Automatic Multiview Synthesis – Prototype Demo

Michael Schaffner [1,2], Frank K. Gürkaynak [1], Hubert Kaeslin [1], Luca Benini [1,3], Aljosa Smolic [2]

[1] *Integrated Systems Laboratory, ETH Zürich, Switzerland*
[2] *Disney Research Zurich, Switzerland*
[3] *Università di Bologna, Italy*

{schaffner,kgf,kaeslin,benini}@iis.ee.ethz.ch, smolic@disneyresearch.com

Overview. Today, most commercially available 3D display systems require the viewers to wear some sort of shutter- or polarization glasses, which is often regarded as inconvenience. Ideally, a 3D display system should not require the users to wear additional gear. In fact, the optimum would be a display that replicates the original light-field of a scene. So-called *multiview autostereoscopic displays* (MADs) represent a step in this direction, as they are able to project several views of a scene simultaneously, enabling a glasses-free 3D experience and a limited motion parallax effect in horizontal direction.

However, appealing content creation for such displays is a difficult task. Moreover, storage and transmission of high definition content with more than two views is costly and even infeasible in some cases. The fact that each MAD model has different parameters (#views, #viewing angles, etc.) exacerbates these problems. In order to bridge this *content-display gap*, so-called *multiview synthesis* (MVS) methods have been developed over the past couple of years, which can generate several virtual views from a small set of input views.

MVS is computationally intensive, yet it should run efficiently in real-time and should be portable to end-user devices to explore the full potential. In this work, we devise an efficient hardware architecture of a complete, *image-domain-warping*-based MVS pipeline, which is able to synthesize content for an 8-view full-HD display from full-HD S3D input at 30 fps. Our hybrid FPGA/ASIC prototype is one of the first real-time systems which is entirely implemented in hardware. Such a dedicated hardware accelerator enables portable and energy efficient MVS, which are both essential properties when considering a deployment in consumer electronic devices. Our system comprises all processing steps including S3D video analysis, calculation of the warp transforms, rendering, and anti-alias filtering/interleaving for the MAD. The algorithms involved in these steps have all been revisited and jointly optimized with their corresponding hardware architectures. The developed hardware IP could be integrated into *systems-on-chip* (SoCs) for 3D TV sets or mobile devices, where it would serve as a power-efficient hardware accelerator.

MVS Based on Image Domain Warping. Common MVS methods are based on *depth image based rendering* (DIBR), where a dense depth map of the scene is used to reproject the input images to new viewpoints. Although physically correct, this approach requires accurate depth maps and additional inpainting of disoccluded regions. Our prototype uses an alternative conversion concept developed at Disney Research Zurich by Smolic, *et al.* which is based on *image domain warping* (IDW). The IDW framework allows to locally warp image regions via a non-linear, two-dimensional transformation which is obtained by solving a least squares problem. The constraints for this problem are formulated using image features extracted from the input images. This technique is robust, accurate and efficient, as it does not rely on pixel-dense depth maps, but only on sparse point correspondences. Further, no inpainting is required as, in contrast to DIBR, no disocclusions are created.

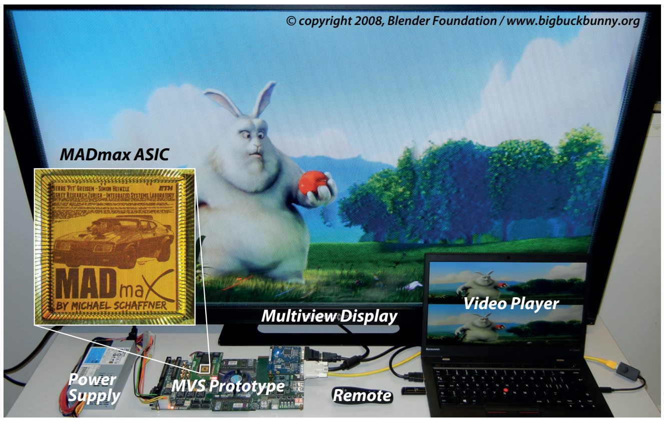

Fig. 1. Photograph of the running system. The MAD used in this setup is an Alioscopy HD 47 LV.

Performance. In this demonstration setup, the system renders 8 novel views from top/bottom packed S3D content with $2\times$1080p resolution at 30 fps. The output views are interleaved into one 1080p frame which can be displayed on the MAD. The system does not require any coded warping information since these transformations are calculated in real-time.

Demonstration. Attendants will be able to watch full-length movies which get converted from S3D to multiview in real-time. The MAD will be a commercial-grade full-HD display with 8 views, and the hardware prototype - comprising an Altera Stratix IV FPGA board and our own rendering ASIC - will be placed inside a plexiglass-box such that interested visitors may draw near and examine it closely. Besides the rendered output, the system can also display the extracted video features. Further, viewing parameters such as the strength of the depth effect (depth volume) can be adjusted interactively using a remote control.