Median Filter Based Adaptive Compensation Method for Depth Map Pre-Processing

Asha Jayachandran, Preetha V.H

Abstract— ***Depth Image Based Rendering (DIBR) is 2D to 3D conversion technology using color image and its corresponding depth image that is widely employed in applications like 3D TV, free view television etc. 3-D viewing is the next most happening technology. Since transmission of 3D video demands a lot of bandwidth, a new technology that renders virtual views using a color image and its corresponding depth image was proposed. If the depth map is incomplete, the virtual views generated will contain holes or disocclusions which affect the quality of 3D viewing. Since holes occur when the intensity in depth map changes significantly, smoothening methods were proposed reduce the number of holes. Since smoothening methods affect the edges and destroys the original information in the depth map, Adaptive Compensation method (ADC) which processes the image in different modes was proposed. Improved Adaptive Compensation method does not produce satisfactory results for images with large number of holes. Though an improvement in PSNR and SSIM improvement is observed, the number of holes in the warped image is increased. A median filtering is incorporated in Adaptive Compensation method to reduce the number of holes. The experimental results indicate an improvement in PSNR and SSIM as well as a reduction in number of holes.***

Index Terms- Depth Image Based Rendering, 3D TV, Inpainting, Adaptive Compensation Method, Disocclusion, Median Filter, Holes, Virtual views.

I. INTRODUCTION

3D viewing gives a very realistic viewing experience and is a trending technology. Transmission of 3D video requires a huge amount of bandwidth. However, a 3-D view is synthesized using more than two images, which costs more processing resources for compression and transmission. Advanced Three dimensional Television System Technologies (ATTEST) [1] of European information society technology have proposed a conversion method that uses a 2-D image and a corresponding depth map, which is called Depth Image-Based Rendering (DIBR), to improve the saving and transmission defects that are inherent with traditional 3-D TV [2]. Depth Image is a grey scale image in which each pixel carries the information about the depth of the scene. Currently there exist three methods to retrieve the depth map: Stereo metric Camera, Image Structure analysis and Depth Camera [3]. Stereo metric Camera captures the same scene from two different angles. But it requires large transmission bandwidth and the depth effects are not easy to customize. Image Structure analysis generates the depth map by analyzing characteristics of the image like texture, relative positions etc. to evaluate the depth. This technology has two main problems: (1) Switching view angles may generate holes in the 3D images (2) Relative movement between scenes and objects may increase the computational cost and slow down the speed [3]. The third method is to capture the depth of the scene using depth sensors and generate virtual views at the receiving end. This method called Depth Image Based Rendering (DIBR) generates right and left virtual views by horizontally shifting pixels in the color image according to the depth value from the depth image. The performance of DIBR relies heavily on the completeness of the Depth map [3]. Holes may be generated in the 3D views synthesized via DIBR if the quality of the depth map is poor [3]. One disadvantage of the DIBR approach is that with this type of data representation, one or more virtual images of the 3D scene have to be generated at the receiver side in real time [4]. The most significant problem in DIBR is how to deal with newly exposed areas (holes) appearing in the virtual images [4]. Occlusion occurs due to regions which are visible in the original view but invisible in the new view. Disocclusions are caused by regions that are invisible in the original view and become visible in the virtual view. Since there is no information corresponding to these regions in the original image, they appear as white or holes in the virtual view. Holes can be reduced by either pre-processing the depth map or post -processing the virtual views. Post processing methods produces undesirable blurring, zigzag effects and other texture artifacts. Hence pre-processing methods are preferred over post-processing methods. Since holes occur when the depth changes sharply, smoothening methods were proposed earlier. Tam et al. [5] proposed a symmetric smoothening method for pre-processing the depth map. But this method affected the vertical edges in the depth map and produced geometric distortions [1]. Zhang et al. [4] proposed an asymmetric smoothening filter which preserves the vertical edges but it affected the intensity of non-vertical lines and increased the computational complexity [2]. To address these issues an adaptive edge oriented smoothening method was proposed by Lee and Effendi [6]. This method applies symmetric filtering to areas having vertical edges and asymmetric filter to regions having non-vertical edges. Although this method preserves the vertical edges the original information of the depth map is lost because of the various size and direction of holes that are created by different sizes and direction of rapid changes in the depth map, which results in a deterioration of 3-D view [2]. Chih-hsien in [2] proposed an improved Adaptive Compensation method which preserves both vertical edges and original information of the depth map.

Manuscript published on 30 June 2016.
* Correspondence Author (s)
Asha Jayachandran*, PG Student, Department of Electronics and Communication Engineering, SCT College of Engineering, Pappanamcode, (Kerala). India.
Preetha V.H, Assistant Professor, Department of Electronics and Communication Engineering, SCT College of Engineering, Pappanamcode, (Kerala). India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

The positions of the holes are determined using a fast parallax look up table. The image is then divided into 16×16 blocks and classified into three modes according to the number of holes. Mode 1 blocks which covers region having large number of holes are inpainted in accordance with analysis of texture direction. Adaptive smoothening method is applied to Mode 2 blocks having medium number of holes. Mode 3 blocks which contains no holes are left as such. Adaptive Compensation method produces low quality rendered images if large holes are present in the depth map. The proposed method uses a median filter to process Mode2 regions. This produces less blurring of the background compared to Gaussian filtering. This results in depth map having high PSNR and SSIM values compared to ADC. The number of holes after warping also reduces considerably compared to that of ADC method.

II. DEPTH IMAGE BASED RENDERING

The block diagram of a traditional Depth Image Based Rendering system is as in Fig.1, where a color image and its associated depth map are input to the system [2]. The color image and pre-processed depth image are inputs to image warping. Warping performs a mapping from the color image to the right and left virtual image using the depth information from the depth map.

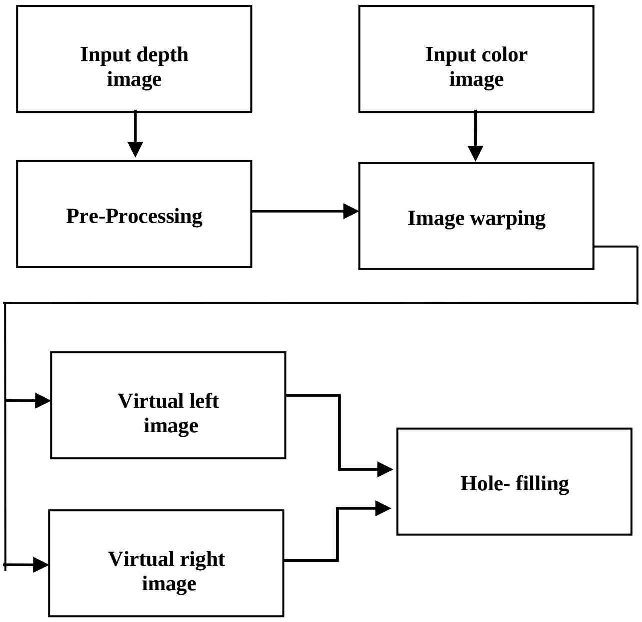


Figure 1: Flowchart of the DIBR system

A. The Pre-Processing of the Depth map

Pre-processing of depth intends to reduce the number of holes generated in the virtual left and right views generated after image warping. Since holes occur when the intensities in the depth map changes abruptly, smoothening methods were proposed initially. The equations for symmetric 2-D filter proposed by Tam et al. [5] and asymmetric 2-D Gaussian filter proposed by Zhang et al. [4] are as below

$$g(x,\sigma_\mu) = \frac{1}{\sqrt{2\pi}\sigma_\mu}\exp\{\frac{-x^2}{\sigma_\mu^2}\}, \quad for\ \frac{-w}{2} \le x \le \frac{w}{2} \qquad (1)$$

$$g(x,\sigma_v) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp\{\frac{-x^2}{\sigma_v^2}\}, \quad for\ \frac{-w}{2} \le y \le \frac{w}{2} \qquad (2)$$

$$s'(x,y) = \frac{\sum_{v=\frac{-w}{2}}^{\frac{w}{2}}\{\sum_{\mu=\frac{-w}{2}}^{\frac{w}{2}} s(x-\mu, y-v)g(\mu,\sigma_\mu)g(v,\sigma_v)\}}{\sum_{v=\frac{-w}{2}}^{\frac{w}{2}}\{\sum_{\mu=\frac{-w}{2}}^{\frac{w}{2}} g(\mu,\sigma_\mu)g(v,\sigma_v)\}} \qquad (3)$$

where g(x,σ_μ) denotes the horizontal Gaussian function, and g(y,σ_v) denotes the vertical Gaussian function, along σ_μ is the standard deviation of the horizontal direction, σ_v is the standard deviation of the vertical direction, w is the filter window size, x and y stands for the horizontal and the vertical axis position respectively, while s(x,y) represents the depth value in the depth image, and s'(x,y) denotes the depth value after smoothing. For symmetric filter σ_μ, σ_v and window sizes in vertical and horizontal directions are the same, but for asymmetric filter σ_μ and σ_v has different values and window sizes also differ. Since symmetric filtering destroys vertical lines and introduces geometric distortions, asymmetric filtering was proposed. Asymmetric filtering, even though maintained vertical edges, affected the non-vertical lines of the depth map. Adaptive edge oriented filtering applied asymmetric filtering to regions with vertical edges and symmetric filtering to other regions. Adaptive compensation method classifies the depth map into 3 modes according to the number of holes. Mode 1 regions with maximum number of holes were inpainted in accordance with the texture direction. Adaptive filtering was applied to medium sized holes in mode2 regions and Mode 3 regions are maintained as such to preserve the original information in the depth map.

B. Image Warping

Image warping is a fundamental concept in DIBR, using the original depth map to compute the parallax of each pixel and then the left and right shift distance are calculated [2]. Lastly, using a shift sensor, two virtual left and right view images are generated, as [2]

$$C_l = C_c + \frac{b_x}{2}\frac{f}{Z}$$
$$C_r = C_c - \frac{b_x}{2}\frac{f}{Z} \qquad (4)$$

where C_l and C_r represents the virtual left and right view images pixel coordinates respectively, and C_c is the original color image coordinate, where b_x denotes the baseline distance, and f represents focal length, Z is the depth distance from object to camera. The baseline distance is assumed as 36 pixels and the focal length is assumed as one for performance estimation.

C. Hole Filling

Even though pre-processing of depth map reduces the number, holes will be present in the virtual left and right views synthesized via image warping. The holes can be filled by linear interpolation, linear extrapolation or neighbor mirroring or inpainting.

III. PROPOSED WORK

The proposed method is a modification of the Adaptive Compensation method proposed by Chih [2]. The Adaptive Compensation method produces images of low quality when the size of the holes is large in size. The flow chart of the proposed Adaptive Compensation method is as in Fig.2 where a median filter is applied to Mode2 region, which preserves both the structure are texture information of the depth map.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

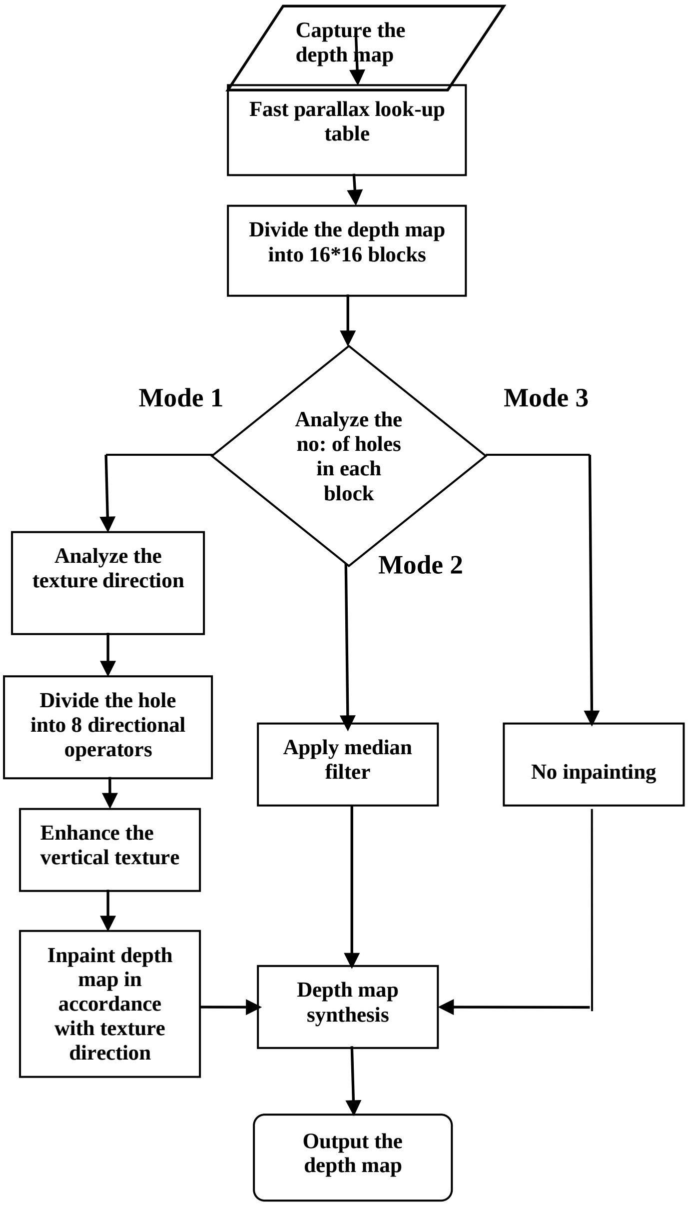


Figure 2: Flowchart of the proposed DIBR system

The locations of holes are determined using a fast parallax lookup table. The parallax is determined using the relation in eqn 5.

$$Parallax = W \times \{1 - \frac{d}{255}\} \quad (5)$$

Where W denotes the greatest parallax, which is generally 5% of the image width, d denotes the depth value in the depth map and Parallax denotes the computed parallax value.
When the difference in the parallax values of d(x,y) and its neighbor is greater than Th_0, d(x,y) is determined as a hole and labeled by $Label_d(x,y) = 1$.

$$Label_d(x,y) = \begin{cases} 1, if\ d(x+1,y) - d(x,y)\ Th_0\ \ for\ left\ view \\ \quad if\ d(x-1,y) - d(x,y)\ Th_0\ \ for\ right\ view \\ 0, otherwise \end{cases} \quad (6)$$

where the threshold value Th_0 is set to 10 and d(x, y) denotes the value of depth at (x, y). The depth map is further divided into 16×16 blocks for further analysis and classified into three modes according to the number of holes.

$$Mode = \begin{cases} 1, if\ total\ hole\ number\ of\ the\ block\ > Th_1 \\ 2, if\ 0\ < total\ hole\ number\ of\ the\ block\ < Th_1 \\ 3, if\ total\ hole\ number\ of\ the\ block\ = 0 \end{cases} \quad (7)$$

the value of Th_1 is taken as 10. Mode 1 regions contain maximum number of holes and hence must be processed first. Mode 1 regions which contain maximum number of holes are inpainted in accordance with the texture direction. For areas with smaller holes, which are determined as mode 2 by the proposed method, a median filter is used. For Mode 3 regions which have no holes, no processing is done.

A. Inpainting of Mode 1 region

Mode 1 regions contain more number of holes and hence must be inpainted. The direction of texture is determined using a Sobel mask. 0, 90,135 and 45 degree Sobel masks are convolved with the depth map to determine the relative degree of effect of greatest texture. The texture direction is determined as per (8)

$$J(x,y) = \begin{cases} 0, & if\ Max(S_0, S_{45}, S_{90}, S_{135}) = S_0 \\ & if\ Label_d(x,y) = 1 \\ 45, & if\ Max(S_0, S_{45}, S_{90}, S_{135}) = S_{45} \\ & if\ Label_d(x,y) = 1 \\ 90, & if\ Max(S_0, S_{45}, S_{90}, S_{135}) = S_{90} \\ & if\ Label_d(x,y) = 1 \\ 135, & if\ Max(S_0, S_{45}, S_{90}, S_{135}) = S_{135} \\ & if\ Label_d(x,y) = 1 \end{cases} \quad (8)$$

where S_0, S_{45}, S_{90} and S_{135} represents the convolution result of the depth map blocks with Sobel masks. Since the human eyes is more sensitive to change in vertical lines [6] and image deformation occurs after image warping, J(x,y) requires extra labeling of the vertical line parts after J(x,y) is found [2]. When $Label_d(x,y)$ is one and J(x,y) is 90, next five pixels are searched in the downward direction. If they are all holes and any of the five pixel is labeled 90, then all pixels from J(x,y) to J(x,y + 5) are labeled by 90. After strengthening the vertical line, the holes with greater size in the depth map is inpainted in accordance with the analysis of the texture directions [2].

B. Mode 2 processing

Mode 2 regions contain moderate number of holes. Mode 2 regions are median filtered. This helps in avoiding the blur caused by the use of Gaussian filter and preserves the details of the image. The median filter preserves the edges of the depth map which otherwise is destroyed by Gaussian filtering.

C. Mode 3 processing

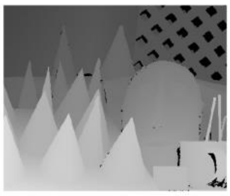
In order to avoid the destruction of the original depth image and to reduce the amount of geometric distortion, when a region is determined as mode 3, no inpainting method is applied by the system [2].

IV. RESULTS

The input images are obtained from the Middlebury Stereo Vision database [8]. The images used are of resolution 696 × 556. The input images used is as in Fig 3. The simulations are done using MATLAB R2014b using MacBook Air with 1.4 GHz Intel Core i5 processor.

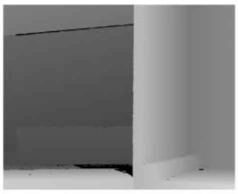
Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

(a)



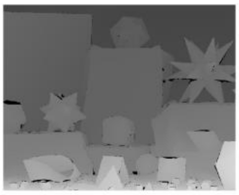
(b)

(c)



(d)

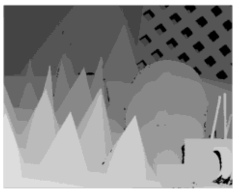
(e)



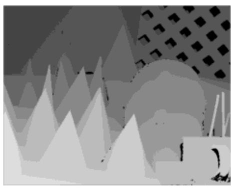
(f)

Figure 3: Test Images (a) Cones Color Image (b) Cones Depth Image (c) Woods Color Image (d) Woods Depth Image e) Gift Color Image (f) Gift Depth Image

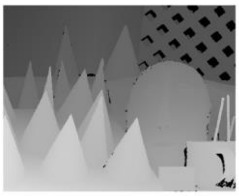
The positions of the holes are located using a fast parallax look-up table using (5). The value of W is assumed to as 5% of the width of the image. The differences in parallax between pixels are compared with a threshold Th_0 which is assumed as 10. If the difference in parallax exceeds the threshold, the pixel is labeled a hole. The depth map is divided into 16×16 blocks and classified according to the number of holes. Mode 1 holes are inpainted according to texture direction. A search is performed in the texture direction until a pixel that is not labeled as hole is found. The depth value is then copied from the current pixel to the hole. Mode 2 holes are median filtered and Mode 3 holes are maintained as such. The Depth map of Cones pre-processed as per Symmetric filtering, Asymmetric filtering, Adaptive Compensation method and proposed method is as in fig.4.



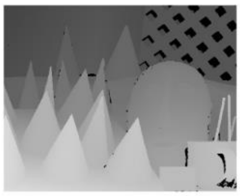
(a)



(b)



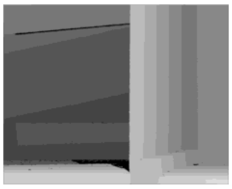
(c)



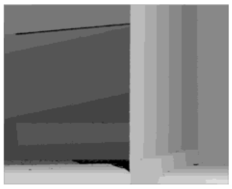
(d)

Figure 4: The Result of Cones with (a) symmetric smoothening (b) asymmetric smoothening (c) Adaptive Compensation method (d) proposed method.

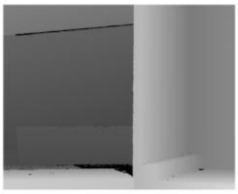
The Depth map of Woods pre-processed as per Symmetric filtering, Asymmetric filtering, Adaptive Compensation method and proposed method is as in fig.5.



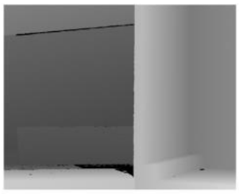
(a)



(b)



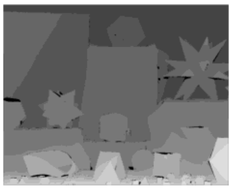
(c)



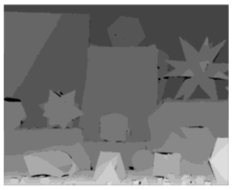
(d)

Figure 5: The Result of Woods with (a) symmetric smoothening (b) asymmetric smoothening (c) Adaptive Compensation method (d) proposed method.

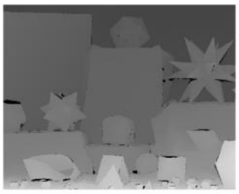
The Depth map of Gift pre-processed as per Symmetric filtering, Asymmetric filtering, Adaptive Compensation method and proposed method is as in fig.6.



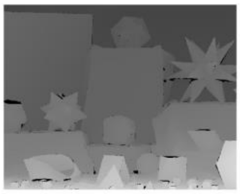
(a)



(b)



(c)



(d)

Figure 6: The Result of Gift with (a) symmetric smoothening (b) asymmetric smoothening (c) Adaptive Compensation method (d) proposed method.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

The virtual left views were generated for Cones, Woods and Gift using (4). The baseline distance is assumed as 36 pixels and the focal length is assumed as one for performance estimation. The left views generated for cones with depth map pre-processed using symmetric filtering; asymmetric filtering, ADC and proposed method are as in fig. 7.

(a)

(b)

(c)

(d)

Figure 7: The virtual left view of Cones with depth map pre-processed with (a) symmetric smoothening (b) asymmetric smoothening (c) Adaptive Compensation method (d) proposed method.

The left views generated for woods with depth map pre-processed using symmetric filtering; asymmetric filtering, ADC and proposed method are as in fig. 8.

(a)

(b)

(c)

(d)

Figure 8: The virtual left view of Woods with depth map pre-processed with (a) symmetric smoothening (b) asymmetric smoothening (c) Adaptive Compensation method (d) proposed method.

The left views generated for gifts with depth map pre-processed using symmetric filtering; asymmetric filtering, ADC and proposed method are as in fig. 9.

(a)

(b)

(c)

(d)

Figure 9: The virtual left view of Gift with depth map pre-processed with (a) symmetric smoothening (b) asymmetric smoothening (c) Adaptive Compensation method (d) proposed method.

The proposed method gives computationally more efficient as well as visually more plausible results than the existing methods. As the edges and details of the images are not blurred out, the proposed method results in a high PSNR and SSIM values compared to other existing methods. The PSNR and SSIM value comparison are as in Table 1. The increase in PSNR and SSIM value indicates that the proposed method produces a high visual quality result than Adaptive Compensation method, symmetric filtering and asymmetric filtering methods.

Table I. PSNR AND SSIM COMPARSION

Image	Method	PSNR	SSIM
Cones	Symmetric smoothening	18.6818dB	0.7323
	Asymmetric smoothening	18.6895dB	0.8527
	ADC	32.1279dB	0.9918
	Proposed method	33.4596dB	0.9927
Woods	Symmetric smoothening	20.1008dB	0.8450
	Asymmetric smoothening	20.1022dB	0.8735
	ADC	37.5981dB	0.9946
	Proposed method	41.7275dB	0.9964
Gift	Symmetric smoothening	16.2901dB	0.7694
	Asymmetric smoothening	16.2935dB	0.8173
	ADC	25.7542dB	0.9772
	Proposed method	26.5379dB	0.9800

The number of holes after rendering also reduces compared to ADC method and the computation time also shows a decrease. The comparison of No: of holes and Computation times are as in Table 2.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

Table II NO: OF HOLES AND COMPUTATION TIME COMPARISON

Image	Method	No: of Holes	Computation time(ms)
Cones	Symmetric smoothening	4108	614
	Asymmetric smoothening	4112	132
	ADC	5012	873
	Proposed method	4966	119
Woods	Symmetric smoothening	2018	87
	Asymmetric smoothening	2023	90
	ADC	2127	194
	Proposed method	2118	125
Gift	Symmetric smoothening	3929	101
	Asymmetric smoothening	3931	98
	ADC	4551	341
	Proposed method	4378	140

The proposed method produces a high visual quality depth map, which produces virtual images with less number of holes compared to other methods. The computation time is also considerably less.

V. Conclusion

The proposed method is a modification over Adaptive Compensation method which adopts a block-based strategy. The Improved adaptive compensation method gives a poor result in case of images having large holes. A new median filtering based ADC method is proposed for Mode 2 region to overcome this shortcoming. Edge preserving ability of median filtering along with texture based inpainting helps achieve speed efficiency, accuracy in synthesis of texture. The experimental results show that the proposed method produces an improvement in PSNR and SSIM values compared to the Adaptive Compensation method. The number of holes produced after warping is also less compared to the previous methods. The proposed method also cuts off the computation time significantly. This method can be adopted to produce better visual image quality for the rendered view.

References

1. Redert et al., "Advanced three-dimensional television system technologies",Proc. IEEE Int. Symp. 3D Data Process. Vis. Transmiss.,Jun. 2002, pp 313-319.
2. Chih-Hsien Hsia, "Improved Depth Image-Based Rendering Using an Adaptive Compensation Method on an Autostereoscopic 3-D Display for a Kinect Sensor",IEEE SENSORS JOURNAL., vol. 15,No.2,Feb. 2015
3. Ming-Fu Hung, Shaou-Gang Miaou, and Chih-Yuan Chiang, "Dual Edge-Confined Inpainting of 3D Depth Map Using Color Images Edges and Depth Images Edges," Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, Nov. 2015, pp. 1-9.
4. L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3D TV," IEEE Trans. Broadcast., vol. 51, no. 2, pp. 191-199, Jun. 2005.
5. W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, "Smoothing depth maps for improved steroscopic image quality," Proc. SPIE, vol. 5599, pp. 162-172, Oct. 2004.
6. P.-J. Lee and Effendi, "Nongeometric distortion smoothing approach for depth map preprocessing," IEEE Trans. Multimedia, vol. 13, no. 2, pp. 246-254, Apr. 2011.
7. Fehn, K. Hopf, and Q. Quanta, "Key technologies for an advanced 3D TV system,"Proc. SPIE, vol. 5599, pp. 66-80, Oct. 2004.
8. Middlebury Stereo Vision Database. [Online]. Available: http://vision.*middlebury.edu/stereo/data/*

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.