Real-time 3D light field transmission

Tibor Balogh, Péter Tamás Kovács
Holografika Kft, Baross u 3., Budapest, Hungary

ABSTRACT

Although capturing and displaying stereo 3D content is now commonplace, information-rich light-field video content capture, transmission and display are much more challenging, resulting in at least one order of magnitude increase in complexity even in the simplest cases. We present an end-to-end system capable of capturing and real-time displaying of high-quality light-field video content on various HoloVizio light-field displays, providing very high 3D image quality and continuous motion parallax. The system is compact in terms of number of computers, and provides superior image quality, resolution and frame rate compared to other published systems. To generate light-field content, we have built a camera system with a large number of cameras and connected them to PC computers. The cameras were in an evenly spaced linear arrangement. The capture PC was directly connected through a single gigabit Ethernet connection to the demonstration 3D display, supported by a PC computation cluster. For the task of dense light field displaying massively parallel reordering and filtering of the original camera images is required. We were utilizing both CPU and GPU threads for this task. On the GPU we do the light-field conversion and reordering, filtering and the YUV-RGB conversion. We use OpenGL 3.0 shaders and 2D texture arrays to have an easy access to individual camera images. A network-based synchronization scheme is used to present the final rendered images.

Keywords: Lightfield, Light-field, 3D video, HoloVizio, 3D capture, Rendering, Real-time, Camera array, 3D display

1. INTRODUCTION

Displaying live 3D imagery is a major step towards realistic visualization. Today most common solutions for 3D displaying are active or passive stereoscopic glasses, which are cheap and easily available. However, like all stereoscopic systems they can only provide 3D view for a single, fixed position (this is also true when multiple people are watching, as they all see the same image). Autostereoscopic displays can show different 3D images to multiple directions, but the most widespread displays (lenticular or parallax barrier systems) have limited light ray count (3D resolution). These can provide continuous view in a narrow FOV, however viewers moving may experience jumps when leaving or entering valid zones. HoloVizio light-field display technology is capable of providing 3D images featuring continuous motion parallax for a wide viewing zone for multiple viewers.

HoloVizio displays are capable of displaying high quality horizontal parallax light fields. The HoloVizio principle can also be extended to have both horizontal and vertical parallax, but this has not been demonstrated yet. As the information content of a light-field is very high, displaying large FOV light field videos is a challenging task. Our camera array consisting of 27 cameras (capturing 18 Million light rays) was connected a 10 MPixel HoloVizio system (HoloVizio 240P) and later a 30 MPixel HoloVizio system (HoloVizio 720RC). This shows the flexibility of our software system, as it can handle an arbitrary number of incoming and outgoing light rays in practically any camera and display configuration. The captured natural content was processed and displayed in real time. Storage and playback of captured 3D content is also possible with the system.

2. HOLOVIZIO TECHNOLOGY

The approach used by HoloVizio technology is quite different from that of stereoscopic, multiview, volumetric and holographic systems. It uses a specially arranged array of optical modules and a holographic screen. Each point of the holographic screen emits light beams of different color and intensity to the various directions. The light beams generated in the optical modules hit the screen points in various angles and the holographic screen makes the necessary optical transformation to compose these beams into a perfectly continuous 3D view, as shown in Figure 1. With proper software control, light beams leaving the pixels propagate in multiple directions, as if they were emitted from the points of 3D objects at fixed spatial locations[1,2].

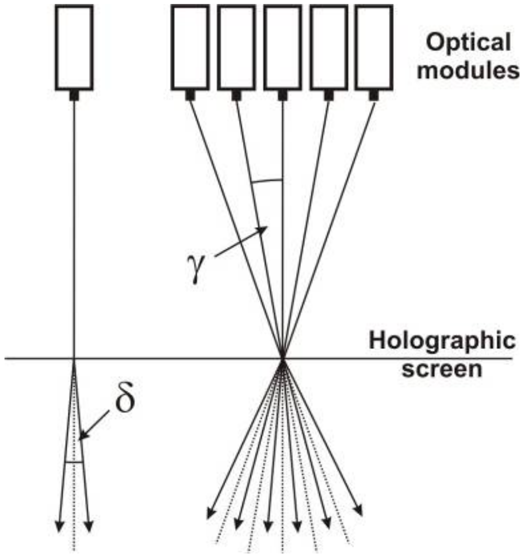
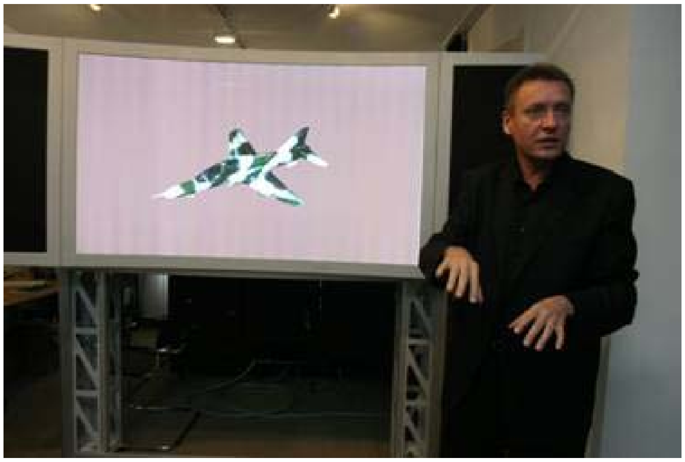


Figure 1. HoloVizio light-field generation principle. The light beams generated in the optical modules hit the holographic screen in various angles.

2.1 HoloVizio Displays

HoloVizio can be implemented in both small-scale and large-scale display systems. A 50 Mpixel large-scale system[3] has been developed with a screen diagonal above 1.8m. The display's optical system consists of compact projection modules arranged in horizontal rows. The system has a high angular resolution; approximately 50 independent light rays originate from each pixel. A PC-based render cluster provides 50Mpixels the display in real-time (using GPUs for most image generation tasks), a control system controls the projectors, PCs, the network, power supplies and monitors all system parameters.



Video 1. Large-scale HoloVizio light-field 3D display projecting 50Mpixels. http://dx.doi.org/doi.number.goes.here

Using HoloVizio technology is possible to build displays that have excellent image resolution of 1920x1080 or beyond, large FOV above 100 degrees, large Field-of-Depth, and at the same time the number of pixels being in the range of hundreds of millions, which demonstrate the scalability of the system very well. Being projection based, and using a high number of optical engines pointing towards the same screen, this technology will always dominate 3D display technologies based on flat screens in terms of pixel count by at least one order of magnitude. Our desktop 3D display is available in 32" size and features 10 Mpixels[4]. This model is in the dimensions of normal TV sets. The size in between is implemented in the HoloVizio 240P display, the first HoloVizio featuring a slim optical design, which, despite being a projection based display, allows it to be only 70 cm deep, and is controlled by 3 built-in PCs.

2.2 HoloVizio Software

There are several possibilities for displaying 3D data on the HoloVizio, the most important is interfacing interactive graphics applications to the holographic displays through the HoloVizio OpenGL wrapper. This library is able to display existing applications without any modification, recompiling, or relinking, thus users can continue using the applications they are used to, but now in 3D. A 3D converter and video player application is also provided, which can be used to create computer generated 3D videos for the display. An application development framework to render directly to the display is also under development, which allows users willing to develop HoloVizio-enabled applications to avoid using an intermediate library, and to use the rendering node's resources arbitrarily.

As clearly visible from the above, our existing software system focused on displaying synthetic content, which was the dominant use case of HoloVizio displays. This is sufficient for professional applications that are working with 3D data, like scientific visualization, engineering, prototyping, oil&gas exploration (see Figure 2) or digital signage. However, to target the public with a 3D display, displaying live 3D content is a necessity, constituting a major step towards 3D Cinema and later on 3DTV.

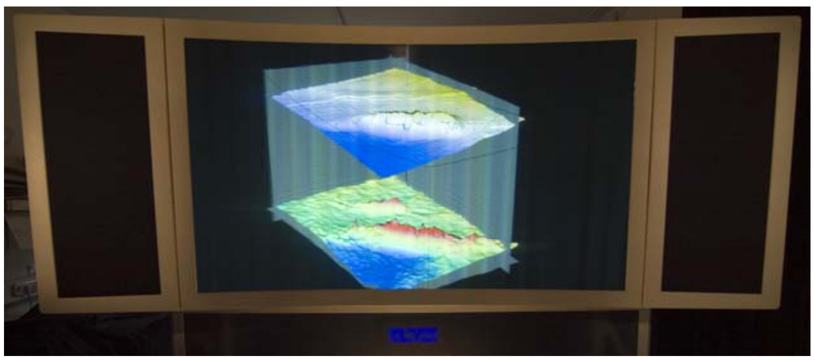


Figure 2. HoloVizio display running an oil&gas exploration application in real-time.

3. REAL-TIME LIGHT-FIELD CAPTURING

3.1 Light-field content

Not surprisingly, publicly available light-field content is not a common resource (some groups providing such content to the public are[6,7]). Using multi-view content results in suboptimal viewing experience on HoloVizios. As these are targeted for multi-view displays, usually very narrow FOV is used for capturing. Angular resolution is also lower than desirable, at least when compared to the capabilities of light-field displays, which provide both wide FOV, and fine angular resolution. The issue of angular resolution can be somewhat compensated by using Depth Based Rendering, Image Based Rendering, or a hybrid approach[8], but on the other hand, increasing the FOV is very challenging after a certain extent, and provides incorrect results. There are companies offering Time Freeze shooting (names may vary)[9], but these are headed towards the movie industry, where real-time acquisition, transmission and playback is not an objective.

3.2 State Of The Art

A number of papers describing real-time 3D video or light-field capture and display have been published in recent years, achieving significant advances. The random access light-field camera provides very good results with 2D or stereo displays due to selective transmission[10], but as the authors pointed out, this approach is less applicable with 3D displays which typically need access to all captured light-field data. On the other hand, as shown later, their observation still holds for light-field displays too, if the rendering process is distributed between a number of processing units. The TransCAIP system[11] uses a single PC and GPU algorithms to achieve interactive speeds with an impressive number of cameras, however, our system provides better results in terms of resolution, frame rate, and angular resolution. Moreover, the strength of that system is that everything is handled in a single GPU (avoiding frequent bus transfers), in contrast our system is designed to be highly scalable to be able to serve a number of cameras and 3D displays with very high light-ray count. The impressive MERL 3D TV System[12] uses a symmetrical system with a high number of PCs and a lenticular-lens based display to create live 3D visuals, however the HoloVizios we used have far better 3D image quality compared to their 3D display, moreover they use excessive number of PCs for capturing, processing and rendering: 16 cameras and 16 projectors are served by 8+8 PCs. In contrast, our system could easily serve such a configuration with using only 3 PCs.

3.3 3D acquisition system and calibration

Our camera system is made up from 27 USB CCD cameras. In the first stages, these were connected to 9 computers, but later on this was reduced to a single capture PC. The cameras are evenly spaced in a linear arrangement on a camera rig, each one capable of capturing at 640x480@15 FPS or 960x720@10 FPS resolutions (see Figure 3).

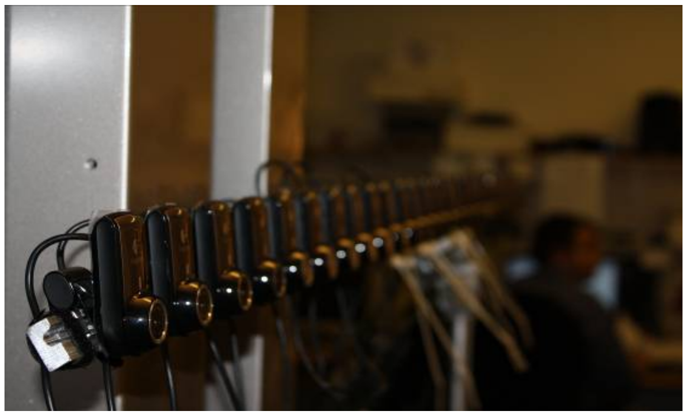


Figure 3. Camera array consisting of 27 CCD USB cameras. Together they capture up to 18MPixels.

The capture computer is directly connected to the demonstration 3D display (HV240RC) through a single gigabit Ethernet connection, supported by a 3 PC computation cluster, which is an integral part of the HoloVizio display. The camera system was calibrated off line using a semi-automatic calibration method, using images of a previously known reference object[14,15]. During the calibration, the parameters of the light field captured by the cameras are estimated based resulting in intrinsic and extrinsic camera calibration parameters. These estimates are further refined by a third refinement step to minimize the error of the estimated model, resulting in very good 3D image quality and field of depth.

The cameras can stream uncompressed YUV or MJPEG output out of which MJPEG has been chosen, as it can also be used for transmission over the network (although not very efficient).

4. REAL-TIME LIGHT-FIELD RENDERING

4.1 Incoming images

The original MJPEG images are arriving to the cluster's nodes on a single gigabit Ethernet channel with approx 10% link utilization. Each individual channel has its own CPU thread that decodes the Huffman encoding and does the inverse DCT algorithm for the incoming JPEG image. This yields a YUV image on the CPU. IDCT has also been implemented on the GPU, but that is not the real bottleneck. These images are then uploaded to GPU memory, where the light-field reordering takes place.

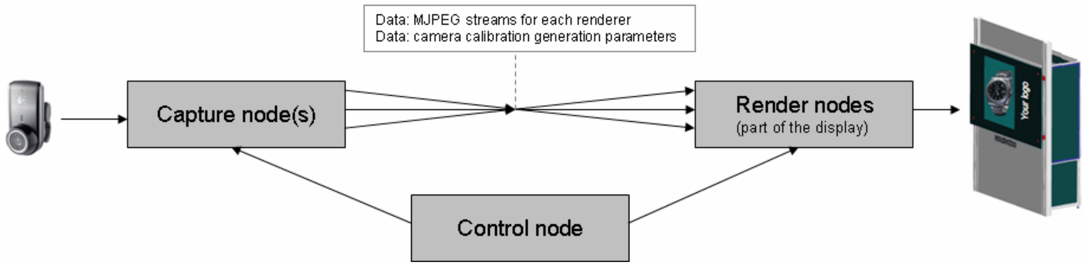


Figure 4. Architecture of the first generation light-field transmission system.

4.2 Light-field reordering and calibration

For the task of dense light field displaying a massively parallel reordering and filtering of the original camera images is required. On one hand, based on camera calibration information, we know which light rays of the scene are captured exactly. On the other hand, we know which light rays are emitted from the display, based on the arrangement of the optical engines and the display calibration information. Thus, we can derive which incoming light rays need to be used for the generation of each outgoing light ray. Once we have that mapping, the reordering of pixels can happen on the GPU very rapidly. On the GPU we do the light-field conversion and reordering, filtering and the YUV-RGB conversion. We use OpenGL 3.0 shaders and 2D texture arrays to have an easy access to individual camera images. This yields the correct bilinear filtering. The additional advantage is that display-specific calibration can also be handled in the same step. There is also a network-based synchronization scheme for displaying the final rendered images.

Video 2. HoloVizio 240P 3D display showing live 3D light-field content. http://dx.doi.org/doi.number.goes.here

5. RESULTS

Our 27 cameras light-field capture system streamed 960x720 resolution video streams with 10 FPS. The image data was converted on-the-fly to light field format and the continuous parallax 3D image stream was displayed on various HoloVizio displays. To our knowledge, this is the first system capable of displaying live 3D video with such quality.

A single PC was used for capturing 27 cameras, and 3 PCs for rendering (which are integral part of the HoloVizio display) were used. With this setup, we reached 15 frames per seconds playback speed for the 640x480 resolution stream and 10 frames per second for the 960x720 resolution stream. The bottleneck here was the camera acquisition speed.

6. FUTURE WORK

Although the bottleneck is camera acquisition speed now, using more and better cameras is desirable in the future. Once we do that, the two performance-critical points will be the Huffman decoding of the JPEG images and the upload speed to the GPU.

There are several ways to improve the solution to overcome these bottlenecks. We have observed that when rendering is distributed to multiple computers, none of them need all parts of all the images (not even when they serve multiple optical engines). Thus, the capture node could transmit only parts of the images needed, based on information received from the renderers. To do that, either uncompressed images, or a compressed image format that can be partially decoded should be used. The capture nodes could then transcode the MJPEG stream to this intermediate format, and transmit only parts requested from the renderers.

To transmit 3D light-field to longer distances, a more efficient compression approach is desirable. Even H.264[16] simulcast coding would help reducing the bandwidth requirement, but applying MultiView Coding (MVC[17]) could decrease the amount of transmitted image information even further.

Combining this two will result in a highly bandwidth efficient and future-proof system. A layered approach for using different encoding during transmission and light-field rendering is being developed, providing good compression on one part, and fast rendering on the other hand, allowing arbitrary number of cameras and very high resolution.

Such a system can be used to implement the most realistic 3D telepresence system ever made, which – being three dimensional and providing very fine angular resolution – also overcomes the problem of missing eye contact between participants [13].

Video 3. HoloVizio 720RC 3D display showing a telepresence situation. http://dx.doi.org/doi.number.goes.here

ACKNOWLEDGEMENTS

The development of the 10Mpixel test display system and camera system has been supported by EU IST-FP6 Integrated Project OSIRIS (IST-33799 IP)[5]. The FP6 project OSIRIS aims to create novel display systems including a high resolution LED based compact display that is capable of real-time playback of live captured natural content. In the same project HoloVizio technology is applied to create a 3D Cinema application.

REFERENCES

[1] Balogh, T., “Method and apparatus for displaying three-dimensional images,” U.S. Patent 6,201,565, EP 0900501, Feb 04, 1997.

[2] Balogh, T., “The HoloVizio system, ” Proc. SPIE 6055, 60550U (2006).

[3] Balogh, T., Forgacs, T., Agocs, T., Bouvier, E., Bettio, F., Gobbetti, E., Zanetti, G., ”A Large Scale Interactive Holographic Display,” IEEE VR2006, Virginia, USA.

[4] Balogh, T., et al, “A Scalable Hardware and Software System for the Holographic Display of Interactive Graphic Applications,” EuroGraphics 2005, Dublin.

[5] OSIRIS Project – Original System for Image Rendition via Innovative Screens, EU IST-FP6 Integrated Project OSIRIS (IST-33799 IP), http://www.osiris-project.eu/

[6] UCSD/MERL Light Field Repository, http://vision.ucsd.edu/datasets/lfarchive/

[7] The (New) Stanford Light Field Archive, http://lightfield.stanford.edu/

[8] Megyesi, Z., Barsi, A., Balogh, T., “3D Video Visualization On The Holovizio™ System,” 3DTV-CON 2008, Istanbul, Turkey

[9] Upstage Productions: Time Freeze, http://www.upstage.com.au/content/standard.asp?name=Time_freeze

[10] Yang, J. C., Everett, M., Buehler, C., McMillan, L., “A Real-Time Distributed Light Field Camera,” 13th Eurographics Workshop on Rendering, 2002.

[11] Taguchi, Y., Koike, T., Takahashi, K., Naemura, T., “TransCAIP: Live Transmission of Light Field from a Camera Array to an Integral Photography Display,” ACM SIGGRAPH ASIA 2008.

[12] Matusik, W., Pfister, H., “3D TV: A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes,” ACM SIGGRAPH 2004.

[13] Lincoln, P., Nashel, A., Ilie, A., Towles, H., Welch, G., Fuchs, H., ”Multi-view lenticular display for group teleconferencing,” IMMERSCOM 2009.

[14] Zhang, Z., “A Flexible New Technique for Camera Calibration,” Technical Report MSR-TR-98-71, Microsoft Research, 1998.

[15] Tsai, R., "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," IEEE J. Robotics and Automation, vol. 3, no. 4, pp. 323-344, Aug. 1987.

[16] Wiegand, T., Sullivan, G. J., Bjontegaard, G., Luthra, A., "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, pp. 560-576, July 2003.

[17] Merkle, P., Muller, K., Smolic, A., Wiegand, T., "Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC," IEEE International Conference on Multimedia and Expo, pp.1717-1720, 9-12 July 2006.