A Novel Hole filling method based on Projection onto Convex Set in DIBR

Weiquan Wang[1] and Yingyun Yang[2] and Qian Liang[3]

Abstract. Depth image based rending (DIBR) is a common technology for synthesis virtual viewpoint in 3DTV. The main problem is how to effectively reduce and fill the holes in the virtual viewpoint. In this paper, a novel hole filling method based on projection onto convex set is proposed to solve this problem. To achieve the hole-filling, the proposed method projects the pixels of the observed frame onto the holes position of reference frame by forming the convex set. The experimental results show that our method can improve the quality of virtual view images and solve the hole-filling problem efficiently.

Keywords: DIBR; Projection onto convex set; Affine Transformation; PSF estimation.

1 Introduction

DIBR [1] is currently used for rendering new views by using one color image and its correlative depth-image. The most significant problem in DIBR is how to deal with newly exposed areas appearing in the virtual images. The main reason is that the occluded region of reference viewpoint would have been visible in the new viewpoint.

To eliminate the tiny holes in virtual viewpoint, an asymmetrical 2D Gaussian filter is proposed by Zhang [3] to smooth the depth image. Chen [4] proposed an edge-dependent Gaussian filter that only smoothed the depth regions with sharp depth transition. Lee [6] ameliorated the filtering method and adopted an adaptive edge-oriented smoothing method. While these methods decrease the number of holes, changes of whole depth values bring the problem of object geometrical dis-

[1] Weiquan Wang (✉)
College of Information Engineering, Communication University of China, Beijing, China
e-mail: hnldwwq@126.com

[2] Yingyun Yang
College of Information Engineering, Communication University of China, Beijing, China

[3]Qian Liang
College of Information Engineering, Communication University of China, Beijing, China

tortion in the warped image. Liu [7] filled the holes by pre-processing depth images and image merging. However, it needs too much desired images and the calculations are too complicated.

In this paper, a novel hole filling method based on projection onto convex set [10] in DIBR is proposed. The proposed method can fill the holes effectively and reduce the distortion of the generated images by forming the convex set and projecting the pixels from the observed frame into the corresponding holes position of the reference frame.

The remaining sections of this paper are organized as follows: Depth-image based rending is briefly introduced in section 2. The proposed algorithm would be presented in section 3. In section 4, we show the experimental results. Conclusion and future work is given in section 5.

2 DEPTH IMAGE BASED RENDERING

DIBR can render target view image by warping one color image and its associated depth-image. DIBR method has three processes: smoothing depth-image, 3D warping and hole filling. Pre-processing of depth-Image and 3D image warping is briefly introduced in section 2 and hole-filling method is depicted in section 3.

2.1 Pre-processing of Depth-Image

Depth image is a gray scale image .The gray values represent the depth information of the point in the world. The sharp changes in depth image cause holes after image warping, so the pre-process of depth image is necessary. In this paper, we perform morphological operations on depth images with 3×3 mask.

2.2 3D image warping

Combined with the depth information and camera parameters, we can map a pixel in reference image into 3D space firstly, and remap it onto target view image by 3D image warping [1]. Assuming the system with two cameras, the perspective projection equation is given as follows:

$$Z^{'} m^{'} = Z P^{'} R P^{-1} m + P^{'} T \qquad (1)$$

Where m and $m^{'}$ are the coordinates of pixels in the reference image and the target image. Z and $Z^{'}$ indicate the depth value in the first camera and the second camera. The 3×3 matrix P and $P^{'}$ specify the intrinsic parameters of the first camera and the second camera. R and T are the rotation matrix of 3×3 and the translation matrix of 3×1 of the second camera.

In 3D warping, several original image points are warped to the same location in the new image plane, which leads to overlapping. A so-called Z-buffer algorithm [5] can be used to solve this problem.

3 THE PROPOSED ALGORITHMS

3.1 Projection onto Convex Sets

Assuming continuous observation frames is $\{y_t(m,n)\}_{t\in[1,p]}$, (m,n) represents the coordinate, t means time and p is the number of the frames. We can define the following convex sets:

$$C_{m,n,i} = \left\{f(k,l): \left|r_i^f(m,n)\right| \le \sigma_0\right\} \ \ 0 \le m \le M, 0 \le n \le N, i = 1\dots p \quad (2)$$

$$r_i^f(m,n) = y_i(m,n) - \sum_{k=0}^{M-1}\sum_{l=0}^{N-1} f(k,l)h_i(m,n,k,l) \quad (3)$$

$f(k,l)$ is the image to be reconstructed. δ_0 represents the credibility of the observation model , which is determined by the point spread function. By using the mapping operator P, the arbitrary of $x(k,l)$ can be mapped to convex sets $C_{m,n,i}$. The definition is described as follows:

$$P_{m,n,i} = \begin{cases} x(k,l) + \frac{r_i^x(m,n)-\delta_0}{\Sigma_0 \Sigma_p\, h_i^2(m,n,k,l)}\, h_i(m,n,k,l) & if\ r_i^x(m,n) > \sigma_0 \\ x(k,l) & if\ -\sigma_0 < r_i^x(m,n) < \sigma_0 \quad (4) \\ x(k,l) + \frac{r_i^x(m,n)+\delta_0}{\Sigma_0 \Sigma_p\, h_i^2(m,n,k,l)}\, h_i(m,n,k,l) & if\ r_i^x(m,n) < -\sigma_0 \end{cases}$$

$$C_A = \{f: \alpha \le f(k,l) \le \beta\}\ \ \alpha = 0, \beta = 255 \quad (5)$$

3.2 The core steps

The algorithm involves three core steps: construction of reference frame and observation frame, pixel projection and point spread function estimation.

3.2.1 Construction of reference frame and observation frame

Aimed at reducing needle-holes, but not altering the whole image, we perform morphological processing on the depth image firstly and process the warped images by median filter. However, there are a large number of holes in the boundary of image as shown in Fig1 (a).

After projecting pixels to the holes positions, it needs to calculate the estimated value in the range of its PSF, but the pixels around holes is unknown, so the reference frame should be pre-filled.

(a)　　　　　(b)　　　　　(c)　　　　　(d)

Fig. 1 Construction of reference frame and observation frame (a) Warped image after median filter (b) reference frame (c) observation frame (d) Blocking Effects

This paper selects the reference frame by using areas interpolation to pre-fill holes. Target view image synthesized by Zhang [3] or Lee [6] algorithm is selected as observation frame as shown in Fig1 (b) and Fig1 (c).

Blocking effects will appear in the edge of holes as shown in Fig1 (d). By using the POCS method and correcting pixel, blocking effects can be eliminated.

3.2.2 Pixel projection

POCS method is intended to recover the detailed information of the image. In this paper, affine transformation [9] is proposed to project the observation frame' pixels onto the hole positions of reference frame accurately.

Image registration based on affine transformation is using the similarity between observed frame and reference frame to obtain affine transformation model with the following formula:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_0 \\ a_3 \end{bmatrix} \tag{6}$$

In the formula, $X(x, y)$ represents the coordinates of pixel in observed frame. $U(u, v)$ represent the incremental coordinates after the affine transformation. a represents a vector which includes six parameters of the affine tion. $\begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix}$ is a matrix of rotation, expansion and contraction. $\begin{bmatrix} a_0 \\ a_3 \end{bmatrix}$ is a translational matrix.

3.2.3 Point spread function estimation

After projecting pixels to the holes positions in reference frame, it needs to calculate the estimated value of each point in the range of its PSF and compare it to the real value of observation frame. If the error is beyond the permitted range, we should amend pixel value of reference frame to make sure that the error between estimated value and real value is reduced to the allowable range.

The mathematical expression of point spread function is as follows:

$$\sigma(x,y) = e^{-\frac{(x-X_0)^2+(y-Y_0)^2}{2}}, X_0 - 2 \leq x \leq X_0 + 2; Y_0 - 2 \leq y \leq Y_0 + 2 \quad (7)$$

Assume that the point spread function is linear time-invariant and $\delta'(x,y)$ represents the normalized function of $\sigma(x,y)$:

$$\sigma'(x,y) = \frac{\sigma(x,y)}{\sum_{m=X_0-2}^{X_0+2}\sum_{n=Y_0-2}^{Y_0+2}\sigma(m,n)} \quad (8)$$

The estimated value of projected pixel is as following:

$$f(X_0, Y_0) = \sum_{x=X_0-2}^{X_0+2}\sum_{y=Y_0-2}^{Y_0+2} f_{ref}(x,y)\sigma'(x,y) \quad (9)$$

Where $f_{ref}(x,y)$ represents the pixel in the reference frame. Therefore, with formulation (9), pixel error between the actual value of observed frame and estimated value can be calculated by the following equation:

$$r = y(m,n) - f(X_0, Y_0) \quad (10)$$

If the absolute value of r is larger than the specified threshold value e, we should amend the pixels of reference frame within the PSF window by the following equation:

$$f_{ref}(x,y) = \begin{cases} f_{ref}(x,y) + (r+e)\sigma'(x,y) & r < -e \\ f_{ref}(x,y) & else \\ f_{ref}(x,y) + (r-e)\sigma'(x,y) & r > e \end{cases} \quad (11)$$
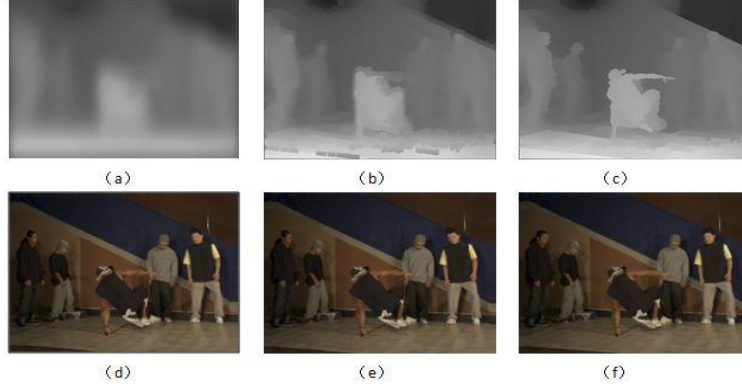
3.3 Implementation

As mentioned above, our proposed holes filling method is described as follows:

1. Construct reference frame by performing morphological processing on the depth image and median filter. Use the image which synthesized by Zhang or Lee algorithm as the observation frame.
2. Convert the reference frame and the observation frame to grayscale image. Calculate and get the affine transformation model
3. Define the convex set by using the affine transformation in step 2.
4. Project the pixel onto the corresponding holes positions of reference frame. Correct the projected pixels of reference frame by using formulation (11).

4 Experimental Results

In order to prove our proposed algorithm, we use the video test sequences Break-dancers of Microsoft research institute. The sequence has 8 views with 20 cm ho-

rizontal spacing between each, and $Z_{near} = 44.0$，$Z_{far} = 120.0$. The video images have a resolution of 1024×768. The depth images are point based. In our research, the fifth camera is considered as target view camera and the fourth camera is considered as reference camera.

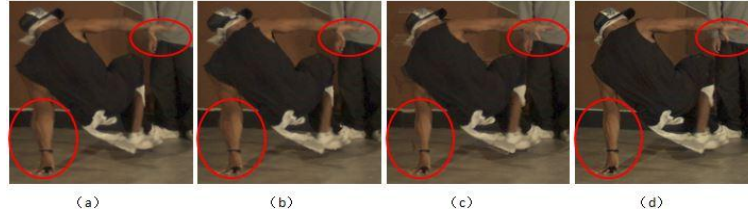(a)　　　　　　(b)　　　　　　(c)

(d)　　　　　　(e)　　　　　　(f)

Fig. 2 Comparison of different algorithms (a) Depth image is processed by Zhang method (b) Depth image is processed by Lee method (c) Depth image is processed by proposed method (d) Synthetic viewpoint by Zhang algorithms (e) Synthetic viewpoint by Lee algorithms (f) Synthetic viewpoint by proposed algorithms

Fig. 3 Segments of images (a) Zhang algorithms (b) Lee algorithms (c) Proposed algorithms (d) Real image

To evaluate performance, we compared the proposed algorithm with Zhang et al. and Lee et al. For Zhang's algorithm, the standard deviation of the Gaussian core was set to 20 horizontally and 60 vertically. The filter window size was set to 91 points horizontally and 161 points vertically. For Lee's algorithm, image edge was detected by using sobel operator. Hole regions with vertical edges is pre-processed by asymmetric smoothing filter of Zhang. For other regions of depth image, symmetric Gaussian filter is adopted.

Fig2 represents the warped images of three algorithms. Fig2 (a) represents the whole depth image is smoothed by an asymmetrical Gaussian filter. Fig2 (b)
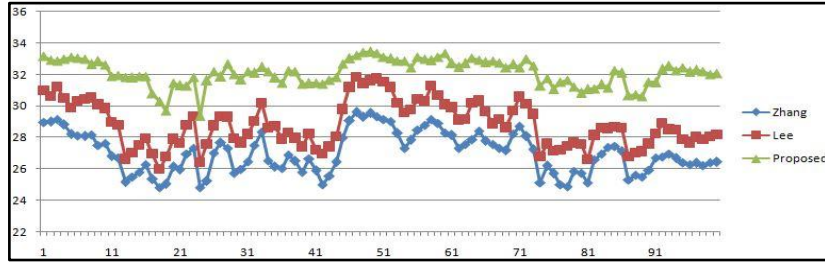

(a)　　　　　　(b)　　　　　　(c)　　　　　　(d)

represents that only vertical edges of the depth image is pre-processed by asymmetric smoothing filter. Overall, three methods can achieve satisfactory subjective results.

Fig3 shows the segments of images. For Zhang or Lee's algorithm, the left arm becomes thick and the right hand fingers have a certain offset as shown in Fig3(a) and Fig3(b). The proposed algorithm does not change the depth image significantly, so the image can be warped with almost no distortion as shown in Fig3(c). However, there are some 'ghost' artifacts around the edge of image and some pixels of right hand fingers are projected to the wrong positions. Though not perfect, the overall image quality is greatly improved.

In this paper, we simulate three algorithms by using Matlab, and the average absolute peak signal to noise ratio (PSNR) is regarded as a measure of objective indicators of test results as is showed in Fig4 and Table 1. We refer to the PSNR of Newly double-side algorithm [10] and algorithm of Xue [11].

The experimental results show that our proposed algorithm is better than all of the above in PSNR.

Fig. 4 PSNR comparison of different algorithms (Breakdancers)

Table 1 PSNR comparisons of different algorithms

Sequences Algorithms	Breakdancers	Ballet
Zhang algorithm	27.0104	23.1532
Lee algorithm	28.8358	25.2848
Newly double-side [10]	30.8569	——
Xue algorithm [11]	31.6136	——
Proposed algorithm	32.1241	28.4878

5 CONCLUSIONS

In this paper, we first perform morphological operations on depth images to reduce needle-holes and preserve the depth information of non-holes regions. Finally, a POCS-based method is proposed for hole-filling. Experiment results show that our proposed method can improve the quality of virtual view images to a large extent.

Because of the hole pixel information is unknown when constructing reference frame, the affine matrix is not entirely accurate. Projection of pixels has a certain deviation. As to the future work, we are focused on improving the resolution for target image.

References

1. MCMILLAN L.An image-based approach to three-dimensional computer graphics [D]. NC，USA：University of North Carolina at Chapel Hill，1997
2. C. Fehn, "Depth-Image-Based Rendering (DIBR),Compression and Transmission for a New Approach on 3D-TV,"Proceedings of SPIE Stereoscopic Displays and Virtual RealitySystems XI, pp. 93–104, San Jose, CA, USA, Jan. 2004.
3. ZHANG Liang，TAM W J. Stereoscopic image generation based on depth images for 3DTV[J].IEEE Trans. Broadcasting，2005，51（2）：191-199
4. CHEN W Y，CHANG Y L，LIN S F，et al. Efficient depth imagebased rendering with edge dependent depth filter and interpolation[C]//Proc. ICME 2005. [S.l.]：IEEE Press，2005：1314-1317.
5. GREENE N, KASS M, MILLER G. Hierarchical Z-buffer visibility[C]. Proceedings of the 20th annual conference on Computer graphics and interactive techniques., 1993:231–238.
6. Lee P, et al. Adaptive edge-oriented depth image smoothing approach for depthimage based rendering[C]. Broadband Multimedia Systems and Broadcasting(BMSB), 2010 IEEE International Symposium on. , 2010:1–5.
7. ZHAN-Wei L, PING A, Su-xing L, et al. Arbitrary view generation based onDIBR[C]. Intelligent Signal Processing and Communication Systems, 2007. IS-PACS 2007. International Symposium on., 2007:168–171.
8. FAN Y, Chi T. The novel non-hole-filling approach of depth image based render-ing[C]. 3DTV Conference: The True Vision-Capture, Transmission and Displayof 3D Video, 2008. , 2008:325–328.
9. De-Xian Z, ZHAO-JiG, Chao-Yang Z. Image Registration Based on Simple Affine Transformation Model [J]. Journal o f the Academy of Equipment Command & Technology, 2005, 16 (1): 84-87.
10. Gao L, Chen H, Liu C, et al. A newly virtual view generation method based on depth image[C]//Signal Processing (ICSP), 2012 IEEE 11th International Conference on. IEEE, 2012, 2: 1088-1091.
11. Xue J, Song G, Li H. Virtual view generation method without pre-processing depth-image for 3DTV[C]//Multimedia and Signal Processing (CMSP), 2011 International Conference on. IEEE, 2011, 1: 362-365.