

# DISPARITY-AWARE STEREO 3D PRODUCTION TOOLS

A.Smolic\*, S. Poulakos\*<sup>†</sup>, S. Heinzle\*, P. Greisen\*<sup>†</sup>, M. Lang\*<sup>†</sup>, A. Hornung\*, M. Farre\*, N. Stefanoski\*, O. Wang\*, L. Schnyder\*<sup>†</sup>, R. Monroy\*<sup>†</sup>, M. Gross\*<sup>†</sup>

\*Disney Research Zurich, Switzerland, smolic@disneyresearch.com

<sup>†</sup>ETH Zurich, Switzerland

## Abstract

*Stereoscopic 3D (S3D) has reached wide levels of adoption in consumer and professional markets. However, production of high quality S3D content is still a difficult and expensive art. Various S3D production tools and systems have been released recently to assist high quality content creation. This paper presents a number of such algorithms, tools and systems developed at Disney Research Zurich, which all make use of disparity-aware processing.*

**Keywords:** Stereoscopic 3D, disparity, production technology.

## 1 Introduction

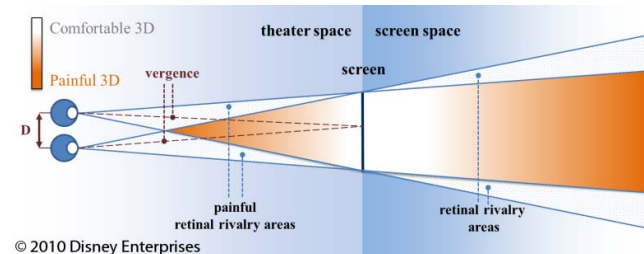
Stereoscopic 3D (S3D) has reached wide levels of adoption by consumer and professional markets. S3D cinema has become mainstream, 3DTVs and blu-ray equipment are available to consumers, S3D broadcast channels are on air, S3D cameras allow consumers to create content, gamers can enjoy playing in a new dimension, mobile devices are equipped with 3D displays, and so on. The current success of S3D technology is due to the fact that technology and understanding about content creation has reached a high level of maturity, ensuring a high quality user experience in most cases.

However, production of high quality S3D content is still a difficult and expensive art [1]. S3D production has to consider fundamentals of human 3D perception as well as capabilities and limitations of 3D displays, and combine them with artistic intent. Section 2 gives a brief overview of related issues. Highly skilled individuals, experienced in the art of S3D have to contribute in many ways to the production processes. In order to assist their challenging job and to ensure best possible experience for the audience, various S3D production tools and systems have been released recently.

Such advanced S3D production tools, algorithms, and systems are also a main research focus at Disney Research Zurich. This paper presents the results of this work. A key component of all of these algorithms is awareness of disparity or depth composition of the input S3D content. In some cases sparse, but highly robust and accurate disparity information is estimated automatically. Other algorithms estimate dense disparity or depth maps. User interaction is part of some of the concepts others are fully automatic.

## 2 Fundamentals and limitations of stereo 3D perception

Our technology presented in the following sections is motivated by the desire to produce content that is comfortable and enjoyable to view. This is fundamentally achieved by restricting disparity to a comfortable range as visualized in **Figure 1**. The illustrated range from Comfortable 3D to Painful 3D reflects the decoupling that occurs between eye-vergence and accommodation. As eye fixation moves farther from the screen plane, it can become more difficult to maintain focus on the screen. This unnatural decoupling has been demonstrated to produce viewer discomfort, fatigue and hinder depth discrimination [2]. The function expressing the zone of comfort is an area of active research, which includes understanding the response of the human visual system to the viewing conditions and image content [3].



**Figure 1: Illustration of the stereoscopic comfort zone**

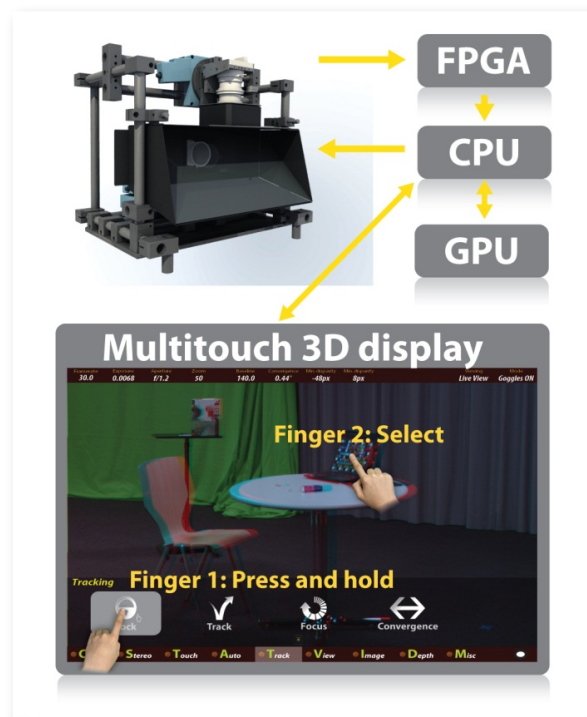
The illusion of stereoscopic depth also introduces potential for perceived depth conflicts and misinterpretations. The binocular disparity cue provided by stereoscopic images can conflict with other cues to depth. For example, stereoscopic window violations (see painful retinal rivalry areas of **Figure 1**) conflict with the strongest depth cue: occlusion. Hyper- and hypo-stereoscopy can conflict with perspective and size cues. Other unnatural phenomena, such as image distortions caused by stereoscopic crosstalk, can inhibit depth interpretation and reduce subjective assessments of image quality and comfort [4].

S3D content creation has to provide a pleasing and expressive mapping of the broad real 3D world into the limited stereoscopic comfort zone, to create the stereo depth illusion. The fact that this is a difficult art, was the motivation for development of the tools, algorithms, and systems, which are described in the following sections of this paper.

### 3 Computational stereo camera

Shooting high-quality stereoscopic live video content remains an art that has been mastered only by a small group of individuals. More specifically, the difficulty arises from the fact that in addition to setting traditional camera parameters (such as zoom, shutter speed, aperture, and focus), camera interaxial distance and convergence have to be set correctly to create the intended depth effect. Adjusting all these parameters for complex dynamically changing scenes poses additional challenges. Furthermore, scene cuts and shot framing have to be handled appropriately in order to provide a perceptually pleasing experience. These problems become even more pronounced for live broadcast of stereo content, such as in sports applications. Capturing high-quality stereo 3D footage therefore requires very sophisticated equipment along with the craftsmanship of an experienced stereographer all of which makes the S3D production inherently difficult and expensive. The cost for S3D movie productions is estimated 10%-25% higher than for traditional productions [5].

We propose a computational stereo camera system [6] that features a closed control loop from analysis to automatic adjustments of the physical camera and rig properties. Our freely programmable architecture comprises a high-performance computational unit that analyzes the scene in real-time (e.g., by computing a dense disparity image or by tracking scene elements) and that implements knowledge from stereography to capture quality S3D video in our control loop algorithms. Since stereography is still a widely open field with a continuously evolving understanding of S3D cinematography, we designed our camera architecture as a freely reprogrammable set of processing units. This enables the utilization of different algorithms depending on the scenes, shots, or artistic intentions. In addition, we support scripting of complex operations to develop and optimize shots within the actual movie production. Thus, some of the postproduction is shifted back into the production cycle. In a live broadcast scenario scripts may be predefined and executed on demand. For efficient camera operation, we devise a set of interaction metaphors that abstract the actual camera rig operations into intuitive gestures. The operator controls the camera using a multitouch stereoscopic user interface. In addition, the interface enables monitoring the S3D content as well as the related stereo parameters instantly. In order to achieve real-time performance, we implemented our custom computational architecture combining FPGA, GPU, and CPU processing close to the sensor to achieve a low latency control loop for full HD (1920x1080) video **Figure 2**.



**Figure 2:** Our computational stereo camera system is designed for fully motorized stereo rigs. A programmable high-performance computational unit (using FPGA, CPU and GPU technology) analyzes the incoming video stream in real-time, and performs automatic rig adjustments depending on the user's setting. User input is performed on a stereoscopic touch screen, which abstracts traditional camera operation into intuitive interaction metaphors.

We built a prototype camera rig with motorized optics as well as motorized interaxial distance and convergence angles, and implemented the computational control loop on a PC based platform. We are able to demonstrate the potential of our system by using three example applications: automatic interaxial distance and convergence plane control, touch-based refocusing and convergence, and subject tracking for follow-focus and follow-convergence shots. Automatic interaxial distance and convergence plane control for a pre-set range of disparities is performed using data from the disparity estimation. Touch-based refocusing and convergence uses local template matching to estimate actual scene distances for automatic, knob-free camera operation. Subject tracking for follow-focus and follow-convergence uses the same concept, for automatic temporal refocusing and convergence **Figure 3**. Our intuitive interaction metaphors automatically abstract and replace cumbersome handling of rig parameters to achieve results that are impossible or difficult to achieve with current systems. Future work will include optimization and extension of all components, including usage of other rigs and cameras, improvement of low-level image processing algorithms, optimization of the control loop implementation and parameters, design of further intuitive interaction metaphors,

improvements to the user interface, as well as general improvements of the software (e.g., partitioning) and hardware. While our system automatically adjusts stereo parameters to limit disparities to a comfortable range, other more sophisticated concepts from stereography (e.g., framing violations, disparity gradients, flatness) still require manual interaction with the provided plug-ins. Such additional plug-ins are left for future work in order to achieve a fully automated stereographer. The basic paradigm of our design, being an efficient computational vision system incorporating advanced image analysis and high-level concepts into a real-time closed control loop, easily extends to other application scenarios as well.



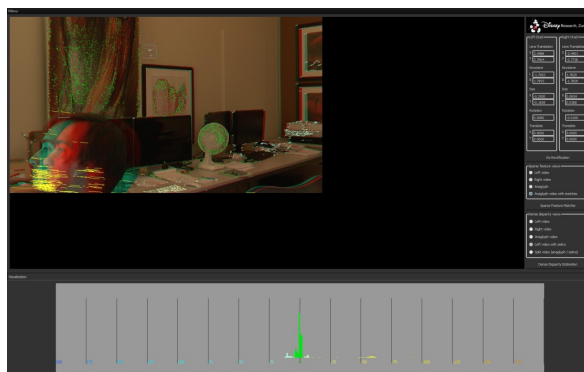
**Figure 3:** Example of automatic subject tracking for follow focus and convergence. Manual interaction is only required in the first frame (top), to initialize the subject of interest. In the following frames, our controller keeps the convergence plane and focus at the depth of the subject

#### 4 Real-time stereoscopic analyzer

On-set analysis and monitoring of stereoscopic video play an important role in S3D productions, especially when the camera rig is not equipped with automated control mechanisms as presented before. The stereographer usually constantly checks and adjusts disparity ranges and other parameters to make sure to create high quality S3D content. Such monitoring is specifically important in live broadcast applications, where no correction is possible anymore. Therefore, a variety of different stereoscopic analyzer systems has been released recently that are able to analyse the stereo parameters of the captured stereo material in real-time [7], [8].

The analysis part of our computational stereo camera system presented in the previous section provides the same functionality. We therefore extracted those modules and designed a powerful stereo analyzer. The system assists the user to detect camera and lens misalignments, and is able to remove vertical disparities as well as keystoneing automatically and in real-time. Our system furthermore analyses the horizontal disparity distribution, and warns the user in case of unpleasant settings such as exceeded disparity budgets.

Figure 4 shows an example of our user interface, including a stereo image with estimated disparities, disparity histogram, and stereo misalignment parameters. The yellow color indicates exceeded disparity range.



**Figure 4:** Our system analyses and displays the disparities, histogram, and stereo parameters in real-time

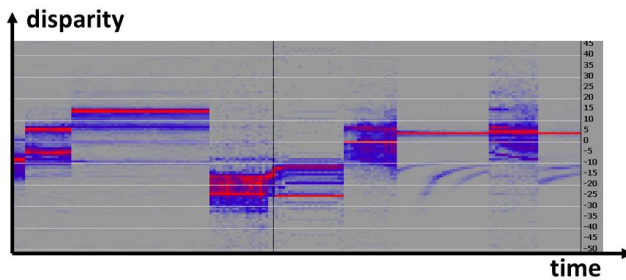
#### 5 Depth script visualization, disparity histograms

As 3D movie making becomes more popular and accepted the artistic desire to use depth as important story telling element increases. Many filmmakers carefully plan the development of depth throughout the movie. This process for scripting depth is very similar to the known art of creating color scripts, which describe the color and lighting throughout a movie. Color scripts can be very easily represented by thumbnails or color palettes (Figure 5) and the resulting movie can easily be compared against those depictions. However, an intuitive visualization for a depth script is not apparent.



**Figure 5: Parts of the color script of the PIXAR movie Up (C) PIXAR/DISNEY**

Therefore, we developed a production tool allowing for visualization of depth over individual takes or through an entire movie. **Figure 6** shows a typical output of our tool. The graph can be interpreted as a top view onto the scene over time. For every frame (along the time axis) it shows the depth histogram of that frame. The Y axis of the graph is the disparity and therefore a measurement of how far in front or behind the screen an object is perceived. The saturation of a point in the graph is proportional to the amount of objects detected at that disparity. With this representation it is easy to judge the depth distribution for individual frames or its variation during a movie segment. The color (red) shows the depth position of the most salient objects in the scene. Often, different layers of salient objects (e.g. foreground/background) can be seen simultaneously.



**Figure 6: Depth Script Visualization for an edited movie. It shows the disparity distribution for the frames over time. This example shows a cut of multiple scenes with different depth composition. The red pixels show the depth location of the salient objects.**

Our depth script visualizer tool has numerous applications. First, it is very valuable during editing. Editing a stereoscopic movie has to consider the depth of the individual scenes. This is not only required to fulfill the desired depth script but also depth continuity across cuts. Depth continuity across shots (especially for fast scene cuts) is required to avoid viewing discomfort and to reduce viewers' adaption time for changing depth compositions after a cut. With our tool an editor can use the additional depth visualizations to more easily select appropriate shots while cutting. The depth visualization tool can also be applied onto a longer movie segment or the entire movie to judge and compare the depth development of the final movie to the depth script from pre-production. It is therefore an important quality assurance instrument.

## 6 Nonlinear disparity mapping by image-domain warping

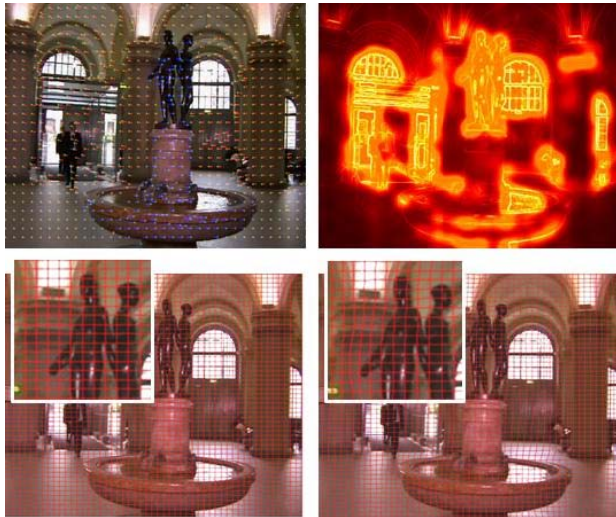
The previous sections demonstrated how stereoscopic footage can be captured and analyzed computationally in order to facilitate the production of “good” stereo. However, most often the captured stereo content still requires modification in post-production:

- Display adaptation: stereo disparities scale linearly with display size, resulting in nonlinear changes of perceived depth. In some cases, disparities can become too large to be fused stereoscopically.
- Artistic modification: by using different interaxial camera distances, depth perception can be influenced locally on a per object basis, e.g., to (de-)emphasize certain scene elements. Merging the different stereo pairs into a final consistent stereo pair, though, requires complex image compositing and processing.
- Problematic disparities: some issues might simply be overlooked during capture or impossible to fix due to physical/geometric constraints of the cameras, the captured scene, etc.

Technically, the assumed standard procedure to remap the range of stereo disparities has been to compute an accurate depth/disparity map from the input, and then synthesize a novel stereo pair with a different interaxial distance. However, the problem of computing accurate disparity maps from general stereo input is highly ill-posed. Moreover, missing scene information in disoccluded scene regions has to be “hallucinated” and seamlessly inpainted into the output images. Hitherto, no sufficiently reliable automatic solution exists for these problems, and most often one has to fall back to cumbersome manual edits. The Foundry's Nuke and Occula software for stereoscopic content processing are available products for that purpose [9].

We developed a novel approach for remapping the disparity range of a stereoscopic image pair that is based on image-domain warping instead of the classical depth-based view interpolation [10]. The basic idea is to compute a sparse set of feature correspondences between the two input images, which effectively span its respective range of stereo disparities (similar to the stereoscopic analysis described in the previous sections). Remapping the range of disparities to a different target range can be interpreted as scaling the pairwise distances between this sparse set of stereo correspondences. The fundamental insight of our method is that the perceived depth of a stereo pair can be effectively changed by smoothly “deforming” the input images subject to these rescaled, sparse disparity constraints. So instead of the classical approaches based on dense depth maps and new-view synthesis, our algorithm synthesizes stereoscopic output with a prescribed target disparity range by deforming the input images, using the modified stereo correspondences as positional constraints for the deformation. Our image warping procedure ensures that the deformations stay below a perceptually noticeable

level by employing a pre-warp image saliency analysis. The image deformation for changing stereo disparities is then applied to the least salient image regions. Additional temporal constraints on the image warps ensure coherent processing for stereoscopic video. The warping procedure is illustrated in *Figure 7*.



*Figure 7: Stereoscopic depth editing by image-domain warping. Our method first computes a sparse set of stereo correspondences between a left and a right stereo input image. The disparities between these correspondences are illustrated in the leftmost image by small line segments. In addition, our method analyses the visual importance of different image regions. It then computes a warp for both input views that remaps the perceived stereoscopic depth. The necessary image deformations are hidden in visually less important regions. The red grid-lines in the close-up images illustrate how the warp changes the image content in order to remap the stereoscopic image disparities to a desired disparity range.*

Our technique is applicable to all three before-mentioned application scenarios such as stereoscopic error correction, optimization of stereo content for different display types and sizes, and artistic modification of content. *Figure 8* illustrates two examples. The fundamental advantage of this approach over previous methods is that it does not require the generation of a dense depth map, but it is sufficient to compute sparse stereo correspondences only. Furthermore, we avoid associated problems such as stereoscopically consistent inpainting, since we deform the input images in order to change disparities rather than synthesizing completely new views. This is possible since in many practically relevant cases for disparity remapping, the required change is only in the order of a few pixels. Several user studies have already confirmed that our algorithm it is able to effectively change the perceived stereoscopic depth without compromising perceived realism and without leading to noticeable visual artifacts.



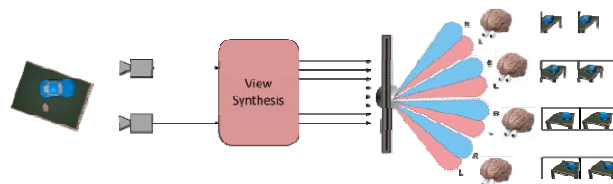
© Car sequence: KUK Filmproduktion GmbH

*Figure 8: Examples of non-linear disparity mapping, left originals, right modified. The cow in the top example is shifted back in depth, while not changing background depth. The car in the bottom example causes a window violation, which is corrected by pushing it back in depth.*

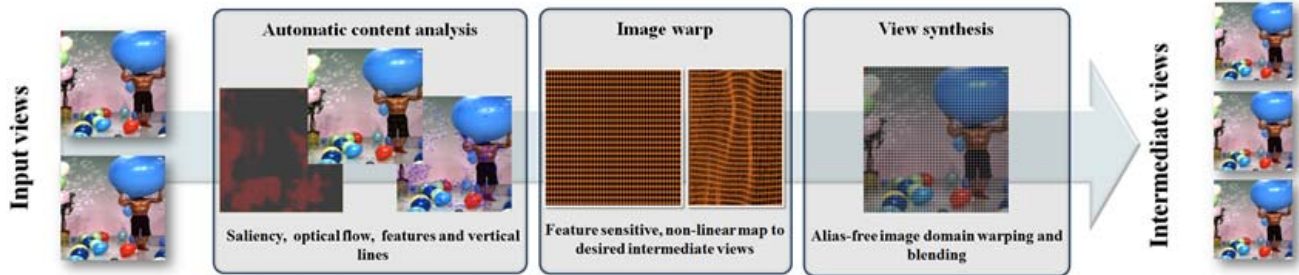
In conclusion, stereoscopic image-warping has turned out to be an effective and practical solution to one of the central problems in stereoscopic post-production. Moreover, several follow-up works have shown that it bears potential for an even wider set of applications. The following sections on stereo to multi-view conversion or 2D to 3D conversion demonstrate some of these additional applications.

## 7 Stereo to multiview conversion

Although S3D is widely adopted today, the necessity to wear glasses and the limitation to two views, which prevents the perception of all natural 3D cues, is often regarded as the main obstacle of today's mainstream 3D systems. These two shortcomings of S3D are addressed by multiview autostereoscopic displays (MAD). The right hand side of *Figure 9* illustrates the principle. MADs support motion parallax viewing in a limited range and do not require glasses. However, content creation for MADs is still widely unresolved. Capturing the required  $N$  views directly is impractical or even impossible due to restrictions on camera placement. Therefore the typical approach is to capture  $M < N$  views and to generate the necessary views in between by synthesis [11] (*Figure 9*). The most important special case in content creation for MAD is  $M=2$ , i.e. stereo input.



*Figure 9: Illustration of the pipeline of stereo content capture, view synthesis, display and perception*



**Figure 10: Stereo to multiview conversion**

A typical approach for M-view to N-view conversion is to use depth image based rendering (DIBR). However, this approach relies on dense depth or disparity estimation, which is an ill-posed and unresolved task so far. In contrast to previous approaches which rely on DIBR, we apply image domain warping (IDW) where we rely only on sparse disparities. The overview of our approach is shown in **Figure 10**. It consists of 3 main building blocks, i.e. automatic video analysis, warp calculation and view synthesis [12]. Automatic video analysis extracts all necessary data from the input data. Instead of depth estimation we use sparse disparity features, combined with saliency and edge information. This data is then used as input for an image warp calculation, which results in a non-linear image domain deformation function of the input images to desired virtual view positions. Finally, view synthesis generates output images by appropriate blending of contributions from input images.

IDW uses only sparse disparities combined with saliency and edge information, to compute a smooth image warping function via constrained energy minimization. The resulting views provide high quality and limited artifacts. Our method is capable to support MADs from S3D input, via view extrapolation in a limited range. Decoder complexity issues can be handled by transmission of warping data [13] or sparse disparity, saliency and edge maps to the receiver. Thus our approach is a very interesting alternative to DIBR from multiview plus depth as widely discussed for support of MADs. The core and principle of this algorithm is very similar to the one presented in section 6 [10], with specific optimizations.

**8 StereoBrush: Interactive 2D to 3D conversion using discontinuous warps**

Conversion of 2D to stereoscopic 3D is one of the most important open problems facing the widespread adoption of 3D technology. Large amounts of legacy footage and well established 2D production pipelines ensure that 2D content will remain significant for some time to come. To retarget this content to modern 3D displays, some conversion must be employed. Automatic conversion methods are currently not sufficiently robust for general applications, while high quality

methods are largely manual, and as a result, are extremely expensive, with prices of up to \$100,000 per minute of converted footage.

We introduce a new workflow for stereoscopic 2D to 3D conversion in which the user “paints” depth onto a 2D image via sparse scribbles [14]. In contrast to existing methods in which the conversion pipeline is separated into discrete steps, including rotoscoping, proxy geometry generation, and rendering (with inpainting), our method accomplishes all steps simultaneously, providing instantaneously intuitive 3D feedback to the user. Our method employs several new technologies. It operates directly on the image domain, creating stereoscopic pairs from sparse, possibly erroneous user input while preserving important depth effects. In addition, inpainting is avoided by means of a stereo-aware stretching of background content to fill in holes.

We tightly integrate all steps of stereo content conversion into a single optimization framework, which can then be solved on a GPU at interactive rates. The instant feedback received while painting depth allows even untrained users to quickly create compelling 3D scenes from single-view footage.



**Figure 11: This is a screenshot of our StereoBrush application, showing user scribbles and two real-time visualizations of scene depth**

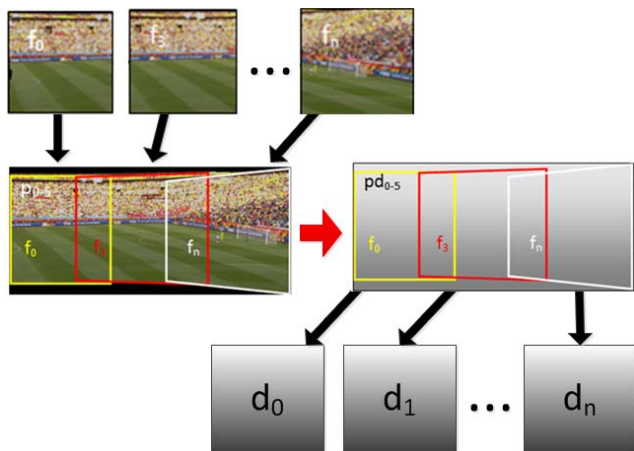
**9 Automatic 2D to 3D conversion for sports**

Given video from a single camera, conversion to two-view stereoscopic 3D is a challenging problem. We present a system to automatically create high quality stereoscopic video

from monoscopic footage of field-based sports by exploiting context-specific priors, such as the ground plane, player size and known background [15]. Our main contribution is a novel technique that constructs per-shot panoramas to ensure temporally consistent stereoscopic depth in video reconstructions. Players are rendered as billboards at correct depths on the ground plane. Our method uses additional sports priors to disambiguate segmentation artifacts and produce synthesized 3D shots that are in most cases, indistinguishable from stereoscopic ground truth footage.

The basic idea of our approach is to separate static and dynamic parts of the scene and process them each using specific algorithms. Our method works for wide field shots and exploits assumptions about the image content. First, each input image is segmented into static background and moving players. So far our system is similar to interactive off-line sports visualizations system used in production [16], [17].

Then, a background panorama is constructed from the whole shot using a classical mosaicing approach, assuming a fixed rotating camera. From this, a depth map is created for the whole panorama using assumptions about the planar structure of the field, and a heuristic, but sufficiently accurate model for the background. Background depth maps for each frame can then be computed by an inverse projection from the panorama depth map using the previous homography as shown in *Figure 12*. By design, these depth maps are temporally stable and consistent throughout the shot. Each segmented player is represented as billboard with depth derived from its location on the background model. Ambiguities in segmentation are corrected so as to not cause noticeable artifacts, giving us a final depth map for each input image. Finally, stereo views are rendered with disocclusions inpainted from known background pixels.



*Figure 12: Overview of our depth map generation,  $f$ : input color frames,  $p$  generated color panorama,  $pd$ : generated depth panorama,  $d$ : per frame output depth maps*

We provide video results in comparison to ground truth stereo

on our web-page<sup>1</sup>, and strongly encourage readers to view our video results on a quality 3D display where possible. While it is possible to notice some differences between footage, such as the graphic overlay, most viewers were not able to distinguish the two videos. Our method is simple, robust, and produces convincing results, but still has some remaining limitations. For one, the segmentation method that we use currently operates mainly on a single frame, and does not use tracking information across sequences. While this step is easily replaceable, it is the main source of artifacts in our results, and a major topic of future work. In addition, our depth assignment assumptions are specific to stadiums with a flat field and rising stadium seats. Other methods for depth map construction would have to be used for different terrain, such as golf courses. One area of future work could be to combine our method with other software or some minimum amount of manual interaction to generate depth maps appropriate to different background structure. However, despite its simplicity, we have found that our method has sufficient accuracy for many cases. This is partially thanks to robustness in human stereo perception, where other cues, such as motion parallax, lighting, and known object size compensate for slight stereoscopic inaccuracies. Our method creates high quality conversion results that are in most cases indistinguishable from ground truth stereo footage, and could provide significant cost reduction in the creation of stereoscopic 3D sports content for home viewing.

## 10 Depth-aware stereo compositing

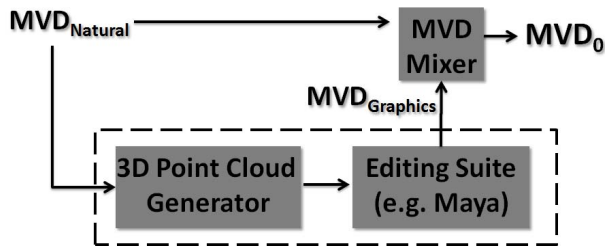
Practically all animated movies are released to theatres in 2D and 3D versions today. In this case artists have all freedom to design 3D scenes that are visually pleasing and exciting without violating the rules of stereography. Even today 2D cinema and TV content contains pixels that are captured by multiple cameras. Images are composed out of content from different sources such as real cameras, graphical elements, animated parts, special effects, etc. This processing can also be automatic or user-assisted, real-time or offline. Algorithms, tools and workflows are established but also continuously extended and improved.

S3D adds a new dimension to this and complicates things further. For instance graphics overlays cannot be simply pasted over other footage. Depth composition must be considered. If a graphics overlay is placed for instance over some scene element but behind it in depth, the results will be annoying, conflicting and unrealistic. To avoid this, graphics or subtitles are placed far out of the screen in many productions today to make sure that all other action stays behind them. Still in live broadcast collisions may happen. In consequence, knowledge about depth composition of all source material is necessary for mixing and composition in 3D.

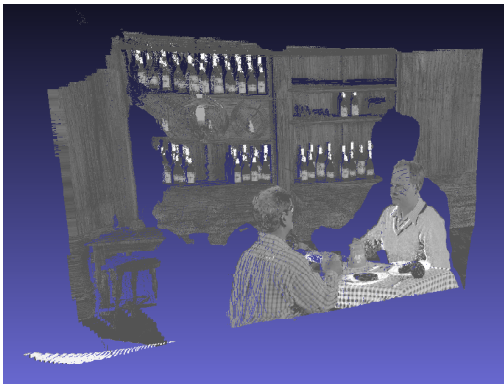
For offline processing in postproduction we developed a set

<sup>1</sup> Supplementary material:  
<http://zurich.disneyresearch.com/videodata/icip/>

of tools and a workflow enabling such depth-aware S3D compositing. This is shown in **Figure 13**. The system operates in general on multiview video plus depth (MVD) as input, i.e. depth data is provided together with the 2 or more input views. From the depth video data, a sequence of 3D point clouds is generated by projection into 3D space, using a priori known camera parameters and geometry. An example is illustrated in **Figure 14**. Such a point cloud is a set of 3D points without connectivity. These data are formatted and stored in a way that they can be read into a 3D editing suite like Maya. Here the animator can create the synthetic content in a way that it does not collide with the live action content and is visually pleasing. Then, equivalent MVD output of the synthetic content is rendered, considering depth ordering and occlusions. Camera parameters and geometry are designed to be the same as those of the live action input. Finally, the MVD mixer can perform the composition of the output MVD data.

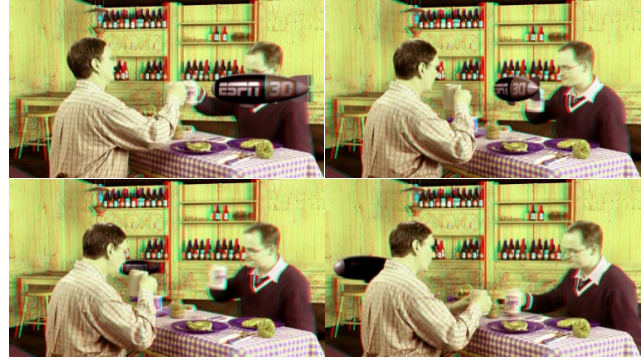


**Figure 13: Compositing workflow.** *MVD data  $MVD_{Natural}$  is given as input, from which an animated point cloud is created. The artist creates synthetic MVD content ( $MVD_{Graphics}$ ) considering the animated point cloud. This synthetic MVD layer is composed with the input to obtain the final composite  $MVD_0$*



**Figure 14: 3D point cloud to be used by the animator**

**Figure 15** shows an example of S3D with depth-aware integration of graphics elements. The zeppelin enters the scene far in the front covering the subjects. Then it dives into the scene, passes behind a subject and exits far back. Depth composition, occlusion and ordering are correctly designed.



**Figure 15: Final composite examples, viewing in red-cyan anaglyph**

## 11 Summary, conclusions and future work

We presented a number of advanced S3D algorithms, tools and systems for high quality content creation, which have a great potential to make future S3D production better, easier, cheaper, and more flexible. All these technologies have in common that they use some kind of disparity awareness of the S3D content. Many of our algorithms use sparse disparities, which can be estimated with very high reliability and accuracy compared to dense depth or disparity maps. Automatic solutions for depth remapping and multiview generation can be created this way using image domain warping.

Still there is room for improvement and optimization in all of these examples, specifically regarding error prone computer vision or image processing algorithms, such as segmentation, feature matching/tracking, dense depth/disparity estimation, etc. Direct depth capture devices such as Kinect gained a lot of attention recently. While the quality of acquired depth data is not yet sufficient for professional view synthesis applications, a lot of momentum can be expected in this area in the future, including more professional hardware and sophisticated algorithms. Development of related products and integration into production systems and workflows will be part of our future work. Development of even more advanced applications such as free viewpoint video or automatic 2D to 3D conversion of general content, with a quality that satisfies highest production requirements, will further be on our future research agenda.

## References

- [1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Mueller, and M. Lang. Three-Dimensional Video Postproduction and Processing, *Invited Paper, Proceedings of the IEEE*, Vol. 99, No. 4, April 2011.
- [2] 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, 2008.



- [3] M. Lambooi, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review, *Journal of Imaging Science and Technology*, Vol. 53, No. 3, 2009.
- [4] J. van Baar, S. Poulakos, W. Jarosz, D. Nowrouzezharai, R. Tamstorf, M. Gross. Perceptually-Based Compensation of Light Pollution in Display Systems, *ACM Symposium on Applied Perception in Graphics and Visualization*, To Appear, August 2011.
- [5] B. Mendiburu. 3D Movie Making – Stereoscopic Digital Cinema from Script to Screen, *Elsevier*, ISBN: 978-0-240-81137-6, 2008.
- [6] S. Heinzle, P. Greisen, D. Gallup, C. Chen, D. Saner, A. Smolic, A. Burg, W. Matusik, and M. Gross. Computational Stereo Camera System with Programmable Control Loop, *ACM Transactions on Graphics (SIGGRAPH 2011)*, August 2011.
- [7] F. Zilly, M. Mueller, P. Eisert, and P. Kauff. The stereoscopic analyzer - an image-based assistance tool for stereo shooting and 3D production, *Proc. ICIP 2010, IEEE International Conference on Image Processing*, Hong Kong, China, Sept. 2010.
- [8] SONY. MPE200 multi image processor and software, <http://pro.sony.com>.
- [9] The Foundry, Nuke, Ocula, <http://www.thefoundry.co.uk>.
- [10] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. Non-linear Disparity Mapping for Stereoscopic 3D, *ACM Transactions on Graphics (SIGGRAPH 2010)*, July 2010.
- [11] A. Smolic, K. Müller, K. Dix, P. Merkle, P. Kauff, and T. Wiegand. Intermediate View Interpolation Based on Multiview Video Plus Depth for Advanced 3D Video Systems, *Proc. ICIP 2008, IEEE International Conference on Image Processing*, San Diego, CA, USA, October 2008.
- [12] M. Farre, O. Wang, M. Lang, N. Stefanoski, A. Hornung, and A. Smolic. Automatic Content Creation for Multiview Autostereoscopic Displays Using Image Domain Warping, *2nd IEEE International Workshop on Hot Topics in 3D (Hot3D 2011)*, Barcelona, Spain, July 2011.
- [13] A. Smolic, Y. Wang, N. Stefanoski, M. Lang, A. Hornung, and M. Gross. Nonlinear Warping and Warp Coding for Content-adaptive Prediction in Advances Video Coding Applications, *Proc. ICIP 2010, IEEE International Conference on Image Processing*, Hong Kong, China, Sept. 2010.
- [14] O. Wang, M. Lang, M. Frei, A. Hornung, A. Smolic, and M. Gross. StereoBrush: Interactive 2D to 3D Conversion Using Discontinuous Warps, *International Symposium on Sketch-Based Interfaces and Modeling (SBIM 2011)*, Vancouver, Canada, August 2011.
- [15] L. Schnyder, O. Wang, and A. Smolic. 2D to 3D Conversion of Sports Content Using Panoramas, *Proc. ICIP 2011, IEEE International Conference on Image Processing*, Brussels, Belgium, Sept. 2011.
- [16] O. Grau, A. Hilton, J. Kilner, G. Miller, T. Sargeant, and J. Starck. O. Grau, A. Hilton, J. Kilner, G. Miller, T. Sargeant, J. Starck, *BBC Research White Paper, WHP 142*, September 2001, <http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP142.pdf>.
- [17] Liberovision Sports Visualization. <http://www.liberovision.com>.