These immersive devices could theoretically work well in a mixed reality setting as their resolution is usually higher than that of the AR glasses and the multiple viewpoint problem may be solved by using multiple units for the different users with the content synchronized. However, these devices also suffer from various problems such as relatively low resolution (compared e.g. to Powerwalls or light-field displays), tracking latency and imprecision, it's uncomfortable to wear them for longer periods of time and they can cause motion sickness in some cases. Furthermore since the user cannot see the ambient environment at all, it makes mixed reality applications unnatural. There are developments to tackle some of these disadvantages (e.g. the Oculus Rift VR headset has high resolution and wide field of view and the NVIDIA Near-Eye Light-Field Display [6] tries to be as lightweight and easy to wear as possible) but these are not yet available commercially and don't address all of the shortcomings mentioned above.

In contrast to HMDs, Powerwalls [3][4] and CAVEs [5] achieve immersion by placing large, fixed projection screens away from the viewer. On these screens usually stereoscopic images are projected; stereo separation for the users is done by wearing active (shutter) or passive (polarization) glasses. These glasses are much lighter and a lot less intrusive than HMDs, but wearing them over a long time can still be uncomfortable. Even though such systems are typically used by multiple users simultaneously, usually only one user is tracked via head-tracking and images are generated from that user's viewpoint. If the other users stand “close enough” to the tracked user, they are still able to view the content from an almost correct perspective. Another advantage of projection based systems is that users can communicate naturally with each other and are able to maintain visual contact as well as observe each other's gestures. In projection systems, head movements such as rotation moves the projection point only slightly whereas in HMDs the whole world is moved around the user, therefore projection-based systems are a lot less sensitive to tracking errors and latency issues. Moreover as projection-based systems are not completely immersive, users can see real-world objects so it is not necessary to model input devices (usually some kind of tracked input device is used) which makes mixed reality applications look and feel more natural.

Light-field displays [6][7][8] offer the same advantages as projection-based stereoscopic displays but overcome their limitations: the users don't have to wear glasses and since they offer continuous view parallax (at least horizontally) tracking is not necessary as all users view the display content with correct perspective. This makes light-field displays well suited for developing mixed reality applications.

In this paper we present a rendering method for mixed reality applications for light-field displays and argue that mixed reality applications are more natural on light-field displays as they don't suffer from the drawbacks detailed above. We will focus on HoloVizio [8] displays but our method should be general enough to be usable on all light-field displays that are driven by a distributed rendering cluster.

2. RELATED WORK

Milgram’s Virtuality Continuum [9] is a good classification of different mixed reality settings. It defines a range where mixed reality extends from completely real environments to completely virtual environments (see Figure 1). On one side of the virtuality continuum are real environments, on the opposite side are virtual environments while augmented reality (AR) and augmented virtuality (AV) applications reside somewhere in the middle. AR applications typically build upon a real environment (either through a video camera or live e.g. via see-through glasses) and add virtual content to this environment, e.g. additional information which is not visible in the real world.

Figure 1: Virtuality Continuum as defined in [9]

A recent typical example of such systems is Google Glass. For a more detailed survey of current AR applications we refer to [10].

Further along the virtual direction are augmented virtuality applications, where a virtual scene is augmented with information coming from the real environment. AV applications have commonly been used for surveillance or telepresence.

For surveillance scenarios the virtual environments are usually augmented by adding textured quads at suitable locations and rendering the video sources on these [13]. Telepresence applications employ a similar approach, although the exact relation between the augmentation and the virtual scene may be more complicated [14][15].

While light-field displays overcome many of the limitations discussed above, they come with their own set of challenges. On the one hand, the content must be rendered from many viewpoints [8] which potentially requires a rendering cluster in case of high-resolution displays. On the other hand, producing either pre-recorded or real-time content suitable for such displays is also not straightforward [17].

3. RENDERING METHODS FOR LIGHT-FIELD DISPLAYS

We call the radiance at a point in a given direction a light-field. This definition is equivalent to the plenoptic function [16]. In the case of light-field displays, we have no occluders, thus we can reduce the dimensionality of the light-field to 4D. We can represent the light-field as lines parameterized by their intersections with two planes in arbitrary positions. This representation is called a light slab [11].

For horizontal parallax only displays, we can have further useful simplifications (see Figure 2). We can replace the *uv* plane with a line parameterized by *u*. We can choose the *st* plane to be the display screen and the *u* line to be the line where the observer moves horizontally. From now on, we will refer to the *u* line as the *observer line* and the *st* plane as the display screen plane.

We can create a right-handed 3D Cartesian coordinate system with the origin display screen, the x axis pointing to the right edge of the screen, the y axis pointing up and the z axis being the normal vector of the screen surface in the center. Due to the horizontal parallax, the display's pixels are diffused vertically. Thus if the observer line is chosen to be parallel to the x axis, the z and y positions of the observer line from the display screen can be chosen arbitrarily.

However the choice of z, i.e. the observer line-display distance will influence the resulting vertical field of view (the further one goes from the display the smaller the vertical FOV becomes under which the display is visible). Depending on the light source type and virtual lighting model used this choice may also influence the lighting calculation while solving the rendering equation. In practice the observer line-display distance can be chosen as the value where the user will be able to observe the entire scene. Values greater than this distance have proven not to influence the light field image significantly.

This simplification allows us to precompute the intersection points and the direction of rays travelling from the pixels of the optical modules to the observer line. During rendering we account for all rays visible in the theoretical display model.

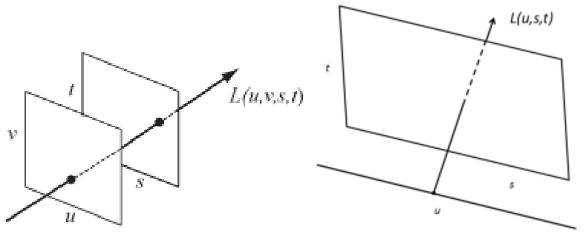

Figure 2: Light slab representation and horizontal parallax only light- field representation

The displacement between the theoretical and the physical rays is usually measured by a camera during the display calibration phase and is applied as a post processing step. The precomputed table of the theoretical display model is called the observer to screen ray lookup table. This consists of two parts: a single channel table containing x position values (as y is 0 and z is the observer line distance so both y and z are constant) for ray origins and a three channel table containing (x,y,z) ray directions.

There are two basic approaches for rendering images for light-field displays; both of which reconstruct the light-field of a scene for a given observer position as accurately as possible. The two methods are:

- capturing the light-field with cameras and reconstructing it for a light-field display [11]
- solving the rendering equation for a scene with surface, lighting and material representation [18]

Our approach merges the advantages of both methods for rendering realistic scenes while adhering to the constraints of real-time rendering. Mixing in a real-world video feed is also possible, which is necessary to achieve our goal of mixed reality rendering.

3.1 Capturing and reconstructing light-fields

For the capture of light-fields, there are several methods utilizing various types of camera arrays. There are several possible camera setups for example:

- linear camera arrays
- arc camera arrays
- matrix camera arrays
- HDR/LDR light probes [12]

Light-field reconstruction consists of selecting pixels from the original camera images that correspond to the rays of the display's light-field stored in the observer to screen ray lookup table. We define a region of interest (ROI) transformation between the visible light rays of the original scene defined by the capture arrangement and the display's light-field.

We are going to use a cube map generated from an HDR light probe [12] to provide real world environmental reflections for the scene. As the pixels of an environment map are treated as being at infinite distance from the scene, we could perform a lookup based on the observer to screen lookup table. However, due to the limited depth of light-field displays the theoretically correct solution would be very undersampled. To achieve an effect similar to the 2D rendering of environment maps, we render the environment onto a skybox geometry.

3.2 Solving the rendering equation

If a sufficiently accurate model of the scene is present one might give an approximate solution of the rendering equation [18].

For backlit horizontal parallax only displays, there are two basic approaches for rendering:

- ray-tracing the rays from the viewing line towards the scene

- using a linear approximation of the light field image by rasterizing the scene with per vertex transformation

The ray-tracing approach is not feasible on the current hardware configuration due to real-time constraints. Therefore we opted for using sufficiently tessellated models to avoid errors from the linear approximation.

In our current implementation, we use the Phong-Blinn lighting model for per pixel lighting of a scene. Directional lights can be calculated without the eye position lookup from the observer to screen ray lookup table. For rendering spotlights and point lights a single eye position lookup per pixel is necessary (since the y and z positions are constant only the x position needs to be looked up). If the object has only tangent space normal maps, the eye position is also used to transform the normal to world space correctly.

For all light types, hardware shadow maps are rendered at a resolution of 512x512 per light direction (1 for directional and spot lights, 6 for point lights), and applied with percentage closer filtering.

It is advantageous to do all lighting calculations in world space as otherwise we would need to transform all vectors necessary to eye space per pixel.

Both mixing the texels of a light probe with the warped camera input and using only the camera input as reflection map for the scene is possible. Reflections are calculated based on Schlick's approximation of Fresnel reflections [20]. This creates an effect of providing visual feedback of the user and / or the environment in the scene. That is, the lights of the real environment can lit the objects of the virtual scene, as well as the user can see his / her own reflections on highly reflective virtual surfaces in the virtual scene.

All textures (including the camera image texture) are used with gamma correction applied to them as this is necessary for the conversion between the sRGB and RGB color spaces. Such conversion is not necessary for HDR environment maps, as they are in linear space. Figure 3 show a typical 2D rendering pipeline as well as our light-field rendering pipeline for comparison. The differences to the 2D rendering pipeline are set in italic.

4. RESULTS

Our implementation runs on the HoloVizio HV80C system and uses its rendering cluster for distributed rendering. We use a single low-cost web camera for augmenting the virtual scene. By using the camera stream to provide reflections of the ambient physical scene on virtual objects we manage to create an immersive effect. Using a single camera stream is sufficient to provide real world interaction such as using a torch to influence the virtual lighting or to visualize real world reflections (e.g. a display cabinet in a museum setting). Due to the cluster based rendering the images produced by the web camera must be distributed over the network creating a significant network load, however, we still achieve interactive frame rates. Lighting and reflection mapping (including the camera texture) is calculated on a per pixel basis, so that the lighting is always correct from the viewer’s position and it seamlessly changes as the horizontal parallax changes when the viewer is moving in front of the display. This was not possible with previous approaches.

Figures 4 and 5 show how the ambient scene can be reflected on virtual models.

Both the material and the structure of the models influence the reflection: the fist model has a less reflective material and its structure scatters the camera texture while the second one has a more reflective material and a much simpler structure therefore the camera image is much more recognizable.

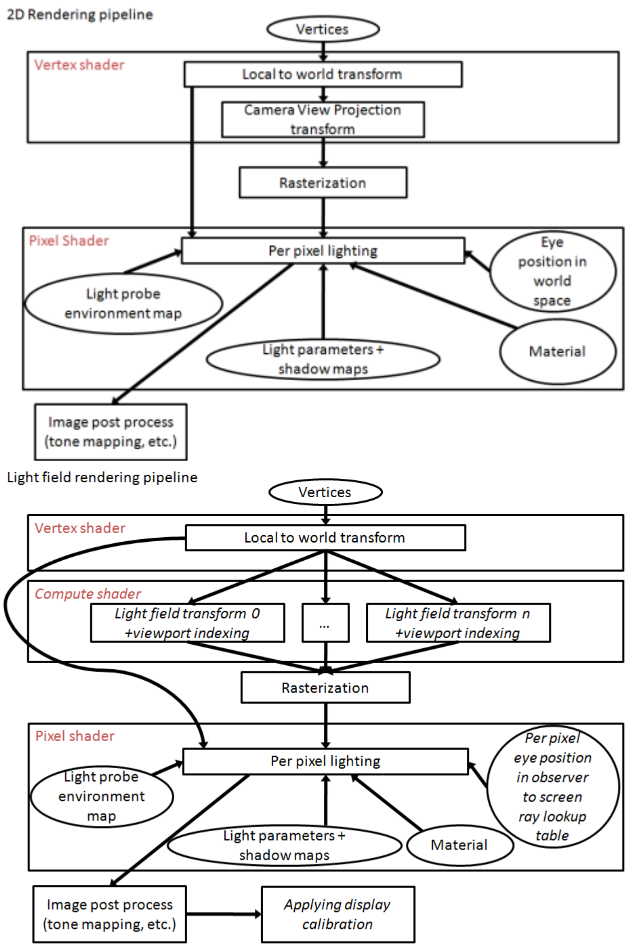

Figure 3: Comparison of a typical 2D rendering pipeline and a light-field rendering pipeline

Figure 4: Reflection of the ambient scene on a complex model

Figure 5: Reflection of the ambient scene on a smoother surface

5. CONCLUSIONS AND FUTURE WORK

We have presented a novel rendering method that is capable of displaying mixed reality content in real-time on light-field displays with per pixel lighting. We have demonstrated the viability of our method on HoloVizio type displays.

As future work we plan to experiment with using multiple cameras to better capture the ambient scene. By using multiple cameras it would be possible to capture ambient lighting and light effects direction selectively and to match the lighting of the virtual scene with the ambient lights precisely even if in the real scene complicated lighting effects are present.

Another future work direction would be to provide physically based lighting based on the incoming camera image(s). This could be done e.g. by calculating a Voronoi-diagram or the Penrose-tiling of the reflection map based on the radiance amount and using the resulting areas as directional lights to add correct lighting based on the camera image. Moreover, in order to provide further realism the surface reflection model could be changed from Phong BRDFs to Cook-Torrence BRDFs [20].

Likewise tone mapping the scene would also help immersion by simulating how the human eye adapts to varying light conditions. In order to achieve the correct effect we would need HDR cameras to capture the ambient scene.

6. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the PROLIGHT-IAPP Marie Curie Action of the People Programme of the European Union's Seventh Framework Programme, REA grant agreement 32449.

7. REFERENCES

[1] J. Rolland and H. Hua, "Head-mounted display systems", in *Encyclopedia of Optical Engineering* , pp. 1–13.

[2] P J. Olson, D. Krum, E. Suma, M. Bolas, "A design for a smartphone-based head mounted display", in *IEEE Virtual Reality*, pp. 233–234.

[3] P. Woodward, "Powerwall", http://www.lcse.umn.edu/research/powerwall/powerwall.html, 2001, [Accessed Apr. 23, 2014].

[4] M. Czernuszenko, D. Pape, D. Sandin, T. DeFanti, G. L. Dawe, M. D. Brown, "The ImmersaDesk and Infinity Wall projection-based virtual reality displays", in *ACM SIGGRAPH Computer Graphics*, 31(2), pp. 46-49, 1997

[5] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the cave", in *Proceedings of SIGGRAPH 93*, pp. 135–142, August 1993

[6] D. Lanman and D. Luebke, "Near-Eye Light Field Displays", in *ACM SIGGRAPH 2013 Emerging Technologies*, July 2013

[7] M. Kawakita, S. Iwasawa, R. Lopez-Gulliver, M. Makino, M. Chikama, M. Panahpour Tehrani, A. Ishikawa, N. Inoue, "Glasses-free 200-view 3D Video System for Highly Realistic Communication", in *Digital Holography and Three-Dimensional Imaging, OSA Technical Digest*, paper DM2A.1.

[8] T. Balogh, "The HoloVizio system", in *Proc. SPIE 6055, Stereoscopic Displays and Virtual Reality Systems XIII*, 60550U (January 27, 2006); doi:10.1117/12.650907

[9] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays", in *IEICE TRANSACTIONS on Information and Systems* 77.12, pp. 1321-1329., 1994

[10] D. W. F. van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations", in *International Journal of Virtual Reality* 9.2, pp. 1-19, 2010.

[11] M. Levoy and P. Hanrahan, "Light Field Rendering" in *Proceedings of SIGGRAPH '96*, pp. 31-42

[12] P. Debevec, "Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography", in *Proceedings of SIGGRAPH '98*, pp. 189-198

[13] O. Ripolles, J. Simo, R. Vivo, "Automatic Population of Scenarios with Augmented Virtuality", in *Computer and Information Sciences II*, pp. 347-353, 2012

[14] K. T. Simsarian and K. P. Åkesson, "Windows on the world: An example of augmented virtuality", 1997

[15] K. P. Åkesson and K. T. Simsarian, "Reality portals", in Proceedings of the ACM symposium on Virtual reality software and technology, pp. 11-18, 1999

[16] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision", in *M. Landy and J. A. Movshon (eds), Computational Models of Visual Processing*, 1991

[17] T. Balogh, P. T. Kovács, Z. Megyesi, A. Barsi, "HoloVizio - True 3D Display System", in *Proc. of 3DTV Conference*, 7-9 May 2007, pp. 1-4.

[18] J. T. Kajiya, "The rendering equation", in *ACM Siggraph Computer Graphics (Vol. 20, No. 4)*, pp. 143-150, 1986

[19] Schlick, C., "An Inexpensive BRDF Model for Physically-based Rendering", *Computer Graphics Forum 13,* pp 233-246, 1994.

[20] R. Cook and K. Torrance, "A reflectance model for computer graphics", *Computer Graphics (SIGGRAPH '81 Proceedings), Vol. 15, No. 3*, pp. 301–316, 1981.

[21] Google Glass, http://www.google.com/glass/start/ Accessed June 10, 2014