ABSTRACT

In this paper we introduce an approach for automatic generation of Layered Depth Video (LDV) content for 3D television (3DTV) purposes. The proposed approach relies on a capturing system consisting of two time of flight cameras (ToF) and five color cameras. The LDV content is generated for the central color camera, based on the depth images provided by the ToF cameras and texture information provided by the color cameras. Through different viewpoints of the ToF cameras and color cameras, we are able to reconstruct depth and texture information not only for foreground objects, but also in the occluded areas. The estimated information is then combined to one consistent LDV representation.

Index Terms — Time of Flight, Layered Depth Video, Depth Image Based Rendering, Bilateral Filter, High Definition

1. INTRODUCTION

Future 3D-TV systems will enhance the user experience beyond HD-TV, because they offer to perceive scene depth and allow realistic user immersion into the scene. Technically, the depth impression arises through observing the scene from slightly different viewpoints. The viewpoint difference then causes a distance dependent shift of the scene objects which is interpreted by our brain as depth. In order to achieve 3D experience, one must provide the viewer with at least two different images of the same scene. In standard stereo approaches, the 3DTV content can be generated directly by capturing the scene with two cameras with slightly different view points. The images are then presented separately to the left and right eye of the viewer enabling a 3D impression [1, 2]. While such approaches have their advantages, they in general lack the flexibility in display choice and are limited to only two views. In contrast to standard stereo, depth based approaches extend the color information with additional depth information representing the scene geometry [2, 3]. Using depth-image-based rendering methods (DIBR), novel views can then be generated [4]. Depth based methods overcome the difficulties of the standard stereo by being independent from display geometry and able to generate multiple virtual views, but are in general more challenging in content creation. To limit the complexity or simply through the lack of data, many DIBR approaches try to use only one color image and one depth image representing the foreground scene geometry. This strategy inevitably leads to annoying artifacts in disocclusion regions appearing in a virtual view, for which no texture and depth information exist.

To overcome this limitations different pre- and post processing techniques were proposed, all of which succeed to some degree, but fail when the disocclusion regions become bigger and more complex in both the revealed texture and geometry [5, 6]. Another way to handle the disocclusion regions is to use more sophisticated formats like LDV or MVD, which provide additional depth and texture information for the occluded areas [7, 8]. In this paper, we propose using the LDV format, as it has its advantages in being more compact and compatible to a standard video. The LDV format extends the standard simple one color one depth image representation by additional color and depth images called occlusion layers. In these layers, the background texture and depth information is coded, so that the occluded areas can be properly filled in a virtual view. The consistency in texture and depth between the layers is very crucial for the rendering quality, requiring very reliable and consistent depth images. However, depth estimation from color images is a difficult and error prone process especially on object boundaries and in textureless regions. Using the ToF camera yields the advantage of getting reliable depth information in these regions without complicated estimation.

In our previous work we demonstrated the use of a ToF camera for LDV generation [9]. However, the occlusion layer was generated through a simple extrapolation method, which can cause significant errors. In this paper, we use a modified camera setup with two integrated ToF cameras and demonstrate that we are able to generate a more reliable occlusion layer based largely on real depth measurements in occluded areas.

2. CAMERA SETUP AND CALIBRATION

The camera setup we use for capturing, consists of five color cameras C1,...,C5 and two ToF cameras T1 and T2 (Fig. 1 (a)) and is described thoroughly in [10]. Here, we only give a short description of the system. The middle camera C5 is a Sony X300 camera with a resolution of 1920 × 1080 pixel. The cameras C1,...,C4 are Point Grey Grasshoppers with a resolution of 1600 × 1200 pixel at 180 mm distance from the middle camera. Both ToF cameras are Pmd - CamCube cameras with resolution of 204 × 204 pixel. The ToF cameras are active range sensors that emit modulated infrared light at 20 MHz and 21 MHz frequency and measure the
time per pixel that the light needs to come back by special correlation elements. The measured time combined with light speed yields the depth information for each pixel (Fig. 2). In addition, a reflectance image measures the reflected IR amplitude.

ToF cameras have an aspect ratio of 1:1 and are slightly rotated outwards to achieve the full coverage of the view of the central camera in 16:9 format. Compare the view of central camera (Fig. 1 (b)) with the views of both ToF cameras (Fig. 2).

Due to bad signal to noise ratio (SNR) of the reflectance image, 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 [11] to calibrate the ToF cameras together with multiple CCD cameras in a joint method, which already proved itself as very reliable in our previous work [9].

3. TOF UPSCALING

ToF cameras capture depth at the very low resolution of 204 × 204 pixel. To generate LDV in HD resolution of 1920 × 1080 pixel, ToF data must be transformed (warped) in the view of the central camera C5, requiring to upscale the image by factor 10.

In order to minimize warping errors and for later generation of occlusion layer, we don’t go to the central camera directly but warp ToF data first to the corner cameras C1, ..., C4. So in a first step: we warp the ToF camera T1 to the color cameras C1 (Top) and C2 (Bottom) and the ToF camera T2 to the color cameras C3 (Top) and C4 (Bottom) according to the method described in [12].

Due to the smaller distance 70 mm to the top and bottom cameras in contrast to 180 mm to the central camera, this warping operation causes in general smaller errors compared to the warping to the central camera directly. However artifacts caused by the resolution change and occlusions on the object boundaries are still clearly visible and very disturbing. To correct such artifacts we perform a color alignment between the warped depth map and the corresponding color image, based on bilateral filtering, as proposed in [13] and as we have already shortly described in our work in [9]. Figure 3 (a) shows warped depth image from camera T1 to the view of camera C2 after color alignment.

4. GENERATION OF LDV

After the ToF upscaling step, we have depth images $D_i$ (Fig. 3 (a)) and the corresponding color images $I_i$ in high resolution of $1600 \times 1200$ pixel for each of the color cameras $C_i, i \in \{1, 2, 3, 4\}$.

Using the four pairs $(D_i, I_i)$ we are now able to generate (render) three virtual views, Left, Central and Right, lying on the baseline corresponding to the x-axis of the central camera. Whereby the Central view is identical to the view of the central camera C5 and Left and Right views are shifted by 100 mm on the x-axis to the left and right of the central view. For the right view, we use the pairs $(D_1, I_1), (D_2, I_2)$ and for the left view the pairs $(D_3, I_3), (D_4, I_4)$. For the central view all pairs are used.

The depth map of the Central virtual view is afterwards color aligned to the original image of the central camera and results in the foreground layer depth of the LDV.
In this paper, we introduced a new camera system and demonstrated that we are able to automatically generate a reliable depth information especially in the occluded areas. Further more, we have demonstrated that the calculated occlusion layer is more reliable, than that based on pure extrapolation methods and that we are able to successfully combine the estimated foreground and background layers to one consistent LDV representation.

In our future work we want to improve the depth estimation on object borders. Our special interests lie thereby in the fusion of stereo matching methods and ToF cameras as for example proposed in [14]. We also want to improve the texture consistency between the foreground and background layer especially in conjunction with more sophisticated LDV rendering methods.

6. CONCLUSIONS AND FUTURE WORK
7. REFERENCES


