FAST AND EFFICIENT DATA REDUCTION APPROACH FOR MULTI-CAMERA LIGHT FIELD DISPLAY TELEPRESENCE SYSTEMS

Vamsi Kiran Adhikarla[1,3], *ABM Tariqul Islam*[2], *Péter Tamás Kovács*[1,4], *Oliver Staadt*[2]

[1]Holografika, Baross u. 3. H-1192 Budapest, [2]Visual Computing Lab, University of Rostock, 18059 Rostock, Germany, [3]Pazmany Peter Catholic University, Faculty of information Technology, Budapest, Prater u. 50/a, [4]Department of Signal Processing, Tampere University of Technology, Finland

ABSTRACT

Cutting-edge telepresence systems equipped with multiple cameras for capturing the whole scene of a collaboration space, face the challenge of transmitting huge amount of dynamic data from multiple viewpoints. With the introduction of Light Field Displays (LFDs) in to the remote collaboration space, it became possible to produce an impression of 3D virtual presence. In addition, LFDs in current generation also rely on the images obtained from cameras arranged in various spatial configurations. To have a realistic and natural 3D collaboration using LFDs, the data in the form of multiple camera images needs to be transmitted in real time using the available bandwidth. Classical compression methods might resolve this issue to a certain level. However, in many cases the achieved compression level is by far insufficient. Moreover, the available compression schemes do not consider any of the display-related attributes. Here, we propose a method by which we reduce the data from each of the camera images by discarding unused parts of the images at the acquisition site in a predetermined way using the display model and geometry, as well as the mapping between the captured and displayed light field. The proposed method is simple to implement and can exclude the unnecessary data in an automatic way. While similar methods exist for 2D screens or display walls, this is the first such algorithm for light fields. Our experimental results show that an identical light field reconstruction can be achieved with the reduced set of data which we would have got if all the data were transmitted. Moreover, the devised method provides very good processing speed.

Index Terms **—** 3D-TV, image processing, 3-D video transmission, light field, HoloVizio, multi-view, telepresence, collaborative virtual environments.

1. INTRODUCTION

During the past few years, the demand of remote collaboration systems has increased firmly in the communication world. The introduction of the large high-resolution displays, into the collaboration space, have added another appealing dimension; now, the collaboration system is capable of integrating multiple cameras in order to capture and transmit the whole scene of the collaboration space, see Figure 1. Projection-based light field displays (LFDs)[1], used for 3D video display, could be one such example of large high-resolution display. The LFDs, in contrast to existing autosterescopic 3D displays (based on lenticular lenses or parallax barrier), do no project 2D views in multiple dimension rather recreate the 3D light field structure of a scene. Inclusion of LFDs

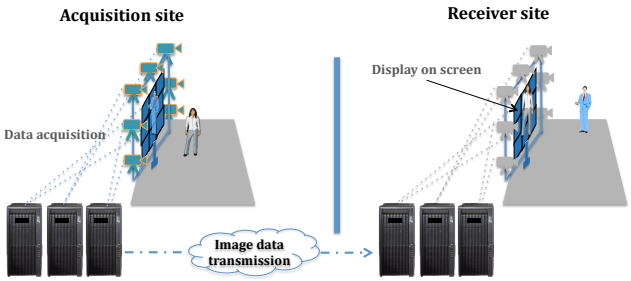

Figure 1. Multi-camera telepresence system.

in telepresence systems is becoming more popular because of the increased Field Of View (FOV) and the extreme feeling of reality. Moreover, there are no practical constraints on the number and the position of the users.

In recent years, we have seen that telepresence systems [2, 3] tend to be equipped with multiple cameras to capture the whole communication space; this integration of multiple cameras causes the generation of huge amount of dynamic camera image data. This large volume of data needs to be processed at the acquisition site, possibly requires aggregation from different network nodes and finally needs to be transmitted to the receiver site. It becomes intensely challenging to transmit this huge amount of data in real time through the available network bandwidth. The use of classical compression methods for reducing this large data might solve the problem to a certain degree, but using the compression algorithms directly on the acquired data may not yield sufficient data reduction. Finding the actual portion of data needed by the display system at receiver site and excluding the redundant image data would be highly beneficial for transmitting the image data in real time. At a later stage, we will use data compression schemes to further reduce the amount of data flow.

In this paper, we propose a fast and efficient data reduction strategy for multi-camera telepresence environments. We present an automatic approach that isolates the required areas of the incoming images, which contribute to the light field reconstruction. More specifically, we take into account the display model and captured and reconstructed light field geometry to devise a precise and automatic data picking procedure. In short, the key contribution of the paper is to reduce the data being transmitted in a light field telepresence system by finding the optimum region of interest from multiple camera images that is used in light field reconstruction.

The rest of the paper is organized as follow: in Section 2, we present the related works, Section 3 discusses about the implemen-

tation details. In Section 4, we show and explain the experimental results and conclusions are drawn in Section 5.

2. RELATED WORK

In [4], Lamboray et al. classified the data stream into several categories such as: bulk data, sporadic-event data, and real-time streaming data. They discussed the aspects of image-based and geometry-based reconstruction systems. They used strategies which allow to transmit selective updates from a collaboration scene. They introduced the idea of a *back channel* between acquisition and receiver site, by means of which positional data of the viewer at receiver site can be sent to the acquisition site for various purposes. In [5], the authors proposed a dynamic camera selection strategy which helps to reduce the number of recording cameras at the acquisition site. The authors also proposed to use a dynamic frustrum selection method in certain cases where the dynamic cameras selection fails. In [6], Lien et al. propose a model-driven data compression. Maimone and Fuchs [7] presented a concise study which suggests that when changes do not occur in all parts of a scene, camera selection should focus on the reduction of the overall amount of data.

In [8], Jones et al. propose a set of rendering methods for an autosteoreoscopic light field display which is able to present interactive 3D graphics to multiple simultaneous viewers 360 degrees around the display. Their method is a multiple-center-of-projection rendering technique for creating perspective-correct images from arbitrary viewpoints around the display. In [9], Magnor et al. have presented two schemes for light field compression. They have applied vector quantization, DCT coding and transform coding using spherical functions to the light field compression technique. In their schemes, the first coder has the advantage of decoding the recorded light-field segments very fast and thus achieves interactive rendering rate; and the second scheme describes a coder which is disparity compensating coder and it incrementally refines the light field during the decoding and predicts the intermediate light field images.

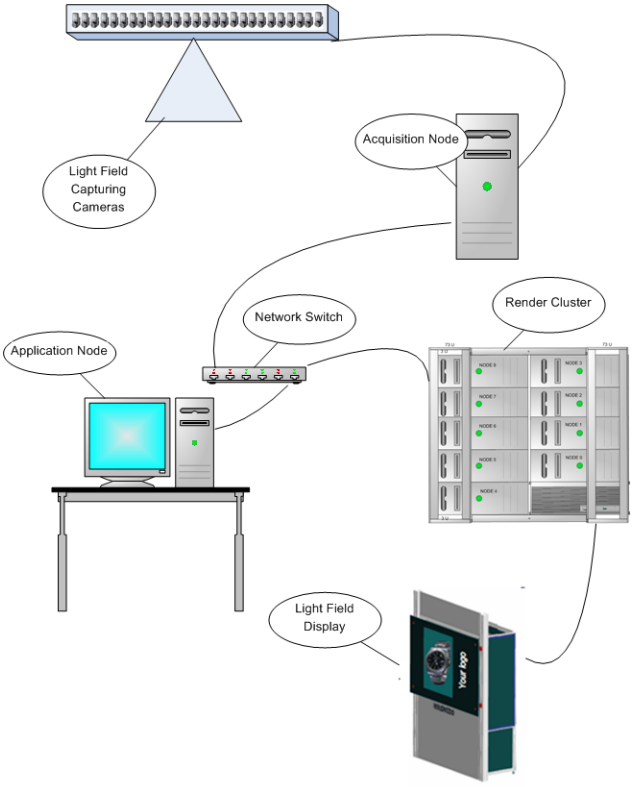

Figure 2. Light field capture, processing and displaying pipeline.

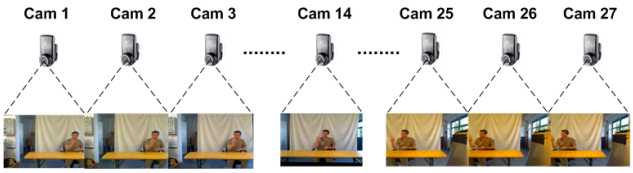

Figure 3. Sample light field capturing.

3. DATA REDUCTION APPROACH

Light field displays present the scene in 3D space. In other words, they do not simply project multiple views in different directions to create a 3D illusion. This primary observation is the basis of current data reduction scheme. Figure 6 shows the whole chain of capturing, processing, rendering and displaying of light field content. The capturing part consists of linearly arranged 27 compact USB cameras (for more information on the system design see [10]). Images from all the cameras are captured from a single acquisition node. This Linux-based acquisition node has extended USB ports to collect information from all the cameras. The captured images can be streamed directly to the rendering cluster via gigabit Ethernet connection. The rendering cluster then drives the optical modules of the display and finally a holographic screen is used to realize the 3D information in the form of light rays projected by the optical modules. The processing done by application node includes controlling operations such as camera calibration and checking the preview from all the cameras. The main part of this processing involves calculating the camera calibration data; a semi-automatic method is adopted for calibrating 27 cameras. Once the calibration is done, the calibration data is made available to the rendering cluster. The render cluster is equipped with light field modelling data built on display projection geometry beforehand. The incoming pixels of captured image stream are reordered on each cluster node's GPU using the available light field geometry and the camera calibration data. This pixel manipulation is handled using look-up tables, which are specific for each node in the render cluster.

The output of the render cluster is the 3D lighfield reconstruction of the scene obtained from multiple 2D images. Figure 3 shows an example light field capture and Figure 4 shows the reconstructed light field realized on a light field display.

An important observation from the light field reconstruction process is that not all the incoming pixels are used from all the cameras. Certain regions in each of the camera images are not used during the light field reconstruction. Another important observation is that the look-up tables used for re-ordering the pixels are constructed once in the beginning of the rendering process and remains same, as far as the mapping between the two light-field remains the same. These key observations forms the basis of the current work. Zooming in and out, or shifting the light field mapping of course forces the recalculation of these tables.

Figure 4. Light field reconstruction.

3.1. Experimental setup

In the current experiment, we used Holografika's HV721RC light field display. The is a large-scale display and can support multiple users simultaneously. The main reason behind choosing the display for the preliminary experiments is it's simplified geometry. As the case with a typical telepresence system, we assume that the capturing is done locally and rendering is done at a remote place. The camera system and demo computer are at a local site and the render cluster together with the optical modules and the display are located at a remote place. As this is first version of the telepresence system, we assume that the local and remote site are not far away from each other and communicate via gigabit Ethernet connection. For further simplicity, we assume a one-way telepresence system in other words, the locally captured images are sent to the remote place and rendered on the display.

3.2. Experimental procedure

As mentioned before the main aim of the current experiment is to reduce the amount of data flow still maintaining the same visual quality and we intend to solve this problem not by exploring image/video coding schemes, but rather taking in to account the display model and camera calibration. In order to achieve this, the first step is to identify the pixels from the input image stream which are discarded after the final rendering. Figure 5 shows significat pixel locations (pixels in white) based on the look-up tables in one of the experimental captures. More precisely, we used the pixel to light ray mapping information to mark the positions of the pixels from each of the camera images used by all nodes in the rendering cluster. In Figure 6, we present the percentage of pixels referred in the look-up tables for pixel re-ordering from each camera image. Note that the asymmetric nature of the curve is the effect of chosen region of interest (can be observed from Figures 3 & 4) and also a part of it is driven by the camera rotation. Also, please note that due to the vertical misalignments of the capture cameras, the top and bottom of the source images are cropped in this mapping between the incoming and outgoing light fields (we essentially zoomed inside the light field), hence the unused pixels on the top and bottom of some images. In synthetic setups, or using a more precise camera system the ratio of used pixels would

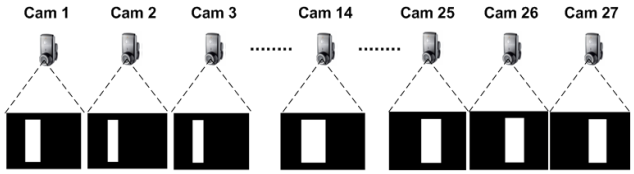

Figure 5. Calculated significant camera image pixels from a sample capture.

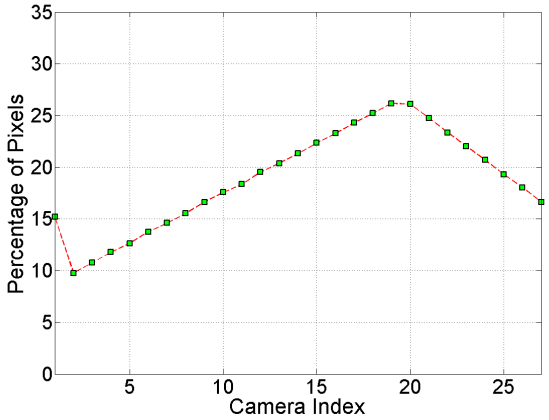

Figure 6. Percentage of pixels from each camera image used in sample light field reconstruction.

be higher.

One of the most important observations from Figure 6 is that the number of pixels utilized from each camera frame is not the same for any two cameras. Moreover, it is more general that these pixels are not chosen in the same pattern and the from same locations on every image. However, the significant pixels from every image form the shape of a rectangular box of varying area across multiple camera images and once set, the shape and the area remains same throughout the capture. This means, light field from the current capture setup can be comfortably reconstructed using the patterns of useful pixels on multiple camera images. In simple terms, it is possible to recreate exactly the same 3D impression with suitably chosen pixel subset on camera images using the masks in a pre-calculated pattern. We exploit this observation in our approach to reduce the amount of data being transmitted.

As the first step, the look-up table generation mechanism is shifted from the remote to the local site and is included as a part of the processing on the transmission side. The look-up table generation is carried out before transmitting the data and after finalizing the calibration. These tables are generated for all the nodes and once this is done, we introduce an additional processing step on all the camera images where we create a mask for each camera image that define a pattern of significant pixels needed by all the nodes in render cluster. The incoming camera images are carefully tailored using the created masks. As soon as the look-up tables are made available, this step can be very fast and apparently does not involve a lot of processing. Thus we can create a one to one mapping of the incoming and outgoing camera images. The processed camera images are now light weight and are sent to the remote site. To speed-up the process of creating masks and accessing the image pixels locally, we introduce an additional processing computer on the local site.

Note that the look-up tables are now available already and thus

the rendering cluster does not require additional time generating them. Also camera calibration data does not need to be transmitted as the information is also included in the form of generated look-up tables. Thus, instead of sending camera calibration data to the rendering cluster, we send the lookup tables for each node before the rendering process. The remote render cluster uses the received look-up tables and the reduced image data to do the light field rendering. As we discard parts of camera images, the resulting image texture coordinates may not coincide with the coordinates in the look-up tables. Thus, we need to store texture coordinate offset values in both X and Y direction for all the 27 cameras. The 54 valued offset texture is also transmitted before the actual rendering starts.

4. RESULTS

We tested the performance of the proposed approach on a pre-recorded 19s footage, 'Telepresence' using a HV721RC light field display. With the given initial conditions, we demonstrated that the proposed approach results in the same light field reconstruction without introducing any temporal or spatial artefacts and yet using only up to 20% of the whole data stream. Thus, the bandwidth resource consumption is effectively reduced by a factor of five. Also because of the reduced image resolution, GPU uploading and hence the overall rendering at the remote site becomes faster. Although the amount of data being uploaded is reduced, for the final rendering the number of texels used remains same and thus there is not any significant speed up in the rendering frame rate.

5. CONCLUSIONS

In this paper, we presented a lossless approach to reduce the data flow in multi-camera telepresence systems using light field displays. The proposed method does not rely on image/video coding schemes, but rather uses the display projection geometry to exploit and eliminate redundancy. We proposed minor changes in the capturing, processing and rendering pipeline with an additional processing at the local transmission site that helps achieving significant data reduction. Furthermore, the additional processing step before transmission, mostly involves simple image processing operations such generating masks and extracting bunch of pixels and needs to be done only once. The processed and transmitted data not only consumes less bandwidth but also speeds up the texture upload process.

In the current work, we showed the use of global masks to reduce pixel data selectively. In practice, each rendering node does not need the whole information, even from the extracted pixel subset. It is possible to customize the masks for each of the render cluster nodes, which can further improve speed. Also, the camera images can be subjected to a 90 degree rotation soon after the capture and now, we access the pixels row-wise for selective transmission. This might bypass any unnecessary passages during the memory access and direct memory offsets can be used.

In the current work, we made an assumption that the local and remote sites are connected via a low latency, relatively high-bandwidth connection. In general this is not the case and in order to transmit the light field data over longer distances, it is possible to incorporate multiview coding schemes such as H.264, MVC and HEVC. Also the capturing speed at the acquisition site is a bottleneck in the current setup. Using constant exposure time cameras with hardware trigger might further increase the accuracy in the camera synchronization.

6. ACKNOWLEDGEMENT

The research leading to these results has received funding from the DIVA Marie Curie Action of the People programme of the European Unions Seventh Framework Programme FP7/2007- 2013/ under REA grant agreement 290227. The research leading to these results has also received funding from the PROLIGHT-IAPP Marie Curie Action of the People programme of the European Unions Seventh Framework Programme FP7/2007- 2013/ under REA grant agreement 324499.

7. REFERENCES

[1] Tibor Agocs, Tibor Balogh, Tamas Forgacs, Fabio Bettio, Enrico Gobbetti, Gianluigi Zanetti, and Eric Bouvier, "A large scale interactive holographic display," in *IEEE Virtual Reality Conference (VR 2006)*, Washington, DC, USA, 2006, p. 57.

[2] Andrew Maimone and Henry Fuchs, "Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras," in *10th IEEE ISMAR*, October 2011, pp. 137 –146.

[3] Benjamin Petit, Jean-Denis Lesage, Clment Menier, Jrmie Allard, Jean-Sbastien Franco, Bruno Raffin, Edmond Boyer, and Franois Faure, "Multicamera real-time 3d modeling for telepresence and remote collaboration," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, pp. 247108–12, 2009.

[4] Edouard Lamboray, Stephan Wurmlin, and Markus Gross, "Data streaming in telepresence environments," *IEEE Tran. on Visualization and Comp. Graphics*, vol. 11, no. 6, pp. 637–648, Jan. 2005.

[5] Malte Willert, Stephan Ohl, and Oliver G. Staadt, "Reducing bandwidth consumption in parallel networked telepresence environments," in *VRCAI*, 2012, pp. 247–254.

[6] Jyh-Ming Lien, Gregorij Kurillo, and Ruzena Bajcsy, "Multi-camera tele-immersion system with real-time model driven data compression," *The Visual Computer*, vol. 26, no. 1, pp. 3–15, Jan. 2010.

[7] Andrew Maimone and Henry Fuchs, "A first look at a telepresence system with room-sized real-time 3d capture and life-sized tracked display wall," in *ICAT*, November 2011.

[8] Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec, "Rendering for an interactive 360 light field display," *ACM Transactions on Graphics (TOG)*, vol. 26, pp. 338–343, July 2007.

[9] Marcus Magnor and Bernd Girod, "Data compression for light-field rendering," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 338–343, April. 2000.

[10] Tibor Balogh and Kovács Péter Tamás, "Real-time 3d light field transmission," pp. 772406–772406–7, 2010.