RENDERING AND INTERACTION ON PROJECTION-BASED LIGHT FIELD DISPLAYS

Theses of the Ph.D. dissertation

Vamsi Kiran Adhikarla

Scientific adviser:

Péter Szolgay, D.Sc.

fides et ratio

Pázmány Péter Catholic University

Faculty of Information Technology and Bionics

Roska Tamás Doctoral School of Sciences and Technology

Budapest 2015

Chapter 1

Introduction

The term *light field* is mathematically defined as a function that describes the amount of light fared in all directions from every single point in space at a given time instance[1]. Over a continuous time, the collection of light rays emitted at all the points in space continuously defines a 3D scene. Presenting a scene in 3D, thus, involves capturing wavelengths of light at all points in all directions at all time instances and displaying this captured information within a given region of interest of a scene. However, this may not possible in reality due to practical limitations such as complex capturing procedure and enormous amount of data every single time instant. Many attempts were made in the past for simplifying and displaying slices of light field such as : stereoscopy (Anaglyph 3D system, Polarized 3D system, Active shutter 3D system), lens based autostereoscopic displays, volumetric displays, head mount displays and displays based on motion tracking and spatial multiplexing etc., The main idea behind was to reduce the number of dimensions of the light field function to derive a discrete segment of light field [2]. However, practical means to produce highly realistic light field with continuous depth cues for 3D perception are still unavailable.

A more pragmatic and elegant approach for presenting a light field along the lines of its actual definition has been pioneered by HoloVizio: projection-based light field displaying technology[3] [4] [5]. Taking inspiration from the real-world, a projection-based light field display emits light rays from multiple perspectives using a set of projection engines. Various scene points are described by intersecting light rays at corresponding depths. A holographic diffuser modulates the emitted light rays and helps to achieve directional light transmission with minimum aliasing. These displays have a high potential to be the future 3D display solution and novel methods to process the massive light field data are very much needed.

1.1   Open-Ended Questions

In order to meet the FOV requirements of a light field display and to have a realistic and natural 3D displaying, we need to capture views from several cameras. In the current generation of light field displays, horizontal only parallax is supported and thus it is enough to consider the camera setups in one dimensional configuration. It is clear that for providing high quality 3D experience and supporting motion parallax, we require massive input image data. The increased

dimensionality and size of the data opened up many research areas ranging through capturing, processing, transmitting, rendering and interaction. Some of the open ended questions are

- **Content creation** - what are the optimal means to acquire/capture suitable light field data to fully meet the requirements of a given light field display?

- **Representation, coding & transmission** - what is the optimal format for storing and transmitting acquired light field data?

- **Rendering & synthesis** - how the captured light field data can be used for rendering and how to create/synthesize the missing light rays?

- **Interaction** - how the interaction can be extended for light field content to appeal to the design, gaming and media industry?

- **Human visual system** - how the human visual system functioning can be exploited to improve the quality of light field rendering?

- **Quality Assessment** - what are the ways to measure/quantify the quality of a light field rendering and how can we automatically detect any annoying artifacts?

1.2 Aims and Objectives

The main aim of the current work is to assess the requirements for suitable and universal capture, rendering and interaction of 3D light field data for projection-based light field displays. Specifically, the emphasis is on *representation*, *rendering* and *interaction* aspects. Considering the existing rendering procedures, the first part of the work aims to derive the requirements for light field representation for given display configurations and also presents a rendering prototype with light weight *multiview* image data. The second part of the work aims to optimize the rendering procedures to comply with the display behavior. In the last part, the objective is to experiment and evaluate interactive rendering using low cost motion sensor device.

1.3 Scope of Work

Concentrating on projection-based light field displays, the work has three branches : assessing the requirements for a suitable representation, rendering visual quality enhancement through depth retargeting and exploring direct touch interaction using Leap Motion Controller.

- By implementing and testing the state-of-art rendering methods, requirements for a future light field representation are presented for projection-based light field displays. The scope of this work does not include deriving a standard codec for encoding/decoding the light field.

- The depth retargeting method solves a non-linear optimization to derive a noval scene to display depth mapping function. Here, the aim is not to acquire/compute accurate depth, instead use a real-time capturing and rendering pipeline for investigating adaptive

depth retargeting. The retargeting is embedded into the renderer to preserve the real-time performance. For testing the performance, the retargeting method is also applied to synthetic 3D scenes and compared with other real-time alternatives. Comparison is not carried out with methods that are not real-time.

- The direct touch interaction setup provides a realistic direct haptic interaction with virtual 3D objects rendered on a light field display. The scope of this work does not include any modeling of an interaction device.

1.4 Workflow

The work began with a study of several light field representations for real-world scenes that were proposed in the literature. The main idea behind investigating a representation is not deriving an efficient and real-time light field encoding/decoding method, but rather explore the light field display geometry to investigate a suitable representation. Existing and well-knowns representations are based on the notion of displaying 3D from multiple 2D. However, projection-based light field displays represent a scene through intersecting light rays in space i.e., instead of several discrete 2D views, any user moving in front of the display perceives several *light field slices*. Thus, the existing light field representations are not optimized for such displays and the required representation should consider the process of light field reconstruction (rendering) on these displays.

A popular approach for light field rendering from real-world scenes is to acquire images from multiple perspectives (*multiview images*) and re-sample the captured database to address the display light rays [6, 7, 8]. Due to the geometrical characteristics of light field displays, it is more reliable to render using images acquired from several closely spaces cameras. Depending on the display geometry, we may need 90-180 images. By carefully examining how several light rays leaving the display are shaded, I derived a data reduction approach that eliminates the unwanted data during reconstruction process. Based on the immediate observations, a light weight representation based on discrete 2D views with each view having a different resolution than the other is followed. These investigations also opened up the requirements for a novel light field representation and encoding schemes.

A different approach for light field rendering in the literature is all-in-focus rendering [9]. Instead of directly sampling the captured data base, this rendering approach also computes the depth levels for various display light rays. A major advantage using this approach is that, it is possible to achieve a good quality light field with less number of cameras. However, the visual quality is highly dependent on the accuracy of the available depth. Currently, there are not any real-time methods that can deliver pixel precise accurate depth for novel light field reconstruction. Thus, this rendering approach is not taken forward for investigations related to light field representation.

One of the important constraints of any 3D display is available *depth of field (DOF)*. If the extent of scene depth is beyond the displayable extent, a disturbing blur is perceived. An important goal of the work was to address the problem of depth retargeting for light field displays (constraining the scene depth smartly to fit to the display depth of field). Warping based approaches are proposed in the literature stereo 3D depth retargeting. These methods do not explicitly consider the scene [10] and they need further adaption to suit to the full parallax behavior of a light field display. Furthermore with methods based on warping, distortions are inevitable, especially,

if there are vertical lines in the scenes. An alternative is to compute and work on the scene depth directly to achieve retargeting. As depth calculation is an integral part of the all-in-focus rendering pipeline, this approach is taken further and the behavior is adapted to achieve content adaptive and real-time depth retargerting on a light field display.

Interaction and rendering are intertwined processes. As light field displays represent a novel technology in the field of 3D rendering, they also require design and evaluation of novel interaction technologies and techniques for successful manipulation of displayed content. In contrast to classic interaction with 2D content, where mice, keyboards or other specialized input devices (e.g., joystick, touch pad, voice commands) are used, no such generic devices, techniques, and metaphors have been proposed for interaction with 3D content. The selected interaction techniques usually strongly depend on individual application requirements, design of tasks and also individual user and contextual properties. One of the goals of the current work is to enable accurate, natural and intuitive freehand interaction with 3D objects rendered on a light field display. For this purpose a basic and most intuitive interaction method in 3D space, known as "direct touch" is proposed. The method directly links an input device with a display and integrates both into a single interface. The main aim was not to research a hand or motion tracking hardware, but use commercially available *Leap Motion Controller* sensor for motion sensing and achieve an interactive rendering for manipulation of virtual objects on a light field display.

All the work is implemented in C++ using OpenGL SDK[11]. The rendering code for light field depth retargeting is written in CUDA[12]. For programming the interaction part, Leap SDK[13] is used.

Chapter 2

Research Methodology

For experiments on assessing the light field representation requirements and implementing the data reduction prototype, a real-time capture and rendering system [8] based on re-sampling the captured light field with 27 USB cameras is used. Glasses-free 3D cinema system developed by Holografika is used for viewing the renderings (see [C2]).

The end-to-end capture and retargeted rendering pipeline is implemented in Linux. On-the-fly light field retargeting and rendering is implemented on GPU using CUDA. The results of the propsed content aware retargeting are teseted on a Holografika 72in light field display that supports $50°$ horizontal Field Of View (FOV) with an angular resolution of $0.8°$. The aspect ratio of the display is 16:9 with single view 2D-equivalent resolution of 1066×600 pixels. The display has 72 SVGA 800x600 LED projection modules which are pre-calibrated using an automatic multiprojector calibration procedure [14]. The front end is an Intel Corei7 PC with an Nvidia GTX680 4GB, which captures multiview images at 15 fps in VGA resolution using 18 calibrated Logitech Portable Web cameras. The camera rig covers a base-line of about 1.5m and is sufficient to cover the FOV of light field display. In the back end, we have 18 AMD Dual Core Athlon 64 X2 5000+ PCs running Linux and each equipped with two Nvidia GTX560 1 GB graphics boards. Each node renders images for four optical modules. Front-end and back-end communicate over a Gigabit Ethernet connection.

For experiments on interaction, *Leap Motion Controller* is used as a tracking device. HoloVizio OpenGL wrapper (see [4]) is used for real-time light field rendering and a small scale light field display of the size comparable to the FOV of Leap Motion Controller is used for visualization and interaction. For accurately acquiring the interaction data, the Leap Motion Controller is placed at 100mm below the center of the display.

Chapter 3

New Scientific Results

The results of this dissertation can be categorized into three main parts:
results dealing with light field

- Representation

- Retargeted rendering

- Interaction

The respective contributions are briefed in the following thesis groups.

3.1 Thesis Group I - Light Field Representation

Examining the light field conversion process from camera images acquired from several closely spaced cameras, I proposed a fast and efficient data reduction approach for light field transmission in multi-camera light field display telepresence environment.
Relevant publications: [C2] [C3] [C4] [C5] [O1] [J3] [O2] [O3]

3.1.1 Fast and Efficient Data Reduction Approach for Multi-Camera Light Field Display Telepresence System

I proposed an automatic approach that isolates the required areas of the incoming multiview images, which contribute to the light field reconstruction. Considering a real-time light field telepresence scenario, I showed that up to 80% of the bandwidth can be saved during transmission.

- Taking into account a light field display model and the geometry of captured and reconstructed light field, I devised a precise and automatic data picking procedure from multiview camera images for light field reconstruction.

- The proposed method does not rely on image/video coding schemes, but rather uses the display projection geometry to exploit and eliminate redundancy.

- Minor changes in the capturing, processing and rendering pipeline have been proposed with an additional processing at the local transmission site that helps in achieving significant data reduction. Furthermore, the additional processing step needs to be done only once before the actual transmission.

3.1.2   Towards Universal Light Field Format

Exploring the direct data reduction possibilities (without encoding and decoding), I presented the preliminary requirements for a universal light field format.

- Simulations were made to see how the *field of view* (FOV) of the receiver's light field display affects the way the available captured views are used. Seven hypothetical light field displays were modeled, with the FOV ranging between 27^0 and 89^0. Source data with 180 cameras, in a 180^0 arc setup, with 1 degree angular resolution has been used.

- Analyzing the pixel usage patterns during the light field conversion, the affect of display's FOV on the number of views required for synthesizing the whole light field image has been realized.

- This analysis has shown that depending on the FOV of the display, the light field conversion requires 42 to 54 views as input for these sample displays. Note the actual number depends on the source camera layout (number and FOV of cameras), but the trend is clearly studied.

- Based on the use cases and processing considerations, three aspects were formulated that need attention and future research when developing compression methods for light fields:

  - The possibility to encode views having different resolution must be added.
  - The ability to decode the required number of views should be supported by the ability to decode views partially, starting from the center of the view, thus decreasing the computing workload by restricting the areas of interest.
  - Third, efficient coding tools for nonlinear (curved) camera setups shall be developed, as we expect to see this kind of acquisition format more in the future.

3.2   Thesis Group II - Retargeted Light Field Rendering

I presented a prototype of an efficient on-the-fly content aware real-time depth retargeting algorithm for accommodating the captured scene within acceptable depth limits of a display. The discrete nature of light field displays results in aliasing when rendering scene points at depths outside the supported depth of field causing visual discomfort. The existing light field rendering techniques: plain and rendering through geometry estimation, need further adaption to the display characteristics, for increasing quality of visual perception. The prototype addresses the problem of light field depth retargeting. The proposed

algorithm is embedded in an end-to-end real-time system capable of capturing and reconstructing light field from multiple calibrated cameras on a full horizontal parallax light field display.
Relevant publications: [C6] [C7] [C8] [J4] [J2] [C9]

3.2.1 Perspective Light Field Depth Retargeting

I proposed and implemented a perspective depth contraction method for live light field video stream that preserves the 3D appearance of salient regions of a scene. The deformation is globally monotonic in depth, and avoids depth inversion problems.

- All-in-focus rendering technique with 18 cameras in the capturing side is considered for implementing the retargeting algorithm and a non-linear transform from scene to display that minimizes the compression of salient regions of a scene is computed.

- To extract the scene saliency, depth and color saliency from perspectives of central and two lateral display projection modules is computed and combined. Depth saliency is estimated using a histogram of the pre-computed depth map and to estimate color saliency, a gradient map of the color image associated to the depth map of the current view is computed and dilated to fill holes. The gradient norm of a pixel represents color saliency.

- To avoid any abrupt depth changes scene depth range is quantized into different depth clusters and the depth and color saliency inside each cluster is accumulated.

- While adapting the scene to display, displacement of depth planes parallel to XY = 0 plane results in XY cropping of the scene background. Thus, in order to preserve the scene structure, perspective retargeting approach is followed i.e., along with z, XY positions are also updated proportional to $\frac{1}{\delta Z}$, as done in a perspective projection. Thus in the retargeted space, the physical size of the background objects is less than the actual size. However, a user looking from the central viewing position perceives no change in the apparent size of the objects as the scene points are adjusted in the direction of viewing rays.

3.2.2 Real-time Adaptive Content Retargeting for Live MultiView Capture and Light Field Display

I presented a real-time plane sweeping algorithm which concurrently estimates and retargets scene depth. The retargeting module is embedded into an end-to-end system capable of real-time capturing and displaying with full horizontal parallax high-quality 3D video contents on a cluster-driven multiprojector light field display with full horizontal parallax.

- While this use is straightforward in a 3D graphics setting, where retargeting can be implemented by direct geometric deformation of the rendered models, the first setup for real-time multiview capture and light field display rendering system incorporating the adaptive depth retargeting method has been proposed and implemented.

- The system seeks to obtain a video stream as a sequence of multiview images and render an all-in-focus retargeted light field in real-time on a full horizontal light field display.

The input multiview video data is acquired from a calibrated camera rig made of several identical off the shelf USB cameras.

- The captured multiview data is sent to a cluster of computers which drive the display optical modules. Using the display geometry and input camera calibration data, each node estimates depth and color for corresponding light rays. To maintain the real-time performance, the depth estimation and retargeting steps are coupled.

3.2.3 Adaptive Light Field Depth Retargeting Performance Evaluation

I evaluated the objective quality of the proposed depth retargeing method by comparing it with other real-time models: linear and logarithmic retargeting and presented the analysis for both synthetic and real-world scenes.

- To demonstrate results on synthetic scenes, two sample scenes are considered: Sungliders and Zenith. The original depths of the scenes are 10.2m and 7.7m, that are remapped to a depth of 1m to match the depth range of the display. Retargeted images are generated by simulating the display behavior.

- To generate the results for logarithmic retargeting, a function of the form $y = a + b * log(c + x)$ is used, where y and x are the output and input depths. The parameters a, b & c are chosen to map the near and far clipping planes of the scene to the comfortable viewing limits of the display.

- Objective evaluation is carried out using two visual metrics: *SSIM and PSNR*. Results show that the adaptive approach performs better and preserves the object dephts, avoiding to flatten them.

- To demonstrate the results on real-world scenes, using a simple hand-held camera, the processes of live multiview capturing and real-time retargeted rendering is recorded. It should be noted that the 3D impression of the results on the light field display can not be fully captured by a physical camera.

- Experiments show that the results from the real-world scenes conform with the simulation results on the synthetic scenes. By following direct all-in-focus light field rendering, areas of the scene outside the displayable range are subjected to blurring. Linear retargeting achieves sharp light field rendering at the cost of flattened scene. Content aware depth retargeting is capable of achieving sharp light field rendering and also preserves the 3D appearance of the objects at the same time.

- The front end frame rate is limited at of $15 fps$ by the camera acquisition speed. The back end hardware used in the current work supports an average frame rate of $11 fps$. However, experiments showed that Nvidia GTX680 GPU is able to support $40 fps$.

- In the back end application the GPU workload is subdivided in this way: 30% to upsample depth values, 20% for census computation, 15% jpeg decoding, 13% to extract color from the depth map, other minor kernels occupy the remaining time. Retargeting is embedded in the upsampling and color extraction procedures.

3.3   Thesis Group III - Light Field Interaction

I designed and implemented two interaction setups for 3D model manipulation on a light field display. The gesture-based object interaction enables manipulation of 3D objects with 7DOFs by leveraging natural and familiar gestures.
Relevant publications: [C10] [C11] [J1] [C1]

3.3.1   HoloLeap: Towards Efficient 3D Object Manipulation on Light Field Displays

I designed and implemented HoloLeap - a system for interacting with light field displays (LFD) using hand gestures. For tracking user hands, leap motion controller (LMC) is used which is a motion sensing device introduced by Leap Motion Inc. and fits very well with the interaction needs. The device is proven to be relatively inexpensive, precise and is useful in tracking the hands and finger inputs than any existing free hands interaction devices. The device has a high frame rate and comes with a USB interface. The device provides a virtual interaction space of about one sq. meter with almost $1/100^{th}$ millimeter accuracy.

- Manipulation gestures for translation, rotation, and scaling have been implemented. Continuous rotation ("spinning") is also included. I designed a custom gesture set, as the increased depth perception of a light field display may affect object manipulation.

- The goal was to enhance the ad-hoc qualities of mid-air gestures. Compared to handheld devices (e.g., a mouse) gestural interaction allows one to simply walk up and immediately begin manipulating 3D objects. Rotation uses a single hand. The user rotates their wrist in the desired direction to rotate the object. It allows for fast.

- Rotation uses a single hand. The user rotates their wrist in the desired direction to rotate the object. It allows fast correction for each rotational degree of freedom and multiple axes of rotation in a single gesture. Moving two hands at once without increasing the distance between them translates the object. Scaling is activated by increasing and decreasing the distance between palms. HoloLeap does not use zooming as LFDs have a limited depth range. Scaling is provided as an alternative to facilitate model inspection and to easily provide an overview. Continuous rotation (spin) is activated with a double-hand rotation gesture.

Freehand Interaction with Large-Scale 3D Map Data

Extending the 3D model interaction framework, I implemented a gesture based interaction system prototype for the use-case - interaction with large-scale 3D map data visualized on a light field display in real time. 3D map data are streamed over Internet to the display end in real-time based on requests sent by the visualization application. On the user side, data is processed and visualized on a large-scale 3D light field display.

- The streaming and light field visualization is done on the fly without using any prerendered animations or images. After acquiring the interaction control messages from the Leap Motion Controller, rendering is performed in real-time on the light-field display's.

- For real-time visualization, HoloVizio OpenGL wrapper library is used which intercepts all OpenGL calls and sends rendering commands over the network as well as modify related data (such as textures, vertex arrays, VBOs, shaders etc.) on the fly to suit the specifications of the actual light field display.

- During interaction, panning in horizontal and vertical directions can be done by translating the virtual camera in the opposite direction by a given amount and is achieved using one hand (either left or right). For rotation, the virtual camera is made to mimic the hand rotation gestures. Zooming is achieved using two hands. Bringing two hands closer is the gesture for zooming out and taking the hands apart results in zooming in.

- The presented system is first of its kind exploring the latest advances in 3D visualization and interaction techniques.

3.3.2 Exploring Direct 3D Interaction for Full Horizontal Parallax Light Field Displays Using Leap Motion Controller

I proposed the first framework that provides a realistic direct haptic interaction with virtual 3D objects rendered on a light field display. The solution includes calibration procedure that leverages the available depth of field and the finger tracking accuracy, and a real-time interactive rendering pipeline that modifies and renders light field according to 3D light field geometry and the input gestures captured by the Leap Motion Controller. The implemented interaction framework is evaluated and the results of a first user study on interaction with a light field display are presented. This is a first attempt for direct 3D gesture interaction with a full horizontal parallax light field display.

- The application generates random patterns of tiles in run time and rendered at a given depth. In parallel, the application receives interaction data from Leap Motion Controller, processes and updates the renderer in real-time. The controlling PC runs GL wrapper and feeds the resulting visual data to optical modules. Hence we can see the same application running on a LCD monitor in 2D and on light field display in 3D.

- Real-time visualization is achieved using OpenGL wrapper library. A controlling PC runs two applications: main OpenGL frontend rendering application for 2D LCD display and backend wrapper application that tracks the commands in current instance of OpenGL (front end application) and generates modified stream for light field rendering. The front end rendering application also receives and processes user interaction commands from Leap Motion Controller in real-time.

- The interaction and display spaces are calibrated to provide an illusion of touching virtual objects. To the best of my knowledge, the is a first study involving direct interaction with virtual objects on a light field display using Leap Motion Controller. The proposed interaction setup is very general and is applicable to any 3D display without glasses. The method is scalable, and the interaction space can easily be extended by integrating multiple Leap Motion Controllers.

Direct touch interaction system evaluation

I have conducted a user study to evaluate the proposed freehand interaction with the light field display through a simple within-subject user study with 12 participants.

Three tiles of the same size were displayed simultaneously and the participants were asked to point (touch) the surface of the red tile as perceived in space. The positions of the tiles varied from trial to trial to cover the entire FOV of the display. 3D and 2D display modes were used representing two different experimental conditions:

- In 2D mode, the displayed objects were distributed on a plane in close proximity of the display surface; and

- In 3D mode, the objects were distributed in a space with the distance varying from 0 to 7 cm from the display.

The 2D mode provided a control environment, which was used to evaluate the specifics of this particular interaction design: the performance and properties of the input device, display dimensions, specific interaction scenario (e.g., touching the objects), etc. Each participant was asked to perform 11 trials within each of the two conditions. The sequence of the conditions was randomized across the participants to eliminate the learning effect. The light field display and the interaction design were evaluated from the following aspects: task completion times, cognitive workload and perceived user experience.

Results show that users did not perceive any significant differences between the conditions in terms of general impression, the easiness to learn how to interact with the content, the efficiency of such interaction, the reliability or the predictability of the interfaces used and the excitement or the motivation for such an interaction. The exception is the novelty subscale where a tendency towards higher preferences for the 3D mode can be observed. The analysis of the post-study questionnaire revealed that the rendered objects were seen clearly in both experimental conditions. However, the users favored the 3D mode in terms of rendering realism. When asked to choose the easiest mode, the user's choices were equally distributed between both modes. However, when asked which mode led to more mistakes in locating the exact object position, two-thirds indicated the 3D mode, which is reflected also in longer task completion times in this particular mode. Finally, when asked about their preference, two-thirds of the participants chose the 3D mode as their favorite one.

Chapter 4

Applications of the Work

3D video is increasingly gaining prominence as the next major innovation in video technology that greatly enhances the quality of experience. Consequently, the research and development of technologies related to 3D video are increasingly gaining attention [15]. The research and development work done during this thesis work mainly finds applications related to the content transmission and rendering parts.

In recent years, the telepresence systems [16, 17] 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. The proposed automatic data reduction approach exploring the target display geometry might be handy in such situations.

The current generation 3D films are mainly based on *Steroscopic 3D*. Every year, more and more movies are produced in 3D and television channels started to launch their broadcast services for events such as sports and concerts in 3D. But despite these technological advances, the practical production of stereoscopic content that results in a natural and comfortable viewing experience in all scenarios is still a challenge [10]. Assuming that the displaying part has no limitations, the basic problem in a stereoscopic setting lies in the human visual perception [18] [19]. With the advances of more natural 3D displaying technologies, the visual performance has greatly increased. However, the problem of depth accommodation has slid to the display side and there is a need for techniques that address the display content adaption. The content adaptive depth retargeting presented in the work can be applied for automatically adapting the scene depth to the display depth. The method works in real-time and can be applied for automatic disparity correction and display adaption.

Light field displays offer several advantages over volumetric or autostereoscopic displays such as adjacent view isolation, increased field of view, enhanced depth perception and support for
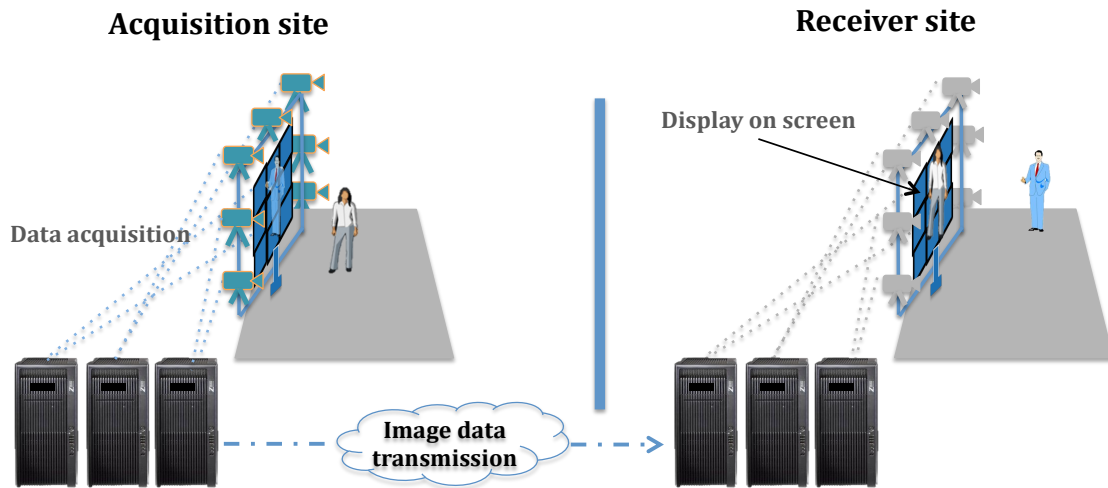

Acquisition site **Receiver site**

Figure 4.1: Multi-camera telepresence system.

horizontal motion parallax. However, it is unclear how to take full advantage of these benefits, as user interface techniques with such displays have not been explored. Google street view development laid an important milestone for online map services. The prototyped interaction setups can be extended for applying to such navigation purposes.

Acknowledgments

This dissertation work is a result of three years of research and development project that has been done in collaboration with Holografika and Pázmány Péter Catholic University, Hungary. It is a delight to express my gratitude to all, who contributed directly and indirectly to the successful completion of this work.

I am very grateful to Tibor Balogh, CEO and founder of Holografika Ltd., Zsuzsa Dobrányi, sales manager, Holografika Ltd., Péter Tamás Kovács, CTO, Holografika Ltd., and all the staff of Holografika for introducing Hungary to me and providing me an opportunity to work with a promising and leading edge technology - the HoloVizio.

My special thanks and sincere appreciation to Attila Barsi, lead software developer at Holografika for holding my hand and walking me through this journey. I would like to further extend my thanks to Péter Tamás Kovács for his unconditional support and supervision despite his hectic schedule. Their mentoring and invaluable knowledge is an important source to this research and so to this dissertation.

I am grateful to Prof. Péter Szolgay, Dean, Pázmány Péter Catholic University for letting me to start my Ph.D. studies at the doctoral school. I thankfully acknowledge him for his guidance and support throughout the work. His precious comments and suggestions greatly contributed to the quality of the work.

I am much indebted to all the members of Visual Computing group, CRS4, Italy and in particular to Enrico Gobbetti, director and Fabio Marton, researcher, CRS4 Visual Computing group for their exceptional scientific supervision. The work would not be complete without their guidance. I feel very lucky to have had the opportunity to work with them.

I am thankful to Jaka Sodnik and Grega Jakus from University of Ljubljana for sharing their knowledge and valuable ideas.

I would like to show my love and record my special thanks to my parents and to my sister who kept me motivated during my work with their unconditional support.

I would like to express my deep love and appreciation to my beautiful Niki for her patience, understanding and for being with me through hard times.

Finally, I would like to thank everybody who was significant to the successful understanding of this thesis, as well as expressing my apology that I could not mention personally one by one.

List of Publications

Journal Publications

[J1] **V. K. Adhikarla**, J. Sodnik, P. Szolgay, and G. Jakus, "Exploring direct 3D interaction for full horizontal parallax light field displays using leap motion controller," *Sensors*, pp. 8642-8663, April, 2015, Impact Factor : 2.245. 10

[J2] **V. K. Adhikarla**, F. Marton, T. Balogh, and E. Gobbetti, "Real-time adaptive content retargeting for live multi-view capture and light field display," *The Visual Computer, Springer Berlin Heidelberg*, vol. 31, May, 2015, Impact Factor : 0.957. 8

[J3] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, F. Dufaux, P. T. Kovács, and **V. K. Adhikarla**, "Subjective evaluation of super multi-view compressed contents on high-end light-field 3D displays," *Signal Processing: Image Communication, Elsevier*, 2015, Impact Factor : 1.462. 6

[J4] B. Maris, D. Dallálba, P. Fiorini, A. Ristolainen, L. Li, Y. Gavshin, A. Barsi, and **V. K. Adhikarla**, "A phantom study for the validation of a surgical navigation system based on real-time segmentation and registration methods," *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, pp. 381 - 382, 2013, Impact Factor : 1.707. 8

Conference Publications

[C1] **V. K. Adhikarla**, G. Jakus, and J. Sodnik, "Design and evaluation of freehand gesture interaction for lightfield display," in *Proceedings of International Conference on Human-Computer Interaction*, pp. 54 - 65, 2015. 10

[C2] T. Balogh, Z. Nagy, P. T. Kovács, and **V. K. Adhikarla**, "Natural 3D content on glasses-free light-field 3D cinema," in *Proceedings of the SPIE: Stereoscopic Displays and Applications XXIV*, vol. 8648, Feb., 2013. 5, 6

[C3] **V. K. Adhikarla**, A. Tariqul Islam, P. Kovacs, and O. Staadt, "Fast and efficient data reduction approach for multi-camera light field display telepresence systems," in *Proceedings of 3DTV-Conference 2013*, Oct 2013, pp. 1-4. 6

[C4] P. Kovács, Z. Nagy, A. Barsi, **V. K. Adhikarla**, and R. Bregovic, "Overview of the applicability of h.264/mvc for real-time light-field applications," in *Proceedings of 3DTV-Conference 2014*, July 2014, pp. 1-4. 6

[C5] P. Kovacs, K. Lackner, A. Barsi, **V. K. Adhikarla**, R. Bregovic, and A. Gotchev, "Analysis and optimization of pixel usage of light-field conversion from multi-camera setups to 3D light-field displays," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 86-90. 6

[C6] R. Olsson, **V. K. Adhikarla**, S. Schwarz, and M. Sjöström, "Converting conventional stereo pairs to multiview sequences using morphing," in *Proceedings of the SPIE: Stereoscopic Displays and Applications XXIII*, vol. 8288, 22 - 26 January 2012. 8

[C7] **V. K. Adhikarla**, P. T. Kovács, A. Barsi, T. Balogh, and P. Szolgay, "View synthesis for lightfield displays using region based non-linear image warping," in *Proceedings of International Conference on 3D Imaging (IC3D)*, Dec 2012, pp. 1-6. 8

[C8] **V. K. Adhikarla**, A. Barsi, P. T. Kovács, and T. Balogh, "View synthesis for light field displays using segmentation and image warping," in *First ROMEO workshop*, July 2012, pp. 1-5. 8

[C9] **V. K. Adhikarla**, F. Marton, A. Barsi, E. Gobbetti, P. T. Kovacs, and T. Balogh, "Real-time content adaptive depth retargeting for light field displays," in *Proceedings of Eurographics Posters*, 2015. 8

[C10] **V. K. Adhikarla**, P. Wozniak, A. Barsi, D. Singhal, P. Kovács, and T. Balogh, "Freehand interaction with large-scale 3d map data," in *Proceedings of 3DTV-Conference 2014*, July 2014, pp. 1-4. 10

[C11] **V. K. Adhikarla**, P. Wozniak, and R. Teather, "Hololeap: Towards efficient 3D object manipulation on light field displays," in *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction (SUI)*, 2014, pp. 158-158. 10

Other Publications

[O1] P. Kovács, A. Fekete, K. Lackner, **V. K. Adhikarla**, A. Zare, and T. Balogh, "Big buck bunny light-field test sequences," in *ISO/IEC JTC1/SC29/WG11 M35721*, Feb., 2015. 6

[O2] **V. K. Adhikarla**, "Content processing for light field displaying," in *Poster session, EU COST Training School on Plenoptic Capture, Processing and Reconstruction*, June, 2013. 6

[O3] **V. K. Adhikarla**, "Data reduction scheme for light field displays," in *Poster session, EU COST Training School on Rich 3D Content: Creation, Perception and Interaction*, July, 2014. 6

References

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. 1

[2] B. Masia, G. Wetzstein, P. Didyk, and D. Gutierrez, "A survey on computational displays: Pushing the boundaries of optics, computation, and perception," *Computers & Graphics*, vol. 37, no. 8, pp. 1012 – 1038, 2013. 1

[3] T. Agocs, T. Balogh, T. Forgacs, F. Bettio, E. Gobbetti, G. Zanetti, and E. Bouvier, "A large scale interactive holographic display," in *IEEE Virtual Reality Conference (VR 2006)*, Washington, DC, USA, 2006, p. 57. 1

[4] T. Balogh, P. Kovacs, and A. Barsi, "Holovizio 3d display system," in *3DTV Conference, 2007*, May 2007, pp. 1–4. 1, 5

[5] T. Balogh, "Method and apparatus for displaying 3d images," Patent US6 999 071 B2, 2006. 1

[6] W. Matusik and H. Pfister, "3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 814–824, 2004. 3

[7] R. Yang, X. Huang, S. Li, and C. Jaynes, "Toward the light field display: Autostereoscopic rendering via a cluster of projectors," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, pp. 84–96, 2008. 3

[8] T. Balogh and P. Kovács, "Real-time 3D light field transmission," in *SPIE*, vol. 7724, 2010, p. 5. 3, 5

[9] F. Marton, E. Gobbetti, F. Bettio, J. Guitian, and R. Pintus, "A real-time coarse-to-fine multiview capture system for all-in-focus rendering on a light-field display," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, May 2011, pp. 1–4. 3

[10] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3d," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 75:1–75:10, Jul. 2010. 3, 13

[11] D. Shreiner and T. K. O. A. W. Group, *OpenGL Programming Guide: The Official Guide to Learning OpenGL, Versions 3.0 and 3.1*, 7th ed. Addison-Wesley Professional, 2009. 4

[12] J. Nickolls, I. Buck, M. Garland, and K. Skadron, "Scalable parallel programming with cuda," *Queue*, vol. 6, no. 2, pp. 40–53, Mar. 2008. 4

[13] Leap Motion Inc. (2012-2014) Leap motion developer portal. [Online]. Available: https://developer.leapmotion.com/ 4

[14] M. Agus, E. Gobbetti, A. Jaspe Villanueva, G. Pintore, and R. Pintus, "Automatic geometric calibration of projector-based light-field displays," in *Proc. Eurovis Short Papers*, June 2013, pp. 1–5. 5

[15] M. C. Frederic Dufaux, Beatrice Pesquet-Popescu, *Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering*. Wiley, May 2013. 13

[16] A. Maimone and H. 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. 13

[17] B. Petit, J.-D. Lesage, C. Menier, J. Allard, J.-S. Franco, B. Raffin, E. Boyer, and F. Faure, "Multicamera real-time 3D modeling for telepresence and remote collaboration," *International Journal of Digital Multimedia Broadcasting*, vol. 2010, pp. 247 108–12, 2009. 13

[18] I. P. Howard and B. J. Rogers, *Seeing in Depth*. Oxford University Press, New York, USA, 2008. 13

[19] D. Hoffman, A. Girshick, K. Akeley, and M. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3, p. 33, 2008. 13