Real-Time Stereo Video Processing

Personnel: Pierre Greisen
Simon Heinzle
Andreas Burg

Funding: Disney HPCC

Partners: Disney

The recent revival of stereoscopic 3D (S3D) content has been enabled by the advances in digital video processing and distribution. However, producing and displaying visually appealing S3D content is still a non-trivial process, requiring computationally intensive pre- and post-processing steps. The doubled data bandwidth for stereo acquisition is not the only requirement. A more significant challenge is the ability to set the stereo camera parameters in real-time and to match the resulting two video streams such as to obtain a pleasing 3D experience.

In this project we investigate and implement the crucial steps of a real-time stereo video processing pipeline. Target application will be a stereoscopic acquisition system, able to process and react to incoming video streams on the fly. Each stream is (at least) full HD, i.e., it has a resolution of 1920x1080 at 30 fps.

So far, the implemented pre-processing steps comprise camera synchronization, noise removal and color interpolation. Stereo correction steps are necessary to compensate for photometric and geometric discrepancies. Depth information can be extracted from the scene via depth matching algorithms. More advanced computer vision algorithms, such as tracking and object recognition, can be added on top of the processing pipeline in view of a smart stereo system. The algorithms and the camera control run on dedicated digital hardware (FPGAs), graphics processors (GPUs) and central processing units, whichever is best suited for the particular task.

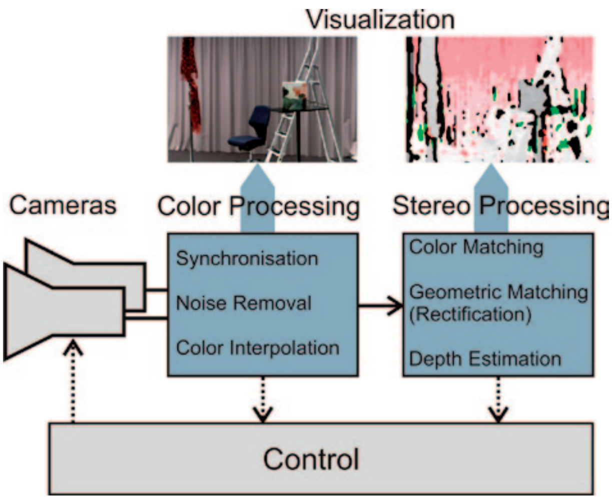


Illustration of the stereo system: first, the two camera streams are synchronized, noise corrected, and color interpolated; next, more advanced algorithms extract information from the scene (a depth maps in the illustration).

EWA Splatting Core for Video Retargeting

Personnel: Marian Runo, Michael Schaffner (students)
Pierre Greisen, Frank Gürkaynak
Simon Heinzle

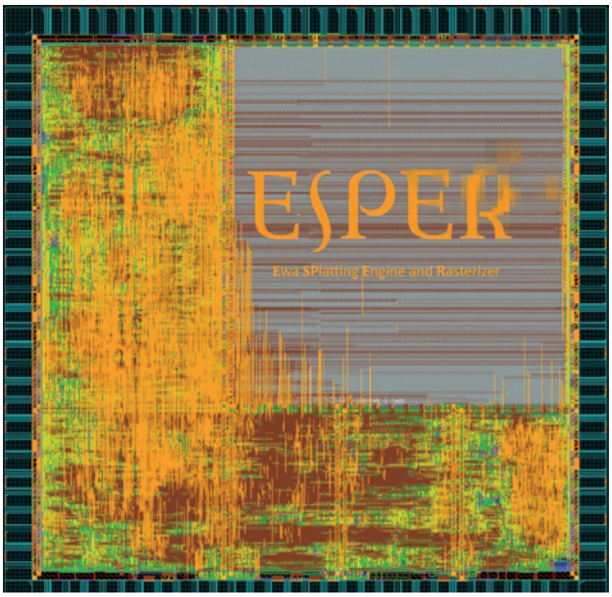
Funding: ETHZ

Partners: Disney

Content-aware image transformations often require non-linear warping: each pixel of a source image is transformed with a different (non-linear) transformation to a destination image. A prominent application for non-linear warping is video retargeting, where an input video stream with a certain image size needs to be transformed into a stream with a different aspect ratio and resolution, without distorting the salient image regions. One technique to perform such spatially varying transformations is elliptical weighted average (EWA) splatting, where each source pixel is spanned by a Gaussian kernel and then transformed into a destination Gaussian, which is re-sampled. The new samples are then accumulated.

In this semester project, a hardware architecture that performs EWA splatting has been designed and implemented in digital hardware. The architecture renders the input pixels and accumulates their contributions. The rendering step consists in evaluating the Gaussian kernels in the destination space; since different non-linear functions are involved, number precision considerations are important. The accumulation part is easy in terms of calculations, but expensive in terms of memory bandwidth. A dedicated memory strategy has been developed to alleviate memory bandwidth requirements.

The architecture was implemented for an ASIC prototype to be fabricated in 180 nm CMOS technology. It is able to process WPAL (1024x576) images in real time and it is scalable to larger resolutions, such as full HD.



Layout of the EWA splatting VLSI chip, sent for fabrication in 180 nm CMOS technology.