Realtime Preview for Layered Depth Video in 3D-TV

Frick A.[1*], Bartczak B.[2*] and Koch R.[3*]

[*]Christian-Albrechts-University, Hermann-Rodewald-Str. 3, D-24098 Kiel , Germany

ABSTRACT

In this paper we propose a hardware capture system for the generation of layered depth video content, and realtime feedback mechanismus required to ensure optimal data acquisition for 3DTV productions. The capture system consists of five color cameras and two time of flight cameras. The time of flight cameras allow direct depth measurements and thus help to overcome difficulties of classical stereo matching. However they suffer from low resolution and must be combined with stereo to achieve acceptable results in a post production process. Realtime previews are hence necessary, so that a dynamic adaption of the scene as well as the capture parameters becomes possible, and a good data quality can be assured.

Keywords: Preview, Layered Depth Video, Real Time, Filter, Disparity, 3D Display, Time of Flight, PMD, Depth Image Based Rendering

1. INTRODUCTION

The new generation of TV systems, 3DTV, will go far beyond the traditional 2DTV, supplying a viewer with the ability to perceive depth which, significantly increases realism in the scene. Technically, the impression of depth arises through perceiving the scene from slightly different viewpoints through our visual system, providing us with binocular cues, which are interpreted by our brain as depth. Therefore, in order to create 3D impression, one has to provide a viewer with at least two different images of the same scene.

Traditional stereo approaches achieve this by capturing the scene with two cameras from slightly different positions and presenting the captured images separately to the left and right eye of the viewer. The pioneer projects, which investigated the possibilies and limitations of such approaches, were such projects as ACTS MIRAGE and RACE DISTIMA.[1,2]

While traditional stereo approaches may have their advantages, they in general lack flexibility in display choice and are limited to the maximum of two views. In order to overcome this limitations, depth based approaches were proposed. These approaches extend the color information with additional depth information describing the scene geometry. Using depth-image-based rendering methods (DIBR), novel views can then be generated.[3–5] Depth based approaches allow to use different display types (for example stereoscopic so as autostereoscopic) and are in general independent of the display size, and of the number of required views. Further more, knowing the scene geometry the video content may easily be adapted to the requirements of a specific user, for example by adjusting such parameters like optimal display distance or eye difference.

One of the first projects to research the possibilities of the depth based approaches (video plus depth) was the EU project PANORAMA, which focused mainly on stereoscopic video conferences, closely followed by the EU project ATTEST, which further adapted depth based concepts to the needs of 3DTV.[2,6]

In 2008 a new EU project 3D4YOU was started. The research activities of this project cover the complete production chain like acquisition, broadcasting and visualisation of 3D content. One of the projects goals is the investigation of such formats like LDV (layered depth video) and MVD (multi video plus depth), which are extensions to the plain video plus depth representation.[7,8]

In our work we rely on the LDV format, because we believe it to be more suited for 3DTV, having some advantages over the MVD in being more compact and compatible to the standard video. In the LDV format, the 3D content is generated for the view of one camera, the so called reference camera. It consists of multiple layers, each containing depth and texture information, describing geometry and color of different parts of the scene, and

[1] Email: africk@mip.informatik.uni-kiel.de, Phone: +49-431-880 4920

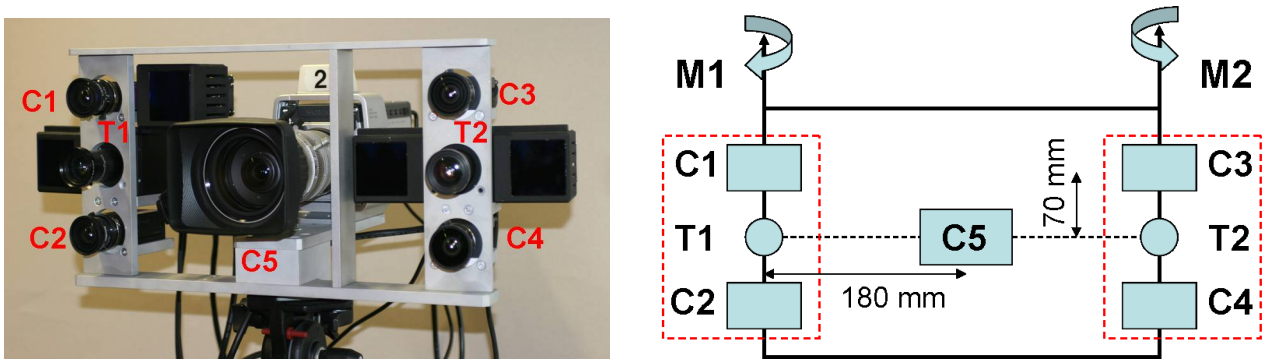
[2] Email: bartczak@mip.informatik.uni-kiel.de, Phone: +49-431-880 4920

[2] Email: rk@mip.informatik.uni-kiel.de, Phone: +49-431-880 4484

is a straight forward extension of the video plus depth format, which can be seen as the first layer of the LDV. The main idea behind the LDV is to describe those parts of the scene which are hidden from the view of the reference camera, through additional layers. In most cases, two layers, one for the foreground and one for the hidden background, the so called occlusion layer, are sufficient.
The generation of LDV is in general a very challenging process, which requires very accurate depth images, that can for example be estimated through stereo matching algorithms. However, despite the progress in this field over the last years, stereo matching still remains a very difficult and ill posed problem, especially in textureless regions. To overcome the difficulties of the stereo matching approaches, we propose using the ToF cameras (time of flight). ToF cameras are active range sensors and overcome the limitations of the stereo approaches by measuring depth directly. The second benefit using the ToF cameras is their ability to capture depth images at 25 frames per second. However, ToF cameras suffer from low resolution and must be combined with the stereo methods in the later processing to achieve optimal viewing results.
In this paper we propose a camera system for capturing the content for LDV generation and introduce a real time preview system, for the content acquisition, consisting of a 3D - Previewer and other mechanisms for online depth evaluation. Such a preview system is of great importance, because one is able to react during capture and can ensure that the optimal data for the later processing can be recorded. Further more, it gives the necessary feedback to ensure that the desired 3D impression can be created from the captured content at all, so, if necessary, additional shooting can be scheduled to improve content. The 3D - Previewer is currently targeted only for the autostereoscopic display of Philips, the WoW - Display, but principally can be adapted to any kind of 3D - Display after an appropriate format conversion.
The rest of this paper is organized as follows. In section 2 we describe the proposed camera setup. In section 3 an overview of the processing chain is given. In section 4 the preview system is introduced and in sections 5 and 6 we demonstrate our results and provide some conclusions.

2. CAMERA SETUP

In this section we describe the proposed hardware camera setup and discuss its benefits for the acquisition of LDV suitable content. We further shortly describe the calibration method, which we used in this work.
The proposed setup consists of five color cameras $C1$,..,$C5$ and two time of flight cameras $T1$ and $T2$. The cameras $C1$,...,$C4$ are Point Grey Grasshopper cameras with a resolution of 1600×1200 pixel. The camera $C5$ is a Sony X300 camera with a resolution of 1920×1080 pixel. The time of flight cameras $T1$ and $T2$ are Pmd-CamCube cameras with a resolution of 204×204 pixel. Figure 1 shows an image of the camera setup together with it's schematic representation.

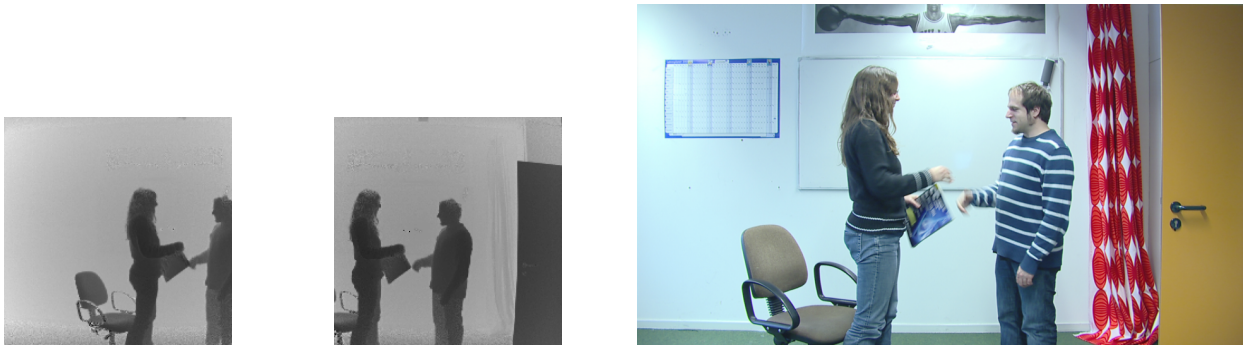


(a) Camera setup (b) Schematic representation

Figure 1. (a) Camera setup: $C1$, ..., $C4$ are Point Grey Grasshopper Cameras, $C5$ is Sony X300, $T1$ and $T2$ are PMD CamCube cameras (ToF) (b) Schematic representation of the camera setup

The ToF cameras are active range sensors which emit modulated infra-red light (IR) at the frequences of 20 MHz ($T1$) and 21 MHz ($T2$) and measure the time per pixel that the light needs to return through special correlation elements. Scene depth can then be computed for each pixel. In addition, a reflectance image measures the

amount of reflected IR amplitude. The effective range of the ToF cameras lies between 1 and 7 meters, so that we are basically limited to close range indoor scenarios. For more information on time of flight cameras we refer to.[9] Figure 2 shows depth images from both ToF cameras together with an image of the central camera $C5$.
The proposed setup is modular. The cameras ($C1$, $T1$, $C2$) and ($C3$, $T2$, $C4$) are organized into rigid modules $M1$ and $M2$. All cameras in a module are approximately parallel aligned, lying on one vertical baseline at a distance of nearly 70 mm to each other. The modules are located to the left ($M2$) and right ($M1$) of the central camera $C5$ at the distance of about 180 mm and can be rotated around their baseline axis. The central camera $C5$ is our reference camera and is aligned to lie approximately on the same horizontal baseline as the cameras $T1$ and $T2$ (see Fig. 1 (b)).



(a) Left PMD camera (b) Right PMD camera (b) Central camera

Figure 2. (a) Depth image from the left PMD camera; (b) depth image from the right PMD camera; (c) image from the central camera

For full coverage of the scene area seen by the reference camera $C5$, with aspect ratio 16:9, we need to align the ToF cameras, with aspect ratio 1:1, properly. Therefore, the modules $M1$ and $M2$ can be rotated as to cover the full viewing area. Compare the coverage of the scene through both ToF cameras (Fig. 2 (a) and (b)) and through the central camera (Fig. 2 (c)).
Positioning the both modules to the left and right of the central camera, with the ToF cameras being on the same horizontal baseline as the central camera, allows us to measure depth and color information in the occluded areas not seen through the central camera, what is essential for the occlusion layer generation of the LDV format. The multiple baselines ($C1$, $T1$, $C2$), ($C3$, $T2$, $C4$), ($C1$,$C3$), ($C2$,$C4$), ($C1$,$C5$,$C4$) and ($C2$,$C5$,$C3$) provide us with different texture orientations which can be used to resolve some ambiguities in stereo matching in the later fusion step between the ToF and stereo data.
For the further data processing, like transformation (warping) of depth images between the views, a reliable internal and external calibration of the complete camera setup is required. Due to bad signal to noise ratio (SNR) of the reflectance image IR, low pixel resolution, and systematic errors in the depth measurement, the ToF cameras are difficult to calibrate with traditional calibration methods. We use therefore the method proposed in[10] to calibrate the ToF cameras together with multiple color cameras in a joint method, which already proved itself as very reliable in our previous work.[11]

3. SYSTEM OVERVIEW

In this section we give an overview of the whole processing chain including image acquisition, storing and previewing. Figure 3 shows an overview diagram of the proposed system.
The diagram consists of active elements in form of arrows, we will call them processes, and passive elements such as image buffers, used for communication and data exchange between the active elements.
As we already mentioned, we use five color cameras $C1$,...,$C5$ and two time of flight (PMD) cameras $T1$ and $T2$. The four cameras $C1$,...,$C4$ are firewire B cameras which deliver images in bayer pattern format in 1600×1200 pixel resolution. Camera $C5$ is a Sony X300 camera, with an HD-SDI interface, and delivers images in YUV format in full HD resolution of 1920×1080 pixel. The time of flight cameras are connected via an USB 2.0 interface and deliver phase images which can be transformed to required depth or amplitude images of 204×204

pixel in the later processing.

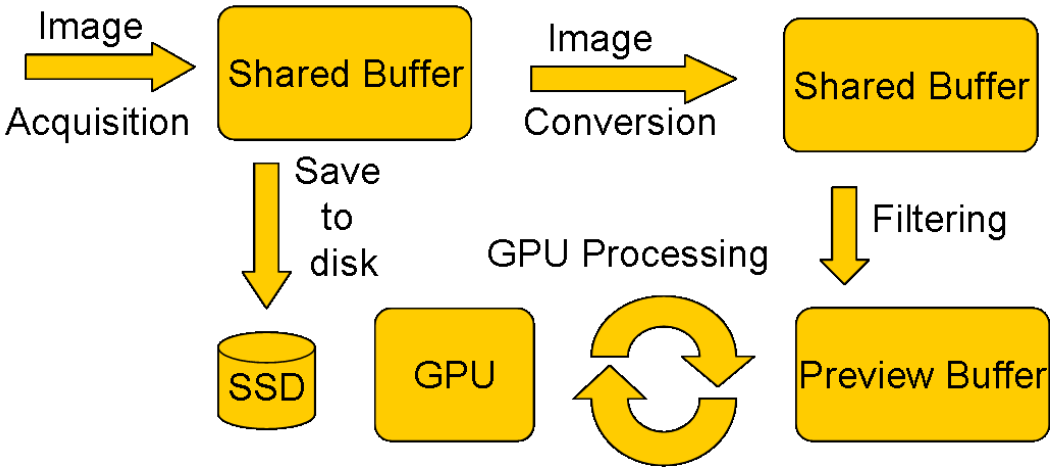


Figure 3. Overview diagramm of proposed system.

In the image acquisition part, each camera is controlled through its own thread, which is responsible for the image transfer from the camera to the shared buffer. All camera threads are currently software triggered to assure synchronized image streams.
A shared buffer holds multiple images from each camera and is a producer - consumer construct which allows synchronized communication between one producer and multiple consumers. We call this strongly coupled communication, because the producer or consumers will block if no data is available or if one of the theads is not ready with processing the data. In the diagramm one can see that the image acquision process is acting as a producer to the first shared buffer and the save-to-disk and image conversion processes are the consumers.
The save-to-disk process, is responsible for storing the captured image data permanently. To ensure the required frame rate, we are using a solid state drive (OCZ Z-Drive p84) of 256 GB capacity which is connected over $\times 8$ PCI-Express interface and achieves writing speed up to 640 MB/s.
The image conversion process is responsible for converting the color images from bayer pattern and YUV format to RGB format and for calculating depth and amplitude images from the ToF phase images. It is a time consuming operation and is divided into several processing threads. The image conversion process acts as the producer for the second shared buffer, which has currently only one consumer, the filtering process.
The filtering process is an optional step which can be switched to simple copying of data to the preview buffer, or to a filtering mode. If the filtering process is switched to the filtering mode, a number of predefined filters is applied to selected image streams before copying images to the preview buffer.
Due to bad signal to noise ratio of the PMD cameras, the filtering step is important for further processing. Currently we are applying a 3×3 or 5×5 median filter for denoising of the PMD depth images, but more sophisticated filtering techniques are possible, for example filtering over time.[11] The median filter also runs in parallel to ensure real time performance.
The preview buffer is used by the GPU - Processing process to transfer the image data to the GPU memory and to run multiple shaders for preview generation. To prevent frame rate drop-down the preview buffer is what we call loosly coupled to his producer and consumers. It means that although the GPU - Processing process may not be ready with processing the current data, the data may be overwritten by the Filtering process. Currently we are using GeForce 9600 GT graphics card for GPU-Processing, but more powerfull graphics cards for better performance can certainly be used.
The proposed system is currently running on Intel(R) Core(TM) i7-860 CPU (SuSe 11.1, 64 Bit). The CPU - Processing part is currently operating with about 20 frames per second, mainly due to the configuration of PMD - cameras, but 25 frames per second are in principal possible. The performance of the GPU - Processing part strongly depends on the number of previewers run simultaneously and can be increased by using a more powerfull graphic card. For the standard configuration we measure over 25 frames pes second.

4. REAL-TIME PREVIEW

In this section we discuss the real - time preview mechanisms supported by our system for online feedback during the shooting. Figure 4 shows an overview diagramm of the GPU - Processing part from Figure 3 in detail.
The images are transfered from the CPU - to the GPU - Memory, where they are stored in form of textures. This is indicated in the diagramm by the arrow Load Textures. The GPU - Memory itself is logically organized into four parts: Original, Warped, Interpolated and Filtered. Each part consists of a number of textures where the results of different shader operations are stored. Images loaded from the CPU - Memory are for example located in the Original block.
Below the GPU - Memory box one can see the shader pipeline box.

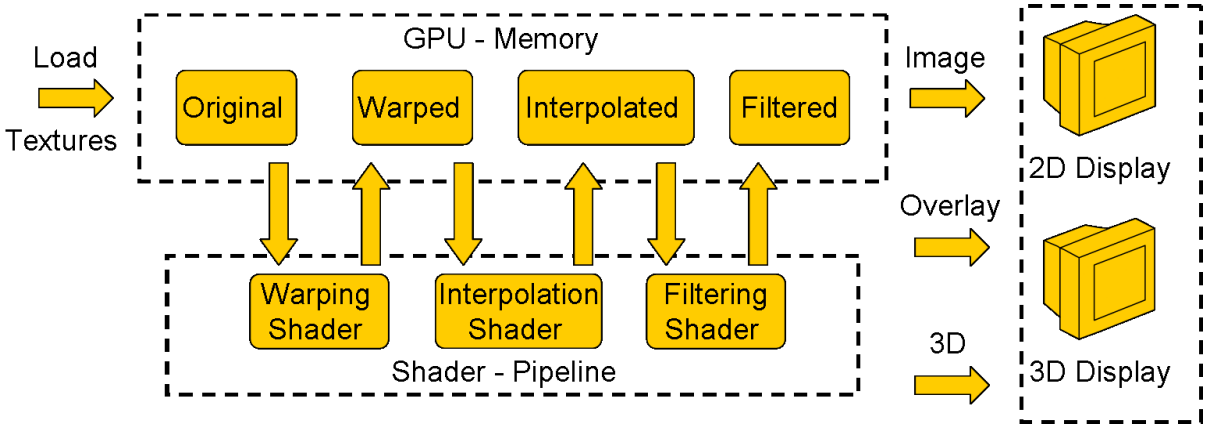


Figure 4. Overview diagramm of GPU - Processing

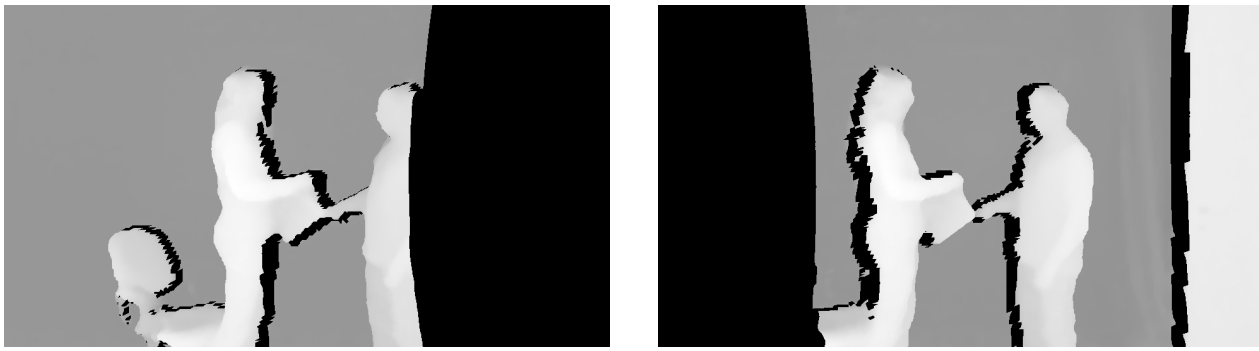
Shaders are programms executed on the GPU. Currently we have three shader programms in our processing pipeline: Warping Shader, Interpolation Shader and Filtering Shader, which are executed from left to right. Each of the shader programms has one GPU - Memory block as input and one GPU - Memory block as output, and consists actually of multiple vertex, fragment or geometry shaders.
On the right side of the diagramm one can see the display side of our preview system, consisting of a 2D - and a 3D - Display. Currently we support three kinds of preview :

1. Image-Preview. One can directly see the depth images from the PMD cameras and the image of the central camera, to control if optimal scene coverage is given.
2. 3D - Previewer for the autostereoscopic WoW - Display from Philips. The LDV format is generated for 3D viewing.
3. Overlay - Previewer. The produced disparity image is transparently overlayed with the image of the central camera by alpha - blending, so one gets a feedback on edge alignment and calibration quality between depth and color.

The rendering to the displays is performed by three different shaders: Image, Overlay and 3D, which are not included in the shader pipeline.
For the Image - Previewer the textures from Original memory are rendered to the 2D - Display, through the Image shader, it includes simple copying of the textures to the display memory.
For the Overlay - Previewer or 3D - Previewer, the depth image from the Filtered memory block is combined with the texture of the central camera from Original block by the Overlay or 3D shader and rendered to the 2D - or 3D - Display.
For the Overlay - Previewer the combination is a simple blending operation: $T = (1 - alpha) * T_{orig} + alpha * D$, whereby T_{orig} is the original texture, D the filtered depth image and $alpha$ a user specified blending parameter.
For the 3D - Previewer the original image and filtered depth image are converted to the LDV format, which can be read by the WoW - Display.
In the following we will shortly describe each of the three processing steps shown in the shader pipeline box.

4.1 Warping

The purpose of warping is to transform the left and right PMD depth images in to the view of the central camera. Changing the viewpoint, however, causes new regions to become visible, the disocclusion regions, where no depth information is available in the target view. When warping both PMD cameras to the central view, such regions appear on the left and right side of the foreground objects, on the right side for the left and on the left side for the right PMD camera. Figure 5 shows an example of the left and right PMD images warped to the central view, where the disocclusion regions are marked in black.



(a) Left warped Pmd depth image (b) Right warped Pmd depth image

Figure 5. Results from warping left and right pmd depth images independently to the view of central camera

In[12] the authors propose a warping method, based on triangle mesh construction, which achieves real - time performance if implemented on the GPU, using vertex and fragment shaders. However, the disocclusion regions are simply filled in by linear interpolation, implicitly given through mesh construction.
While this may be sufficient for a single PMD camera, where no depth information in the disoccluded areas is available, it is certainly not sufficient if using two PMD cameras with different viewpoints, as we have in our camera setup. As one can clearly see from Figure 5, both PMD cameras complete each other in the overlap area so that the disocclusion regions caused by the warping of the left PMD image can be filled through depth information from the right PMD image and vice versa.
Taking into acount this consideratio, we extended the method in[12] to warp multiple depth images simultaneously, and added a simple mesh reduction technique for disocclusion handling.
To warp N depth images, we construct N triangle meshes in the coordinate system of the target camera and render them as a single geometry, whereby occlusions between the meshes are resolved through a standard ZBuffer test. The mesh reduction is implemented in a geometry shader and consists of calculating the normal for each triangle and discarding the triangle if the angle between the normal and the viewing direction of the PMD camera exceeds a predefined threshold. The threshold for the mesh reduction can be changed by the user during runtime. Figure 6 shows some results from warping of left and right PMD depth images to the central view with mesh reduction (a) and in comparison without mesh reduction (b). There are still some artifacts on the object boundaries (Fig. 6 (a)) which are due to mesh - reduction in conjuction with large resolution difference between the central camera and both ToF cameras, but one can cleary see that correct depth has been pasted on the left and right of both persons in the foreground. While the advantage of warping with mesh reduction may seem to be not very big in comparison to warping without, the disocclusion areas become bigger with shorter distances to the camera, so that at a certain distance the difference becomes quite noticeable.
Another issue is that we use the WoW - Display from Philips for our 3D - Preview, which requires the disparity instead of depth images. Therefore we extended the warping to produce disparity images, based on user parameters, like distance to zero parallax plane and maximal allowed disparity. Throughout this paper we do not further differ between the disparity and depth for simplicity matters, because each information can in general be transformed into another and is basically different representation of the same scene geometry.

(a) Warped with mesh reduction

(b) Warped without mesh reduction

Figure 6. Results from warping left and right PMD depth images to the central view simultaneously (a) with and (b) without mesh reduction; Pixels without depth information are coded in black

4.2 Interpolation

After the warping step, disocclusion regions can still be visible in some areas (see Figure 6 (a)). To fill this regions with reasonable depth values, we implemented two simple methods: linear interpolation and background extrapolation, between which the user can freely choose during the runtime. Activation of one of these methods during image acquisition ensures a dense depth image for the central view in later processing steps. Figure 7 (a) shows the depth image from Figure 6 (a) after the background extrapolation.

(a) After background extrapolation

(b) After filtering with asymmetric Box - Filter

Figure 7. Results after (a) Interpolation - and (b) Filtering - step.

4.3 Filtering

After the Interpolation step, depth and texture information is available only for the foreground layer of the LDV format. Without the occlusion layer, however, disocclusion regions can not be filled in correctly. This results in annoying artifacts, when rendering novel views, which can significantly disturb the 3D experience.
In[4,13] the authors propose to reduce such artifacts through filtering of depth image with gaussian filter. We followed this idea but implemented a separable Box - Filter instead of Gaussian - Filter, because it provides results of sufficient quality for our purposes while having less computational complexity compared to the Gaussian - Filter. The results of filtering the depth image after the Interpolation step with an asymmetric Box - Filter from size (20×30) are shown in Figure 7 (b).

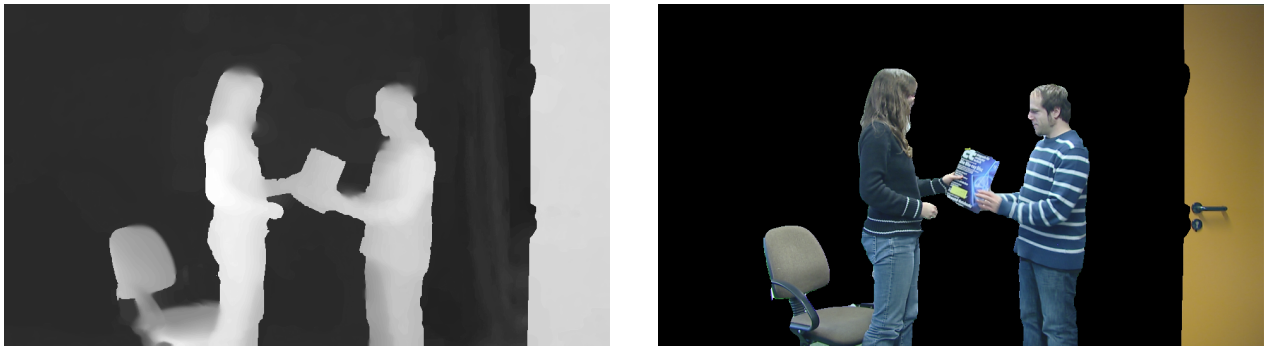
5. RESULTS

In this section we present some experimental results for the proposed system. For the evaluation purposes we captured two videos (Video 1 and Video 2) at 20 frames per second, with preview system running at 25 frames

per second.
During the shooting the system was configured as follows:

1. 5×5 Median - Filter for depth denoising before warping (Filtering part of the CPU - Processing chain, see Figure 3).
2. background extrapolation for disocclusion areas (Interpolation part of the GPU - Processing chain, see Figure 4)
3. asymmetric Box - Filter with size 20×40 (Filtering part of the GPU - Processing chain, see Figure 4)

In order to demonstrate the results, both videos were processed after the capturing but with the same configuration.
Figure 9 (a) and (b) shows the Foreground Layer of LDV format provided by the 3D - Previewer for the Philips WoW - Display. Each frame of the LDV format consists of a color image left and a disparity image right.
Figure 9 (c) and (d) shows the output of the Overlay - Previewer. The Overlay - Previewer blends the color image and disparity image of the central view according to a blending parameter alpha specified by the user during the runtime. The purpose of this previewer is to show the user how good the disparity image (depth image) matches the color image. This provides necessary feedback for the quality of camera calibration and depth measurements by the ToF camera.
To show the quality of the produced disparity images, we additionally segmented the corresponding color image through disparity thresholding. Thereby all pixels in the color image with the disparity smaller than the predefined threshold were marked as black, so that only the foreground objects remained visible. Figure 9 (e) and (f) shows the results of the segmentation. One can clearly see that although there are some errors on object boundaries, the two foreground persons are mainly segmented correctly.
Errors seen in real - time preview can be later corrected in the offline process through more sophisticated filtering methods[11] or fusion with stereo.[14] Figure 8 (a) shows the disparity image of the foreground layer after performing color alignment from[11] and Figure 8 (b) shows segmentation of foreground persons through disparity thresholding on (a). Compare Figure 8 (b) to Figure 9 (e).

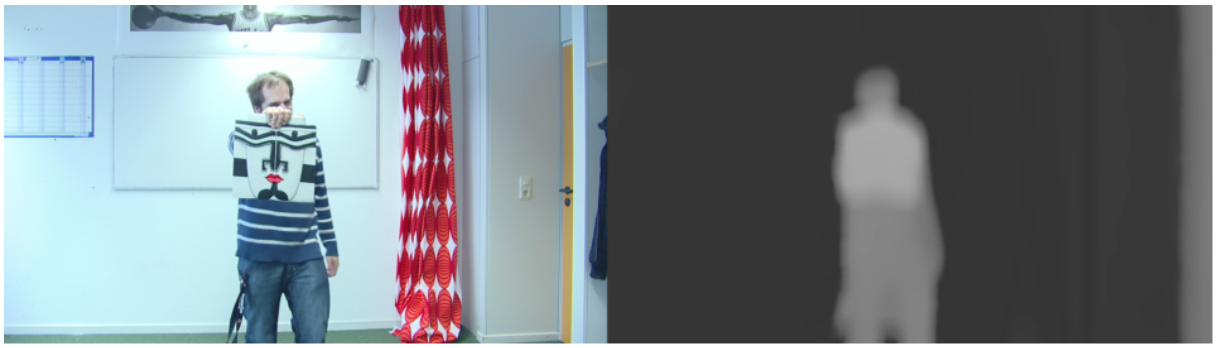


(a) Disparity image after color alignment (b) Segmentation through depth thresholding

Figure 8. Results from not real - time offline processing

6. CONCLUSIONS AND FUTURE WORK

In this paper we introduced a new camera setup for the acquisition of LDV suitable content. We demonstrated that using the proposed setup we are not only able to capture the required content in real - time but also to provide an immediate feedback during the shooting, so that optimal data quality for later processing can be ensured. Currently all cameras are synchronized through software. In our future work we plan to extend our system by a hardware synchronization, using an external trigger for all cameras.

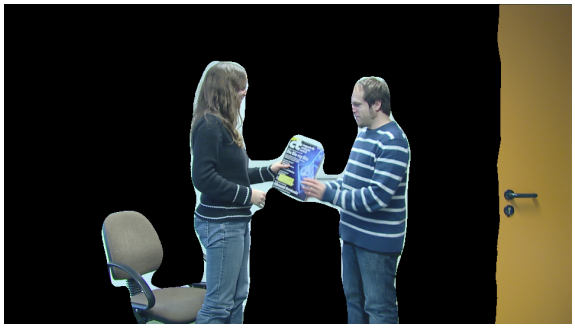


(a) Foreground Layer for 3D - Previewer (Video 1)

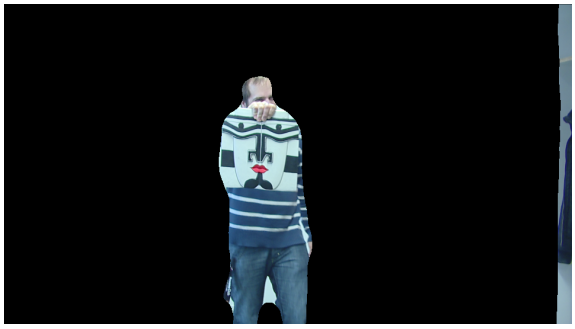
(b) Foreground Layer for 3D -Previewer (Video 2)

(c) Overlay - Previewer (Box - Filter on)

(d) Overlay - Previewer (Box - Filter off)



(e) Depth thresholding (Video 1)



(f) Depth thresholding (Video 2)

Figure 9. Results from the proposed Preview - System: (a) and (b) show the foreground layer for two images from video 1 and 2 (this representation is required by the 3D - Previewer); (c) and (d) show segmentation performed on the texture of the foreground layer through depth thresholding; (e) and (f) show the Overlay - Previewer with and without Box - Filtering activated

Further we are investigating the possibilities of automatic calibration of the central camera based on feature points and fix calibrated side modules ($M1$ and $M2$). This will allow to zoom the central camera in or out of the scene and to adjust the coverage of the scene through the ToF cameras during the shooting.
We are also working on fusion methods between time of flight data and stereo matching approaches for the offline processing.

ACKNOWLEDGMENTS

This work was partially supported by the German Research Foundation (DFG), KO-2044/3-1 and the Project 3D4YOU funded under Seventh Framework Programme, Theme ICT-2007.1.5 Networked Media, Grant 215075.

REFERENCES

[1] Girdwood, C. and Chiwy, P., "MIRAGE: An ACTS Project In Virtual Production And Stereoscopy," in [*Broadcasting Convention International*], 155–160 (September 1996).

[2] Ziegler, M., Falkenhagen, L., ter Horst, R., and Kalivas, D., "Evolution of stereoscopic and three-dimensional video," *Image Communication* **14** (November 1998).

[3] Fehn, C., "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," in [*Stereoscopic Displays and Virtual Reality Systems XI.*], *Proceedings of the SPIE* **5291**, 93–104 (May 2004).

[4] Zhang, L. and Tam, W., "Stereoscopic image generation based on depth images for 3d tv," *Broadcasting, IEEE Transactions on* **51**, 191–199 (June 2005).

[5] Kauff, P., Atzpadin, N., Fehn, C., Müller, M., Schreer, O., Smolic, A., and Tanger, R., "Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability," *Image Communication* **22**(2), 217–234 (2007).

[6] Redert, A., de Beeck, M., Fehn, C., Ijsselsteijn, W., Pollefeys, M., Van Gool, L., Ofek, E., Sexton, I., and Surman, P., "Advanced three-dimensional television system technologies," *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on* , 313 – 319 (2002).

[7] Shade, J., Gortler, S., He, L.-w., and Szeliski, R., "Layered depth images," in [*SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*], 231–242, ACM, New York, NY, USA (1998).

[8] Merkle, P., Smolic, A., Muller, K., and Wiegand, T., "Multi-view video plus depth representation and coding," *Image Processing, 2007. ICIP 2007. IEEE International Conference on* **1**, I: 201–204 (16 2007-oct. 19 2007).

[9] Kolb, A., Barth, E., Koch, R., and Larsen, R., "Time-of-Flight Sensors in Computer Graphics," *Computer Graphics Forum* **28** (2009).

[10] Schiller, I., Beder, C., and Koch, R., "Calibration of a pmd camera using a planar calibration object together with a multi-camera setup," in [*The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*], **Vol. XXXVII. Part B3a**, 297–302 (2008). XXI. ISPRS Congress.

[11] Frick, A., Kellner, F., Bartczak, B., and Koch, R., "Generation of 3d-tv ldv-content with time-of-flight camera," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009* , 1–4 (May 2009).

[12] Bartczak, B., Schiller, I., Beder, C., and Koch, R., "Integration of a time-of-flight camera into a mixed reality system for handling dynamic scenes, moving viewpoints and occlusions in real-time," in [*Proceedings of the 3DPVT Workshop*], (June 2008).

[13] Tam, W. J., Alain, G., Zhang, L., Martin, T., and Renaud, R., "Smoothing depth maps for improved stereoscopic image quality," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], B. Javidi & F. Okano, ed., *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference* **5599**, 162–172 (Oct. 2004).

[14] Bartczak, B. and Koch, R., "Dense depth maps from low resolution time-of-flight depth and high resolution color views," in [*ISVC09*], II: 228–239 (2009).