ANALYSIS AND OPTIMIZATION OF PIXEL USAGE OF LIGHT-FIELD CONVERSION FROM MULTI-CAMERA SETUPS TO 3D LIGHT-FIELD DISPLAYS

Péter Tamás Kovács [1, 2], *Kristóf Lackner* [1, 2], *Attila Barsi* [1], *Vamsi Kiran Adhikarla* [1,3], *Robert Bregović* [2], *Atanas Gotchev* [2]

[1]Holografika, Budapest, Hungary
[2]Department of Signal Processing, Tampere University of Technology, Tampere, Finland
[3]Pazmany Peter Catholic University, Faculty of information Technology, Budapest, Hungary

ABSTRACT

Light-field (LF) 3D displays require vast amount of views representing the original scene when using pure light-ray interpolation to convert multi-camera content to display-specific LF representation. Synthetic and real multi-camera setups are both used to feed these displays with image-based data, however the layout, number, frustum, and resolution of these cameras are mostly suboptimal. Storage and transmission of LF data is an issue, especially considering that some of the captured / rendered pixels are left unused while generating the final image. LF displays can have significantly different requirements for camera setups due to differences in Field of View (FOV), angular resolution and spatial resolution. An analysis of typical camera setups and LF display setups, and the typical patterns in pixel usage resulting from the combination of these setups are presented. Based on this analysis, an optimization method for virtual camera setups is proposed. As virtual cameras have wide range of adjustment possibilities, highly optimized setups for specific displays can be achieved.

Index Terms— 3D display, light-field display, multi-camera capture, camera rig optimization

1. INTRODUCTION

The current generation of Light-field (LF) 3D displays [1][12][13][14] typically reconstruct the equivalent of 100+ viewing directions, and up to 100 million light rays today. One of the possible input formats for such 3D displays is a multitude of images, captured by means of real or virtual cameras. By using a multi-camera setup one can capture or render the necessary number of images as well as estimate or calculate camera calibration information that allows transforming the camera pixels into a common 3D space, and consider the pixels as light-rays captured by the cameras. As there is no direct correspondence between cameras and imaging components in the LF display (even if the number of cameras matches the number of light ray emitters [2]), correspondences between light rays emitted by the display and light rays captured by cameras have to be found, and based on these correspondences, ray interpolation is used to generate the light rays generated by the display. In this case a large number of input views is assumed, and thus additional information like depth maps are not used to perform view synthesis.

In this study, Horizontal Parallax Only (HPO) LF displays are used. Such displays show different views of the scene when the viewer is moving horizontally in front of the display, however the image does not change with vertical movements. This implies that the considered capture setups do not have vertical displacements either.

In Section 2, analysis of previous work on camera setups is presented, mostly in relation with stereoscopic and multiview (MV) displays. Section 3 shows why LF displays require different, most notably wider capture setups. It also describes typical camera setups in use today for content creation. Section 4 provides an analysis of how many pixels are actually utilized during a typical LF conversion process. These lead us to the discussion of ideal camera setups in Section 5. Such ideal camera setups are only applicable in the synthetic case, and even then, only practical in special cases. Therefore, Section 6 discusses our requirements for practical camera setups, thus defining the constraints and search space that are utilized in Section 7 to generate optimized camera setups. In Section 8, the effectiveness of the optimized camera setups are analyzed, and Section 9 concludes the paper.

2. RELATION TO PRIOR WORK

Content creation for stereoscopic 3D displays is mostly concerned with providing the human visual system with the two images directly presented to the eyes [3] without causing discomfort. Stereoscopic 3D shooting is nowadays assisted with automatic tools to ensure that the cameras are properly aligned, that disparity range and convergence are

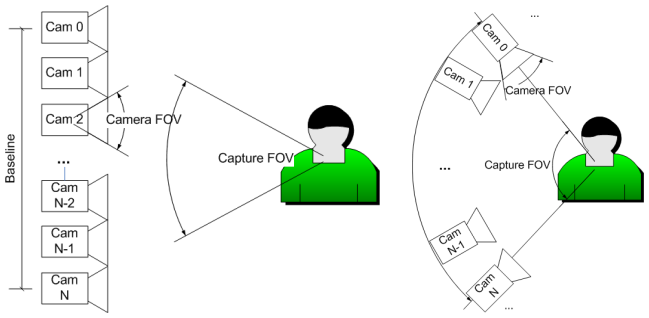

Figure 1: Left: Linear camera setup Right: Arc camera setup

within the desired limits [4], and tools that assist the adjustment of stereoscopic content after it has been shot [5].

MV 3D displays present multiple views directly (two of which are visible at the same time with the two eyes), thus, generating the necessary views for such displays is similar to stereoscopic content creation in the sense that two selected views, which are supposed to be seen at the same time, should obey the same rules as stereoscopic content. However the views are created over a bigger baseline (the distance between leftmost and rightmost camera) [6][7].

Current LF content creation and conversion tools [8][9] use multiple-camera setups (linear, arc), and perform ray interpolation based on the multitude of images and their supplementary information of the cameras used (position, orientation, FOV, distortion), generating a display-specific LF. However, the authors are not aware of any literature that discusses camera setups used for content generation for LF displays.

3. LF DISPLAY AND CAMERA SETUPS

LF displays require substantially different content compared to those required by stereoscopic and MV displays. The reasons for this are twofold.

First, there is a large difference in viewing angles: due to the displacement between human eyes, the viewing angle reproduced by stereoscopic displays is typically fairly small (few degrees). MV displays, as they provide some amount of motion parallax, reproduce a slightly wider viewing angle, thus requiring a bigger capture baseline. However, this baseline is far from the baseline required by LF displays that have viewing angles between 45º and 180º, 70º being a typical value. To avoid view extrapolation, capture setups shall reflect this wide angle, resulting in wide capture setups. In practice, camera setups based on rule of thumb have been used. This typically means a linear or arc camera setup (see Figure 1), the capture FOV which roughly corresponds to the display's FOV, and the number of cameras matching or closely matching the number of directions reproduced by the LF display. In this case, capture FOV means the angle between the leftmost and rightmost cameras as visible from the scene center, as

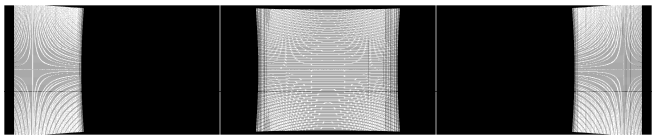

Figure 2: Pixels used during LF conversion from camera number 15, 45 and 75, respectively. Camera setup is 45º arc with 0.5º angular resolution (91 cameras), LF display has 45º FOV. Pixels marked with white are used.

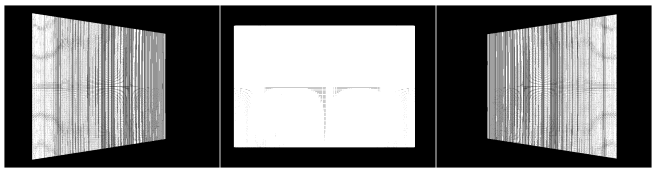

Figure 3: Pixels used during LF conversion from camera number 45, 90 and 135, respectively. Camera setup is 180º arc with 1º angular resolution (181 cameras), LF display has 180º FOV.

opposed to the opening angle of the individual cameras. Typical examples include a 112-camera linear rig, a 180-degree, 180-camera arc rig, and a 45 degree 90-camera rig, each targeting specific LF display layouts.

Second, there is no direct correspondence between cameras and viewing directions, thus the captured views are not used "as is", but have to go through ray interpolation.

4. ANALYSIS OF PIXEL USAGE

As it has been noticed [2] and exploited [11] earlier, such simple camera setups result in suboptimal usage of pixels. Typically only portions of the rendered images are actually used for generating the displayed light field, and the rest of pixels are left unused. Figure 2 and Figure 3 show some typical patterns of pixel usage. Pixels which are used are shown in white, while black areas are unused. From these figures it is clear that the ratio of used pixels is relatively low especially for the side cameras, and thus it is expected that the camera layouts could be improved to enhance the efficiency of both real and synthetic content generation in terms of having the best possible ratio of used pixels.

5. IDEAL CAMERA SETUPS

Based on the above, an ideal camera setup would be one which has many single-pixel sensors exactly matching each emitted ray, and capturing light in the required direction.

Such a set of pixels / light rays can be calculated in a virtual environment provided that rendering single pixels does not have much overhead (for example, via ray tracing [15]). However it is unusual to use such a complex camera setup in a rendering tool due to the scene set-up time typically associated with rendering each image (even when that image is a single pixel). When using real cameras to capture live scenes, such setups consisting of millions of

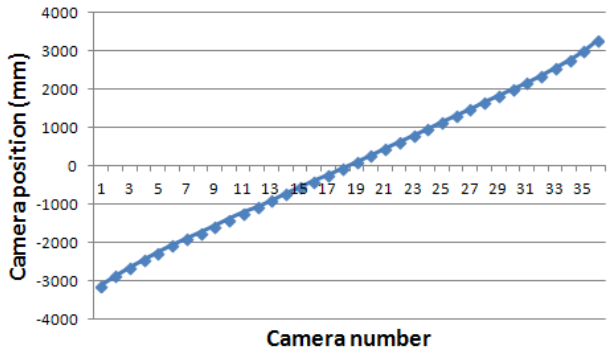

Figure 4: Horizontal position of 36 cameras after optimization for a sample 36-channel LF display

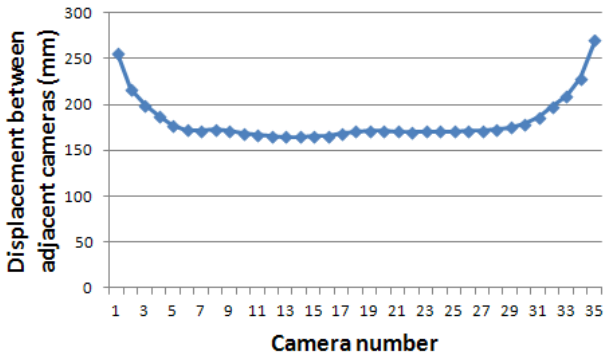

Figure 5: Horizontal displacement between 36 cameras after optimization for a sample 36-channel LF display

6. PRACTICAL CAMERA SETUPS

distinct sensors are clearly out of scope. The desired camera setups rather consist of some tens to hundred cameras (on the order of the number of the directions reconstructed by the display), but are arranged so that the captured pixels are better utilized, and also better match the individual displayed rays, resulting in less interpolated values.

Autodesk 3ds max has been used as a use case for synthetic content generation, as it is commonly used for creating content for 3D displays. While it is not possible to render arbitrary rays by overriding the ray generation step in the rendering engines bundled with 3ds max, it is possible to generate a camera setup consisting of an arbitrary number of cameras, each having custom (even highly asymmetric) FOV and custom resolution. These obviate the need to move the cameras out of the linear setup, as the same effect can be achieved by adjusting the FOV. Such arrangements can be utilized to create highly optimized synthetic camera rigs that render the 3D scene from multiple viewpoints.

7. OPTIMIZING CAMERA SETUPS

As a closed form solution for optimal camera positions could not be derived for real LF displays, optimization has been performed for synthetic camera setups. The set of rays emitted by the display in question has been generated. For simplicity, 1 to 1 matching between the display's physical space and the captured real or synthetic scene is assumed (that is, the scene is depicted in its real scale). A linear camera system model is used with a fixed number of cameras, optimize camera positions and calculate all other parameters. The rays to be captured are those that start from the display and cross the line where the viewer is assumed to be moving [10].

The objective function is defined so that for each displayed ray, the closest matching camera is found, and the distance determined. The sum of squared differences for the distance between all rays and the closest camera is minimized. The ParadisEO metaheuristics framework [16] has been utilized to implement a genetic algorithm that, starting from a large population of randomized linear camera setups, finds the optimal solution that best satisfies the objective function. As the vector describing camera positions is real valued, the genetic algorithm uses mutations to shift the camera positions, and crossovers to combine camera setups. To make convergence faster, the amount with which positions are updated is gradually decreased during the optimization. That is, cameras are randomly displaced with a bigger offset, and when the optimization cannot generate a better population over many iterations, the extent of shifts is decreased, similar to simulated annealing.

The optimized camera positions reflect the slightly higher ray density in the center of the FOV, thus cameras are placed more densely in this area. Figure 4 shows the horizontal positions of the cameras, while Figure 5 shows the displacement between adjacent cameras over the linear rig. This suggests that camera setups better than the equidistant linear setup can be found, and that cameras can be slightly sparser at the sides compared to the center.

After the position of cameras is optimized, the FOV of cameras is determined so that their leftmost and rightmost rays horizontally enclose the display's screen, as rays outside that area are clearly not used. An example of such a setup with calculated FOVs is shown on Figure 6. The vertices on the top of the figure represent the 36 cameras placed, the blue lines represent the edges of the cone captured by each camera, and the red box is the volume of the display. As visible from the figure, the sides of the captured frustum correspond to the sides of the screen of the LF display.

Once the geometry of the capture cameras is found, the desired resolution of the cameras is determined to avoid oversampling the scene by rendering many pixels, which are then left unused. If the resolution of the rendered image is higher than necessary, the unused pixels appear as holes in the pixel usage patterns shown on Figure 2 and Figure 3. The same tool that generates pixel usage maps is capable of determining how many times each pixel has been used during the LF conversion process. If pixels are read multiple times, that is caused by the resolution of the rendered image

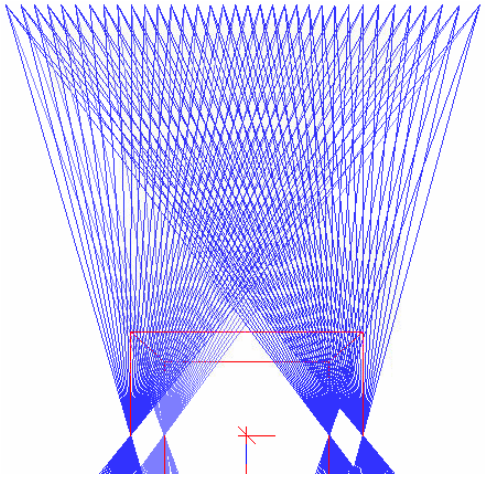

Figure 6: Camera FOVs enclosing the LF display's screen. The top-left, top-right, bottom-left and bottom-right captured ray of each camera is visualized

being lower than desired. Therefore, by summing the number of used pixels over a horizontal line of the camera, a good approximation of the needed horizontal camera resolution can be found,and the virtual camera can be configured to render just as many pixels as necessary.

8. RESULTS

The resulting camera setups can improve the utilization of rendered and captured pixels during LF conversion, therefore improving the efficiency of the content rendering process. As shown on Figure 7, the utilization of pixels is high even in side cameras after the optimization, which is visible from the lack of solid black areas.

The generated camera setup has been created with the assumption that the physical scene is represented in its full size. However, LF displays can be used to visualize scenes of different physical sizes, which are controlled during the LF conversion process with the Region of Interest (ROI) box: the ROI box represents the volume that is reconstructed by the 3D display. Assuming that the aspect ratio of the ROI box does not change, the optimized camera setup determined for the display's real screen size can be used for capturing scenes with a different scale after rescaling the camera system, therefore the camera setup need to be calculated only once for a specific LF display, and can be applied to any scene.

9. CONCLUSIONS AND FUTURE WORK

The presented results will be included in our LF content generation tool chain [8], complementing the 3ds max tools with optimized non-equidistant linear camera setups with sheared frustums, enabling shorter rendering times for synthetic content. The natural next step is to extend our analysis to real camera setups. In capture setups involving

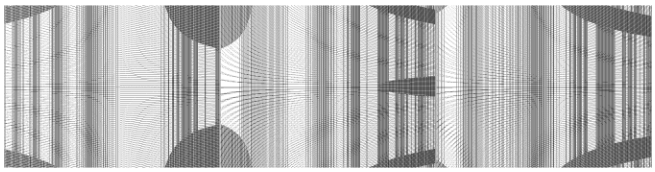

Figure 7: Pixels used during LF conversion from the optimized 36-camera setup. Camera number 9, 20 and 30 are shown.

real cameras, optimization possibilities are more restricted. The number of cameras is scalable, but within budgetary limits, and resolution of the cameras has an upper limit imposed by the resolution of the sensor.

With cameras having a lens mount, FOV can be selected from a set of available lenses, however the FOV is typically symmetric and equal among all cameras. While cropping the captured images with Area Of Interest setting is possible in some cameras (thus creating smaller capture frustums), this does not bring practical benefits in our case. In the past a 27-camera rig has been used [2] for live LF capture, which has been assembled to form a 1.5m wide, equidistant, parallel linear rig. The number, resolution and FOV of these cameras is fixed, but optimizing the placement of these cameras (position and orientation) is planned to provide a better coverage for specific LF displays. In case of real cameras, using an arc (or other nonlinear) camera setup can be advantageous to increase the coverage of the cameras to compensate for the lack of custom FOV settings. In case of modeling and optimizing rigs of real cameras, more complex modeling and optimization is necessary, also considering the vertical position and direction of rays.

The approach used for optimizing synthetic camera setups does improve the utilization of pixels, however future work should prove that this also results in improved perceived image quality for the same amount of cameras. As an extreme example, if the number of cameras is small, and the camera spacing becomes bigger than a stereo baseline, viewers may prefer having the small number of cameras in the center, and lose the sides of the FOV, instead of not experiencing a stereoscopic effect at all. Subjective tests will be performed to check that the perceived quality of rendered LFs does improve with the optimized camera setups, keeping the number of rendered pixels constant.

10. 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.

The research leading to these results has also received funding from the DIVA Marie Curie Action of the People programme of the European Unions Seventh Framework Programme under REA grant agreement 290227.

11. REFERENCES

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

[2] T. Balogh, P. T. Kovács, "Real-time 3D light field transmission". *Proc. SPIE 7724, Real-Time Image and Video Processing* 2010, 772406 (May 04, 2010); doi:10.1117/12.854571.

[3] F. Zilly, J. Kluger, P. Kauff, "Production Rules for Stereo Acquisition," *Proceedings of the IEEE*, vol.99, no.4, pp.590,606, April 2011. doi: 10.1109/JPROC.2010.2095810

[4] F. Zilly, K. Muller, P. Eisert, P. Kauff, "The Stereoscopic Analyzer — An image-based assistance tool for stereo shooting and 3D production," Image Processing (ICIP), 2010 17th IEEE International Conference on , vol., no., pp.4029,4032, 26-29 Sept. 2010. doi: 10.1109/ICIP.2010.5649828

[5] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, M. Gross, "Nonlinear disparity mapping for stereoscopic 3D". In *ACM SIGGRAPH 2010* papers (SIGGRAPH '10), Hugues Hoppe (Ed.). ACM, New York, NY, USA, , Article 75 , 10 pages. doi: 10.1145/1833349.1778812

[6] A. Boev, K. Raunio, M. Georgiev, A. Gotchev, K. Egiazarian, "Opengl-Based Control of Semi-Active 3D Display," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008 , vol., no., pp.125,128, 28-30 May 2008. doi: 10.1109/3DTV.2008.4547824

[7] C. González, J. Martínez Sotoca, F. Pla, M. Chover, "Synthetic content generation for auto-stereoscopic displays", In *J. Multimedia Tools and Applications*, Feb. 2013. doi: 10.1007/s11042-012-1348-x

[8] Holografika Software System. http://www.holografika.com/Software-and-system-compatibility/Software-system-and-compatibility.html Visited 06 June 2014

[9] J. Park, D. Nam, S. Y. Choi, J.-H. Lee, D. S. Park, C. Y. Kim, "Light field rendering of multi-view contents for high density light field 3D display". *SID Symposium Digest of Technical Papers*, 44: 667–670, 2013. doi: 10.1002/j.2168-0159.2013.tb06300.x

[10] A. Said, B. Culbertson, "Virtual object distortions in 3D displays with only horizontal parallax," *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, vol., no., pp.1,6, 11-15 July 2011. doi: 10.1109/ICME.2011.6012200

[11] V. K. Adhikarla, ABM T. Islam, P. T. Kovács, O. Staadt, "Fast and Efficient Data Reduction Approach for Multi-Camera Light-Field Display Telepresence Systems". In *Proceedings 3DTV Conference*. October 2013

[12] J.-H. Lee, J. Park, D. Nam, S. Y. Choi, D.-S. Park, C. Y Kim, "Optimal Projector Configuration Design for 300-Mpixel Light-Field 3D Display", *SID Symposium Digest of Technical Papers*, 44: 400–403. doi: 10.1002/j.2168-0159.2013.tb06231.x

[13] G. Wetzstein, D. Lanman, M. Hirsch, R. Raskar, "Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting", In *Proc. SIGGRAPH 2012*

[14] D. Lanman, D. Luebke, "Near-Eye Light Field Displays", In *ACM Transactions on Graphics (TOG)*, Volume 32 Issue 6, November 2013 (Proceedings of SIGGRAPH Asia), November 2013

[15] J. A. Iglesias Guitián, E. Gobbetti, F. Marton, "View-dependent Exploration of Massive Volumetric Models on Large Scale Light Field Displays". *The Visual Computer*, 26(6--8): 1037-1047, 2010

[16] S. Cahon, N. Melab and E-G. Talbi, "ParadisEO: A Framework for the Reusable Design of Parallel and Distributed Metaheuristics", *Journal of Heuristics*, vol. 10(3), pp.357-380, May 2004.