ABSTRACT

View synthesis offers a great flexibility in generating free viewpoint television (FTV) and 3D video (3DV). However, the depth-image-based view synthesis approach is very sensitive to errors in the camera parameters or poorly estimated depth maps (also called depth images). Because of these errors, three kinds of artifacts (blurring, contour, hole) are possibly introduced during the general synthesis process. Comparing to conventional methods which implement the view synthesis only in ideal case, in this paper, we propose to design an error compensation and reliability based view synthesis system where the potential errors are considered. The main contributions are highlighted as follows: Firstly, the camera parameter errors are compensated by a global homography transformation matrix. Secondly, the depth maps are classified into both reliable and unreliable regions and the reliability based weighting masks are built to blend synthesized images from two different views together. Finally, a reliability depth map based hole-filling technique is used to fill the existing holes. The experimental results demonstrate that these artifacts are efficiently reduced in the synthesized images.

Index Terms— view synthesis, artifacts reduction, reliability based, camera parameter error, depth map error

1. INTRODUCTION

3D video (3DV) has caught enormous attention in consumer market due to the progress in technical maturity of the 3D displays, digital video broadcastings and computer vision algorithms. Based on the view synthesis technology, arbitrary views of 3D scene can be synthesized, which is very useful in the free viewpoint television (FTV), 3D video compression, and other multiview applications. One of the great challenges in the view synthesis process is to generate virtual views with minimal artifacts. In the depth-image-based rendering (DIBR) approach [1], three kinds of artifacts are often introduced: a) the blurring artifacts due to the errors in camera parameters; b) the contour artifacts owning to both the depth map errors along the boundaries and the digital camera properties; c) the holes due to disocclusions or mapping process. Zhang and Tam [2] used a low-pass filter to smooth the noise in depth map and also the sharp edges to do view synthesis. Zitnick [3] used a layered representation in high-quality video view interpolation. Hasinoff et al. [4] proposed a method called boundary matting which represents each occlusion boundary as a 3D curve to reduce matting artifacts. Oh et al. [5] proposed to fill holes using depth based inpainting method. Previous methods are able to deal with some of the artifacts. However, the visual quality or computation complexity can not be satisfied at the same time and none of the previous works considered the artifacts caused by camera parameter errors.

In order to efficiently and robustly synthesize images in virtual views, we design an error compensation and reliability based view synthesis system. The proposed system considers both the camera parameter errors and depth map errors and deals with them separately. For the camera parameter errors induced artifacts, we compensate them by a global homography matrix. In addition, we deal with the depth map induced errors and holes based on the depth map reliability.

The rest of the paper is organized as follows: Section 2 presents a system overview of our proposed view synthesis algorithm. In section 3, we explain camera parameter error compensation in details. In section 4, we describe depth reliability based blending and hole-filling. The experimental results are reported in section 5. Finally, section 6 concludes the paper.

2. A FRAMEWORK OF PROPOSED VIEW SYNTHESIS SYSTEM

In this section, we present an overview of our proposed view synthesis system. The synthesis approach can be treated as a mapping process followed by some post-processing steps, which aim to reduce the artifacts introduced during the mapping process. The mapping process between two images are carried out by first projecting the points of one image plane to the 3D world and then projecting the 3D points to the other image plane.

Before the view synthesis procedure, the captured color images or videos associated with corresponding depth maps and all camera calibration data are given. There are two kinds of basic mapping: forward-mapping and backward-mapping.
Forward-mapping directly maps the reference data onto the virtual image plane. However, since many pixels in the reference image are likely to map to the same pixel location or non-integer pixel locations, an additional filter computation and interpolation are needed at the virtual image side. These extra processes increase the complexity and may reduce the subjective visual quality of final synthesized image. On the other hand, backward-mapping refers to the virtual image data to the reference data, therefore, all the pixels in the virtual image plane are guaranteed to find their correspondences. However, the virtual depth map has to be synthesized before backward-mapping.

Taking the advantages of both forward-mapping and backward-mapping, we design our view synthesis system as shown in Fig. 1.

The depth maps are firstly synthesized by forward-mapping and then median filtered to eliminate small blank points caused by forward-mapping. Two synthesized color images are derived by backward-mapping from the left and right reference images. At last, the reliable pixels from two synthesized images are blended to generate final virtual view in the post-processing operation. The post-processing operation includes camera parameter error compensation followed by a reliability based blending and hole-filling step.

3. CAMERA PARAMETER ERROR COMPENSATION

In this section, we will prove that the camera parameter errors are able to be compensated by a homography transformation. When a point $\hat{M}$ in the world coordinates is projected to camera coordinates, a pixel $\hat{m}$ in the image can be found by

$$w\hat{m} = P\hat{M}$$  \hspace{1cm} (1)

where a single point $\hat{M} = [X, Y, Z, 1]^T$ in world coordinates and a projected point $\hat{m} = [x, y, 1]^T$ in camera coordinates represent their homogenous pixel positions. $w$ is a non-zero scalar and $P$ is a $3 \times 4$ camera matrix.

We further denote a $4 \times 4$ matrix $\hat{P}$ as an extension of $P$ and $\tilde{m}$ as an extension of $w\hat{m}$, where

$$\hat{P} = \begin{bmatrix} P & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tilde{m} = \begin{bmatrix} w\hat{m} \\ 1 \end{bmatrix}$$  \hspace{1cm} (2)

In this notation, equation (1) can be rewritten as

$$\tilde{m} = \hat{P}\tilde{M}$$  \hspace{1cm} (3)

In the DIBR process, we first get the point in the real 3D world.

$$\hat{M} = \hat{P}^{-1}\tilde{m}$$  \hspace{1cm} (4)

$$w = \frac{Z - \hat{P}^{-1}(3, 1 : 3)}{\hat{P}^{-1}(3, 1 : 3) \times \tilde{m}}$$  \hspace{1cm} (5)

where $Z$ is the depth value of the pixel $\hat{m}$. $\hat{P}^{-1}(3, 1 : 3)$ is a row vector containing the first three elements of the third row in $\hat{P}^{-1}$. A pixel-to-pixel correspondence is then built. Let $\tilde{m} = [x', y', 1]^T$ denote the pixel position in the other image with $3 \times 4$ projection matrix $P'$. We have $\tilde{m} \sim P'\hat{P}^{-1}\tilde{m}$  \hspace{1cm} (6)

where $\sim$ means the equation is up to a scaling factor. Considering insignificant error in the projection matrix $P$, the perturbation of scalar $w$ is negligible compared to the global projective transformation. And there exists a homography matrix $H$ such that

$$P'\hat{P}^{-1} = H(P'\hat{P}^{-1} + P_e)$$  \hspace{1cm} (7)

where $P_e$ is the accumulating error matrix in 3D mapping process. According to the above discussion, we can see that the camera parameter error is compensated by a global homography matrix $H$. In practical situation, because we use two reference images to synthesize a virtual image, this homography matrix helps to match the two synthesized images (synthesis from left and right views) well for further blending. $H$ is determined between the two synthesized images by a 4-point algorithm. Each pair of corresponding pixels determines two equations. Then $H$ is fixed by finding at least 4 pairs of reliable correspondences since there are 8 unknowns in $H$.

4. DEPTH RELIABILITY BASED BLENDING AND HOLE-FILLING

Depth map is generally a quite smooth image with some strong edges compared to the traditional color images. View synthesis is not very sensitive to noise in smooth and texture regions of depth map because insignificant artifacts would appear in the synthesized image. However, the errors along the object boundaries or depth discontinuities in depth map will cause significant contour artifacts in the synthesized image. One example of contour artifact is circled in Fig. 2. Moreover, even if the depth map is perfect, there are still...
contour artifacts due to the capture device in the synthesized image. This phenomenon is illustrated in Fig. 3. Our captured color image contains the foreground $F$, background $B$ and a transition region $T$ mixed by foreground and background. After synthesis, part of the transition region will be mapped with foreground and the others be mapped with background. The disocclusion region denoted as $D$ in the left synthesized image means that this region is not visible in the left reference view. Because two reference images are used for synthesis, this disoccluded region is easily compensated by the other view. However, the transition region $T$ (circled in Fig. 3) will cause undesirable contour artifacts after blending from the right synthesized image.

We study the images taken by different cameras and find the transition region between the foreground and background is usually around $t = 6$ pixels in width. Therefore, in order to avoid these contour artifacts and keep the naturalness along the boundaries at the same time in the synthesized image, we propose to classify the forward-mapped depth map into both reliable and unreliable regions by dilating the disocclusion regions around by $t$ pixels in width as shown in Fig. 4 (a)(b).

The reliability based masks $M_l$ and $M_r$ are built for the synthesized images from two reference views as shown in equation (8) and (9), where $\alpha$ equals to the distance ratio.

\[
M_l(u,v) = \begin{cases} 
\alpha, & d_l(u,v) \& d_r(u,v) \text{ are reliable;} \\
1, & \text{only } d_l(u,v) \text{ is reliable;} \\
0, & d_l(u,v) \text{ is unreliable.}
\end{cases}
\]

\[
M_r(u,v) = \begin{cases} 
1 - \alpha, & d_l(u,v) \& d_r(u,v) \text{ are reliable;} \\
1, & \text{only } d_r(u,v) \text{ is reliable;} \\
0, & d_r(u,v) \text{ is unreliable.}
\end{cases}
\]

Then, the virtual view image is rendered by blending two neighboring views using their mask sheets.

\[
I = I_lM_l + I_rM_r
\]

Although most of the disocclusions are covered after blending, there are still some holes left because of either real occlusion or mapping process. We fill these holes using the real background information in the synthesized color image. The real background information means that the pixel is not a mixture of background and foreground but only background. The real background is easily identified using the proposed reliability based masks. This approach is efficient, especially for those holes lying along the boundaries.

5. EXPERIMENTAL RESULTS

Our research mainly targets to improve the subjective visual quality of the synthesized image. As a result, we compared the subjective visual quality of three resultant images: an image without camera parameter error compensation, an image without reliability based blending and hole-filling, and an improved image using our proposed method. As an example, we show the results of a typical sequence named 'breakdancers' generated and distributed by Interactive Visual Group at Microsoft Research [6]. The captured images have a resolution of $1024 \times 768$ pixels. The camera arrangement in "breakdancer" is shown in Fig. 5. These cameras are neither parallel nor exactly in the same horizontal line.

For the convenience of comparison, we generate center view with $\alpha = 0.5$ from captured view by camera 3 and camera 5. The two rows in Fig. 6 are different parts of the image in the view of camera 4. Fig. 6 (a) and (a’) are real captured
images in the view of camera 4. Fig. 6 (b) and (b’) without homography compensation seem to be blurred because of the wrongly registered views. We get the Fig. 6 (c) and (c’) by directly blending the synthesized images without reliability based blending and hole-filling after the process of homography compensation. Clearly, the contour artifacts are observed along the boundaries. Fig. 6 (d) and (d’) demonstrate the final synthesis results using our proposed approach. The differences between the final results using our proposed approach and the real captured images are negligible.

6. CONCLUSION AND DISCUSSION

In this paper, we introduce a novel view synthesis framework by taking the advantages of both forward-mapping and backward-mapping. In the post-processing part of the framework, we propose a homography-based camera parameter error compensation and a depth map reliability-based blending and hole-filling approach. These contribute to an efficient and robust artifacts reduction of synthesized view.

We successfully reduced the main noticeable artifacts, however, some other artifacts such as those caused by transparent or reflectance surfaces need to be carefully dealt with. These problems will be addressed in our future work.

7. ACKNOWLEDGEMENT

This work has been supported in part by the Applied Science and Technology Research Institute (ASTR52-20A00210/11PN) of Hong Kong, China.

8. REFERENCES


