A Novel Filling Disocclusion Method Based on Background Extraction in Depth-Image-Based-Rendering

Zhenguo Lu, Yuesheng Zhu, Jian Chen
ShenZhen Graduate School, Peking University, China
Email:zglu@sz.pku.edu.cn, zhuys@pkusz.edu.cn

Abstract: DIBR is a key technique in 2D-to-3D conversion systems. But in the traditional DIBR algorithm there is an inherent issue which leads to disocclusion areas like holes in the warped images. The holes can be reduced by using depth map preprocessing, however, resulting in some geometric distortions and a low depth perceived quality. In this paper, a novel method for disocclusion filling without depth map preprocessing is proposed. A new background extraction method is presented for accumulating background information for hole-filling. Cooperated with simple linear interpolation and the extrapolation, the disocclusion areas can be satisfactorily removed. The experiment results show that a good stereo perception can be obtained without any geometric distortion by using the proposed method.

Key-Words: DIBR; Background Extraction; 3D warping; View Synthesis

1 Introduction

With the rapid development of display and the digital signal processing technologies, 3D-TV has been expected to be the next generation advancement of television. The traditional approach produces 3D contents with the transmission of left-eye and right-eye view images. However, it has two main drawbacks in regard to large transmission bandwidth and depth reproduction flexibility. To overcome the above shortcomings [1], a new “view plus depth” representation format is proposed in the 2002 European Information Society Technologies (ITS) project “Advanced Three-Dimensional Television System Technologies” (ATTEST).

Based on this new format, multiple virtual views can be regenerated by DIBR (depth image based rendering) [2]. However, DIBR has an inherent vice that many disocclusion areas like holes will appear after the image warping because of the lack of information in the original image. As a result, the qualities of the synthesized virtual views are degraded. In general, these holes can be dismissed from two aspects. One is preprocessing the depth map with a smooth filter, such as symmetric Gaussian filter [2], asymmetric Gaussian filter [3] and edge dependent depth filter [4], and the other is filling the remaining holes by linear interpolation and extrapolation [5]. Although preprocessing the depth map reduces the number of small holes and lowers the system complexity [6], the original depth information is modified, and therefore produces some geometric distortions [7] and depth-resolution attenuation [8]. It is noted that these algorithms only consider the current image information for hole-filling and ignore the other video sequences information. In this situation, our approach tries to draw into the temporal video information for disocclusion filling.

In this paper, a new hole-filling method by extracting video background without depth prepressing is proposed. Allowing for the
complexity of extracting background information for different scenes, our method only considers a specific scenario of moving foreground and stationary background. In general, a scene can be divided into foreground and background. If the foreground moves, the new exposed background information can be extracted, and the accumulated background information can be used to fill the holes after image warping. Combined with simple linear interpolation and extrapolation, the disocclusion areas can be fairly eliminated.

The remainder of this paper is organized as follows. The proposed method is illustrated in details in section 2, and then the experimental results are shown in section 3. Finally, the conclusions are given in section 4.

2 Proposed method

Warping images with the original depth maps directly will result in a lot of holes in the final synthesized images. Two main reasons lead to the appearing holes. One is the discrete depth intensity, which usually generates small holes, and the other is the abrupt depth intensity changing at objects edges which produces many large holes. In nature, the large holes appear as a result of lacking of the background information, which often appear around the foreground edges, as black region shown in Fig.1. Therefore, these large holes can be filled with pixels from the background rather than the foreground [9]. The large holes in the current warped images may be filled with the corresponding pixels of the neighbor video sequences.

![Fig. 1. Stereo images of the warped image without smooth depth maps and hole-filling](image1)

(a) (b) (c) (d) (e)

In the following, we will introduce our method in details. As shown in Fig.2, our method consists of three parts: (1) extracting the background information from the video (2) 3D image warping (3) holes filling. In the proposed method, background is firstly extracted from the video sequences, and then followed by image warping and holes filling of each frame. The details will be given as follows.

![Fig. 2. Flowchart of the proposed method.](image2)

2.1 Extracting background

In the section 1, it has been assumed that the background are extracted only in the scene of moving object and stationary background. The
foreground region usually has a large depth value while the background depth value is often very small, so the background can be easily extracted by depth map.

As shown in Fig.3, the extracting background method can be divided into three steps. The first step is to dilate the depth map. As the depth map is not very accurate at the object edge, so the foreground edge may be mistaken as the background region. After the dilation of the depth map, some background information may lose, but the lost information of current frame can be compensated from the next frames.

Fig. 3. Flowchart of the extracting background.

Then the current frame background can be extracted as follows:

\[
I_{bg}(x, y) = \begin{cases} 
I_{image}(x, y) & I_{depth}(x, y) \leq th \\
0 & \text{otherwise}
\end{cases}
\]

(1)

Where \(I_{bg}(x,y)\) indicate the background of the current frame, \(I_{image}(x,y)\) and \(I_{depth}(x,y)\) specify the current frame and its corresponding depth map respectively.

The threshold value is defined as

\[
th = d_{\min} + \frac{d_{\max} - d_{\min}}{K}
\]

(2)

Where \(d_{\max}\) and \(d_{\min}\) are the max and min values of the current depth map. The factor \(K\) is manual positive value.

Finally, the last step result is used to update background information. In order to update the background quickly, an update background mask is generated in the first frame. If the intensity of first depth map frame is larger than \(th\), the intensity of the corresponding coordinates in the mask will be set zero, on the contrary, set one.

In the next frames, we will update the background by the formula (3).

\[
I_{N}(x, y) = \begin{cases} 
\alpha \cdot I_{N-1}(x, y) + (1-\alpha) \cdot I_{n}(x, y) & I_{n}(x, y) \leq th \\
I_{n}(x, y) & I_{n}(x, y) > 0
\end{cases}
\]

(3)

Where \(I_{N}(x,y)\) and \(I_{N-1}(x,y)\) indicate the current and the last update background respectively. \(I_{n}(x,y)\) indicates the current extracting background. \(I_{n}(x,y)\) indicates the update background mask. The factor \(\alpha\) is manual positive value less than 1.

For the slowly changing scene, the method extracts background from some key frames rather than each frame. The experimental result shows an efficient and excellent background image. In addition, the depth map of the background is extracted in the same way.

2.2 3D Image Warping

As the compression artifacts exist in the origin depth map, human factor is taken into account [10] and redefines a new formula of depth distance which can be expressed as follows:

\[
\frac{1}{Z_v} = \frac{1}{Z_{\text{far}}} \left( \frac{v}{255} \right) + \frac{1}{Z_{\text{near}}} \left( 1 - \frac{v}{255} \right) \quad , 0 \leq v \leq 255
\]

(4)

where \(Z_v\) is the \(v\)-th depth distance between an object and a viewer. \(Z_{\text{far}}\) and \(Z_{\text{near}}\) are the farthest and nearest depth distances between an object and a viewer. Although the step may spend little time, it can save much time to fill holes.

Two different stereoscopic camera system configurations, the “toed-in” and the “shift-sensor”, have been introduced in the literature [11]. To balance the simplicity and the viewing experience, the “shift-sensor” configuration is applied in this paper. Then the virtual left-eye and right-eye images can be generated as follows.
In the formula (5), \( X_L \) and \( X_R \) are the pixel coordinates of the generated left-eye and right-eye images. \( X_C \) is the pixel coordinate of the origin image. \( t_x \) is the baseline distance between two virtual cameras. \( f \) is the focal length. \( Z_v \) is the converted depth distance between the object and the two virtual cameras.

### 2.3 Hole-filling

In this section, the extracted background image will play an important role. Before filling the holes, the extracted background is warped to left-eye and right-eye background images which will save much time to use the background information to fill the holes in the next steps.

The hole-filling specific steps are illustrated in Fig.4. \( I_v \) is the generated virtual left-eye or right-eye image. For simplicity, the virtual left-eye image is taken as an example. \( I_m \) is a pixel of \( I_v \). If \( I_m \) isn’t a hole, the method will calculate the next pixel directly. Otherwise, it will be determined as a small or a large hole. A small hole can be filled by simple linear interpolation. For a large hole, it can be deal with two steps. Suppose \( I_b(x,y) \) is a pixel in the warped left-eye background image which corresponds to \( I_v \) in the virtual left-eye image. If \( I_b(x,y) \) is not zero, then \( I_i(x,y) \) is set to be \( I_b(x,y) \). Otherwise, an extrapolation method is used to fill the large holes. Considering of the efficiency and viewing experience, the multidirectional extrapolation hole-filling method [12] is used to fill the big hole.

In this paper, only the horizontal directional pixels are considered and the large holes can be filled by formulas (6) and (7).

\[
I_v(x, y) = \sum_{i=1}^{m} w_i \cdot I_b(x_m + \alpha \cdot i, y) \tag{6}
\]

\[
\alpha = \{ \frac{1}{-1}, \frac{m-1}{-m}, \frac{1}{+m}, \frac{m-1}{+m} \} \sum_{i=1}^{m} w_i = 1 \quad w_i > 0, w_j \geq w_{j+1} \tag{7}
\]

Where \( I_b(x_m + \alpha \cdot i, y) \), \( i = 1, 2, 3, 4 \) of the corresponding warped background image are the nearest four pixels to \( I_i(x,y) \).

### 3 Experiment result

The video sequence “interview” provided by the MPEG 3DAV group [13] is tested in our experiment. To evaluate the performance of the proposed method, the symmetric and asymmetric [3] depth smoothing method are implemented to compare with the proposed method.

The parameter selections of the depth smoothing methods are according to the result of [3]. In the formula (2), \( K \) is set to be 10. The factor \( \alpha \) of the formula (3) is 0.25. In the formula (6), the values of \( w_i \) are 0.4, 0.3, 0.2 and 0.1. Z-buffer algorithm is applied in the three methods above to handle the visibility.

The background extracted from the “interview” sequences is shown in Fig.5. By comparing Fig.5 (a) with (b), it’s obvious that some useful background information at object edge is recovered. As some parts of the foreground are motionless or always overlap the background, the full background cannot be completely recovered.
Fig. 5. The extracting background from the “interview” sequences. (a) the extracting background of the first sequences, (b) the final extracting background information.

The enlarged segment of the virtual left-eye image produced by symmetric depth smoothing method is shown in Fig. 6 (a) and (d). It’s easy to find the geometric distortions in the white circle regions.

Fig. 6 (c) and (f) shows the results created by the proposed method. Without depth map preprocessing, the proposed method handles the geometric distortions very well.

Fig. 7 shows the final stereo frame of the test sequences. With a red-green glass, we can perceive a good stereo effect.

Fig. 7. The final stereo image of the proposed method.

4 Conclusion

This paper presents a novel hole-filling algorithm in 2D-to-3D conversion systems. A new background extraction method is proposed to accumulate the background information for hole-filling. Combined with the simple linear interpolation and the extrapolation, the disocclusion can be satisfactorily filled. Considering the complexity of extracting background information for different scenes, the proposed method only focuses on a specific scenario of moving foreground and the stationary background. No additional depth map preprocessing is adopted, the proposed method can achieve a fairly satisfactory stereo effect without any geometric distortions.

References:


