3D Capturing using Multi-Camera Rigs, Real-time Depth Estimation and Depth-based Content Creation for Multi-view and Light-field Auto-Stereoscopic Displays

Peter Tamas Kovacs[1], Frederik Zilly[2]

Overview

The wide variety of commercially available and emerging 3D displays - such as stereoscopic, multi-view and light-field displays - makes content creation challenging as each displays technology requires a different number of views available of the scene. As consequence, the content creation pipelines differ considerably and involve different camera setups such as beam-splitter rigs with small baselines and high quality cameras used for stereo 3D productions or camera arrays for auto-stereoscopic displays which usually use small lower quality cameras in a side-by-side arrangement. Converting content shot for a specific display technology into a different format usually impairs the image quality and is very labor-intensive.

Against this background a generic method for capturing and rendering live 3D footage for stereoscopic, multi-view and light-field displays is presented. The system consists of a wide-baseline multi-camera rig, a camera assistance system, a real-time depth estimator, a real-time view generation and rendering engine, and multiple displays, one multi-view auto-stereoscopic and a light-field display. The system features several innovative components: the professional-grade multi-camera assistance and calibration system; a real-time depth estimator producing convincing depth maps; a real-time and generic depth-image based rendering (DIBR) engine that is suitable for generating imagery for a range of 3D displays; and the largest auto-stereoscopic light-field display to date.

The Experience for SIGGRAPH Attendees

The exhibited system will allow attendants (watching one of the 3D displays) to see other attendants (standing in front of the multi-camera rig) in 3D, without glasses, with the possibility to walk around the perceived 3D image, experience smooth motion parallax and large depth-range. Visitors are interested in the technical details will also have the opportunity to see the camera calibration and assistance system in operation, as well as the output of the real-time depth estimation system (as gray-scale depth maps), and also other 3D content – including interactive 3D applications – on the light-field display.

Core Technical Innovations

The first 4-camera rig built from professional HD cameras (Sony HDC-P1) is presented. In order to support a wide range of 3D displays while being backwards compatible with already established stereo displays, two of the captured views are shot using a beam-splitter which allows showing them directly on a stereoscopic 3D display without any further processing. The additional satellite cameras placed outside the mirror box provide the information that is needed to create a generic depth-based 3D representation format and content for other wide baseline applications.

The first professional-grade multi-camera calibration and assistance system is demonstrated as part of the system. The precise calibration of a multi-camera system is a demanding task as the system has many degrees of freedom. However, to keep the system easy to use and robust, a dedicated assistance system has been developed as an extension of the stereoscopic analyzer (STAN) towards a quadrifocal setup. For the multi-camera setup we bring all four cameras in a position such that each pair of two cameras is rectified. The system has shown its maturity and suitability for productions during two field trials.

We present the first real-time multi-view depth estimation system based on line-recursive matching that generates depth maps for the visualization components. We have extended the existing depth estimator Hybrid Recursive Matcher (HRM) towards parallelization. Although the HRM is able to generate depth maps in real-time for smaller resolutions, its recursive structure prevents it to take advantage of multiple CPU cores. We broke the recursive structure of the HRM and limited the recursion on a line-wise level. Thus, each line can be processed in a different thread, resulting in a significant speed-up when executed on multi-core CPUs. The estimation is applied independently for the images with subsequent filtering of left/right consistent disparities. Temporal stability avoiding flickering artifacts is achieved by incorporating temporal disparity candidates in the estimation process.

The first real-time Multi-View plus Depth (MVD) based view generation & rendering system targeted for wide-baseline light-field displays is presented. The view generator renders interpolated views (between original cameras) as well as heavily extrapolated (outside the original cameras) novel views. The interpolation process detects and keeps gap area information from the content using depth layers. Extrapolation is hierarchical, using each image from the closest to the furthest. Holes are filled using information coming from the other images, where available. Inpainting techniques are used where no information is available, during which texture and structure is rendered, propagating contour gradients with prioritized matching costs.

The largest glasses-free light-field 3D display to date (140" screen diagonal) will be shown. The display presents natural 3D light-field to a larger audience on a cinema-sized screen size previously not possible with auto-stereoscopic displays. The display itself consists of a complex hardware and software system, being the first front-projected light-field 3D display, controlling 63 Mega-Pixels in total. It consists of an array of optical engines, projecting light rays onto a reflective holographic screen, in front of which viewers can see 3D content with an exceptionally wide Field Of View and depth range.

The Future of this Work

Today's glasses-based stereoscopic 3D display systems can be seen as stepping stones towards more advanced 3D display technologies. The generic Multi-View plus Depth (MVD) representation used inside the system can serve as the future 3DTV format, which is generic enough to drive a multitude of 3D displays, independent of the underlying technology. The presented approach is also in line with MPEG's efforts towards future 3DTV formats.

Acknowledgments

The demonstrated system is based on work partially supported by the MUSCADE European FP7 project (EU-FP7-247010).

[1] Holografika
[2] Fraunhofer HHI

SIGGRAPH 2012, Los Angeles, California, August 5 – 9, 2012.
ISBN 978-1-4503-1435-0/12/0008