



Original research article

## Fast multi-view image rendering method based on reverse search for matching



JianPu Lin, WeiXing Wang, JianMin Yao, TaiLiang Guo, Enguo Chen\*, Qun Frank Yan\*

National and Local United Engineering Laboratory of Flat Panel Display Technology, Fuzhou University, Fuzhou, 350116, People's Republic of China

### ARTICLE INFO

**Keywords:**

Reverse search  
Matching  
Multi view  
Image generation algorithm  
Depth-image-based rendering  
Parallax matrix  
3D display  
3D image

### ABSTRACT

In a traditional multi-view image generation algorithm, partial image information might be lost at the pixel mapping step during the 3D image acquisition. A lower hardware cost and shorter operation time can be realized if an effective algorithm is employed. Therefore, a fast multi-view image rendering method based on reverse search is proposed in this paper, it uses pixel mapping information to derive a rendering image. Compared with the conventional depth-image-based rendering (DIBR) which using multiple images, the method only generates one image with reverse mapping in the image rendering step. First, the parallax range of the original image is calculated. Then, a warping image is generated through reverse mapping; this image has contained information for eight different viewpoints. Finally, the image modification can be made. Experimental results show that the proposed method effectively reduces rendering time and memory size by 58.82% and 83.78%, respectively. In addition, the mean structural similarity between the two images created by DIBR and our method is 1, which means that our method has the same effect as DIBR.

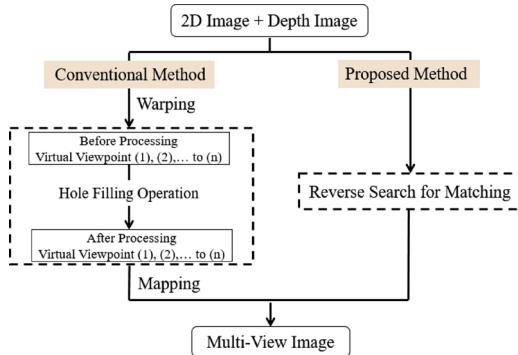
### 1. Introduction

Three-dimensional (3D) films have many advantages, such as strong telepresence and a more natural feel. In recent times, 3D stereo videos are gradually gaining interest among the public. However, compared with 3D glasses, naked-eye 3D perception without the aid of additional equipment has received increasing attention from researchers, especially the multi-view naked-eye 3D display [1]. Presently, multi-view 3D video resources can be created using DIBR (depth-image-based rendering) which is a multiple virtual-view synthesis production technology [2]. Compared with the manual generation method for multi-view 3D video, DIBR has several advantages, such as narrow transmission bandwidth [3], low cost [4], and arbitrary viewpoint image production capability [5].

A multi-view 3D video can be generated using DIBR based on naked-eye 3D display parameters, such as  $n$  number of views [6]. The algorithm for the construction of multi-view 3D video is described as follows. First,  $n-1$  virtual view images are obtained using a depth map; these virtual view images exhibit disocclusion after image warping. However, the disocclusion can be eliminated by using a hole filling algorithm. Then,  $n$  view point images are synthesized to a single multi-view image by combining original image and  $n-1$  processed virtual view images. In general, the original image, the virtual viewpoint images, and the multi-view image have the same resolution, i.e. only  $1/n$ th of the pixel information of each view image is used to generate the multi-view image, and the remaining pixel information is lost. This unused pixel information in the DIBR stage results in a large number of useless operations and leads to

\* Corresponding authors.

E-mail addresses: [ceg@fzu.edu.cn](mailto:ceg@fzu.edu.cn) (E. Chen), [qunfyang@gmail.com](mailto:qunfyang@gmail.com) (Q.F. Yan).



**Fig. 1.** Comparison diagram of the principle of multi-view 3D image generation between conventional DIBR method and our proposed method.

disadvantages of such as long computing time and high memory space consumption during the operation process [7]. This further results in increased hardware cost eventually hindering the real-time generation of multi-view images.

The existing approach use real-time image generation involves hardware design optimization [8]. For example, Lee et al. [9] put forward a method for hardware-accelerated DIBR; they suggested the use of surface rendering to account for varying illumination. The surface rendering is performed inside a vertex shade, while adaptive point splatting is performed inside the fragment shade. Do and Bravo [10] proposed a combination of highly parallel programming architecture, called Compute Unified Device Architecture (CUDA) and a graphics Application Programming Interface (API), along with signal processing blocks to share memory usage. Further, Jin et al. [11] used a Field-Programmable Gate Array (FPGA) for achieving a 3DTV system, which is based on the hardware cost and uses fast locally consistent dense stereo. However, the mentioned methods do not completely solve the fundamental problem of how to achieve real-time 3D video generation [12].

As shown in Fig. 1, we proposed algorithm uses the principle of reverse search for matching to generate multi-view 3D images directly from 2D images, simplifying the generation process of virtual viewpoint images, and effectively reducing operation time and memory usage; consequently, this can also provide a lower hardware cost.

## 2. Depth-image-based rendering

### 2.1. DIBR principle

DIBR is a virtual multi-view image generation technology based on depth images [13]. In brief, a depth map is used to generate virtual viewpoint images with hole filling to eliminate the existing disocclusion; then, all viewpoint images are mapped for a multi-view image.

The schematic diagram of the typical DIBR algorithm is shown in Fig. 2. Here, we take an 8-view image as an example. Using the original image and depth image, the algorithm will generate virtual image of views 1–8 based on the depth map; here, view 5 is the original image. Some disocclusion occurs because of the shielding effect [14] and can be eliminated using a hole filling process. Then, these viewpoint images are mapped as a multi-view image based on naked-eye 3D mapping rules. Because the viewpoint images have the same resolution, each viewpoint image occupies only 1/8th of a pixel and the remaining 7/8th of the pixel is left out. Consequently, one multi-view image contains the pixel information of all viewpoint images. However, in DIBR, because each viewpoint image leaves 7/8th of the pixel information unoccupied, an increase in program computing time occurs, along with the high memory usage to buffer all virtual viewpoint images.

### 2.2. Image mapping principle

Fig. 3(A) shows a mapping matrix [15] of the 8-view naked-eye 3D display screen. It can be concluded that sub-pixels with red colour among the pixels of the first row and the first column display red colour information of view 5, the sub-pixels with green colour among the pixels of the first row and the first column represent green colour information of view 6, and the sub-pixels with blue colour among the pixels of the first row and the first column indicates blue colour information of view 7, and so on. Fig. 3(B) shows the display state of view 1 after mapping where the black areas represent the unoccupied sub-pixels as only 1/8th of the sub-pixels are occupied. Therefore, the mapped image only retains 1/8th of the sub-pixel information for each viewpoint image with 7/8th of the sub-pixel information being abandoned.

### 2.3. Parallax value computation

Fig. 4 shows the calculation principle of parallax (8-view).

In Fig. 4,  $u$  represents the binocular distance and  $D$  represents the distance between the observer and the 3D screen.  $h$  and  $h'$  indicate the position of the object deep into and the object pointing out from the 3D screen, respectively.  $p$  is the parallax value.  $Z_{near}$

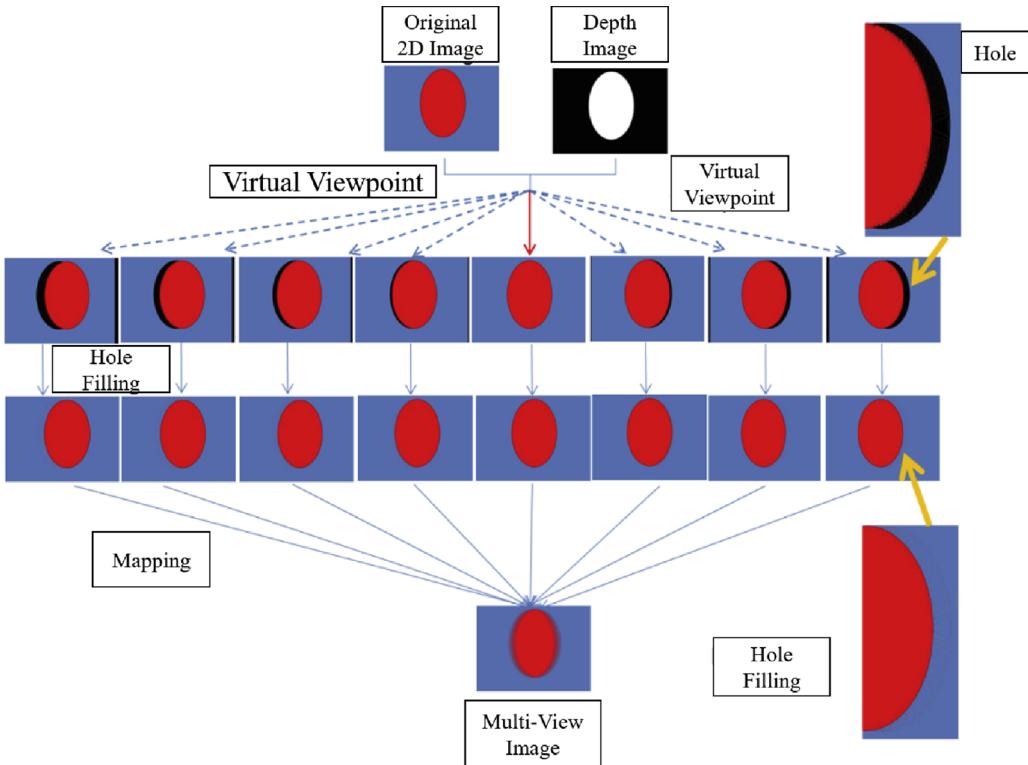


Fig. 2. Schematic diagram of a typical DIBR algorithm.

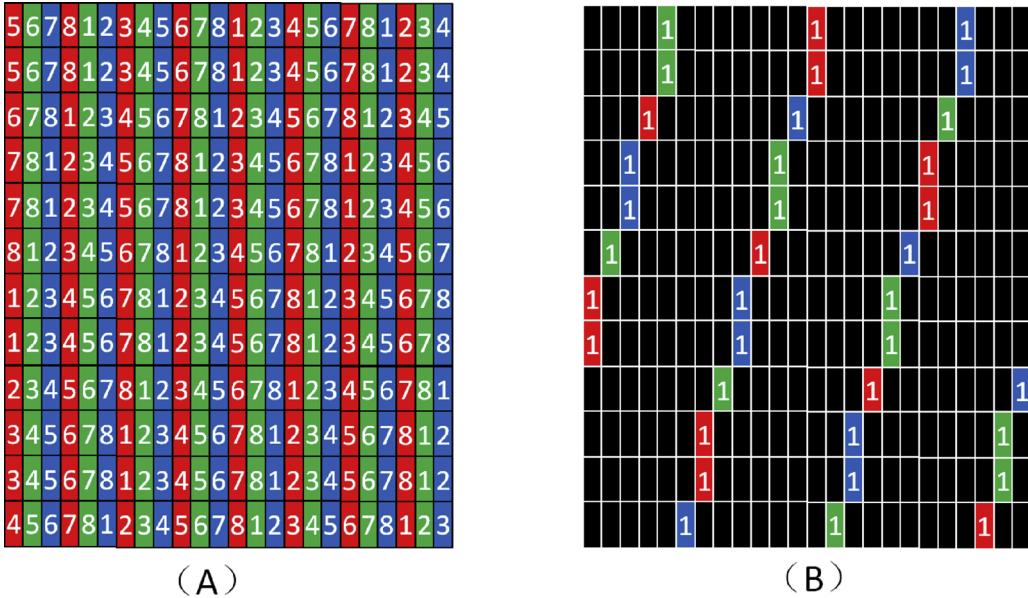
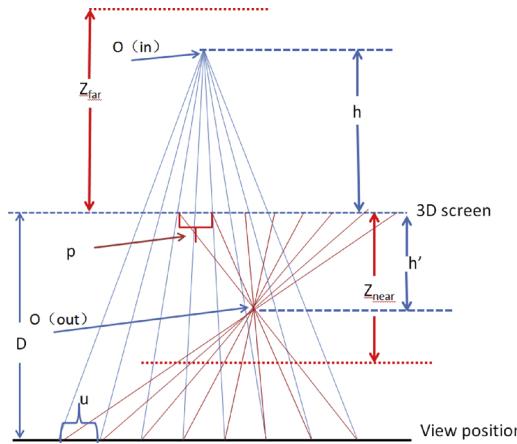


Fig. 3. Pixel arrangement of the 8-view naked-eye 3D screen. (A) A mapping matrix and (B) the view #1 display state after mapping.

and  $Z_{far}$  represents the image points in which one is the nearest and the other farthest from 3D screen, respectively. From Fig. 4,  $Z_{near}$  and  $Z_{far}$  can be derived as follows:

$$Z_{far} = \frac{a_1 * D}{u - a_1} \quad (1)$$



**Fig. 4.** Calculation principle of parallax.

$$Z_{near} = \frac{a_2 * D}{u + a_2} \quad (2)$$

where  $a_1$  is the maximum parallax value when an observer perceives the object jumping out from the 3D screen, and  $a_2$  is the maximum parallax value when an observer perceives the object deep inside 3D screen.

Discrete depth value can be quantified as follows:

$$j = \frac{(Z_{near} + Z_{far}) * g}{255} \quad (3)$$

where  $g$  represent the depth value. Then,  $h$  and  $h'$  can be computed as follows.

$$h = Z_{far} - j \quad (4)$$

$$h' = j - Z_{far} \quad (5)$$

Thus, the parallax value of the object deep into and the object jumping out of the 3D screen can be approximated using Eqs. (6) and (7), respectively.

$$p = \frac{u * h}{D + h} \quad (6)$$

$$p' = \frac{u * h'}{D - h} \quad (7)$$

Using Eqs. (1)–(7),  $p$  and  $p'$  can be re-derived as follows:

$$p = \frac{u * [(255 - g) \frac{a_1 * D}{u - a_1} - g * \frac{a_2 * D}{u + a_2}]}{255 * D + [(255 - g) \frac{a_1 * D}{u - a_1} - g * \frac{a_2 * D}{u + a_2}] \quad (8)}$$

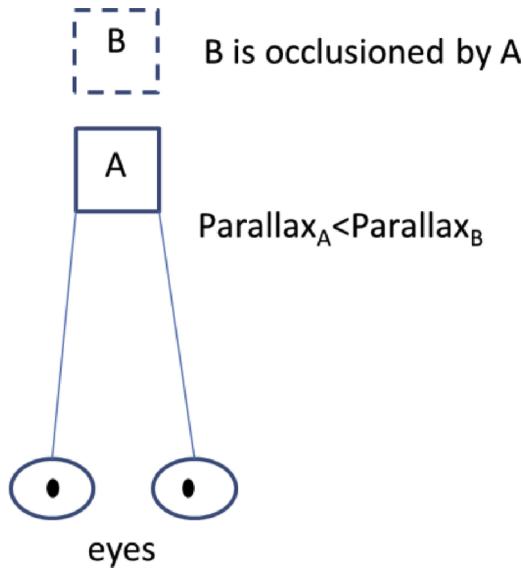
$$p' = \frac{u * [(g - 255) \frac{a_1 * D}{u - a_1} + g * \frac{a_2 * D}{u + a_2}]}{255 * D - [(g - 255) \frac{a_1 * D}{u - a_1} + g * \frac{a_2 * D}{u + a_2}] \quad (9)}$$

## 2.4. Shielding effect

As there is a possibility that a number of objects in the original image might meet the matching condition and warped toward the same position in the virtual image [16], those objects might result in the overlay area [17]. As can be seen in Fig. 5, the viewer is only able to view object A; however, there is only a limited view of object B because it is hidden behind object A. Therefore, using Eqs. (8) and (9), it can be seen that the parallax of object A is smaller than that of object B. Here, the selection matching process is based on the ascending order of parallax for pixel matching. The pixel matching will be terminated once the selection process meets the requirements calculated using reverse search matching equation to avoid overlapping [18].

## 2.5. Mean structural similarity (MSSIM)

MSSIM [19] is a function that is commonly used to examine the structural similarity between two images. If the value of MSSIM is



**Fig. 5.** Schematic of shielding effect.

close to 1, it indicates that the pixel information between two images have high similarity. Further, if the MSSIM value is 1, it indicates that the two images are the same [20].

## 2.6. Number of operations

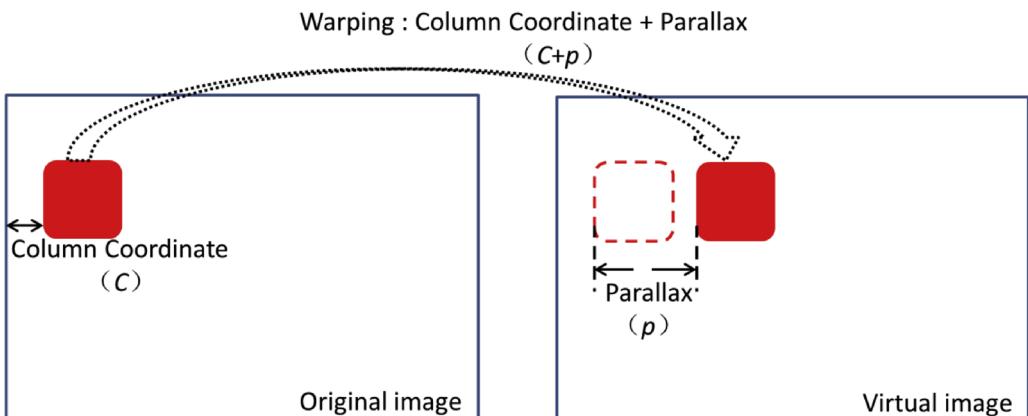
The typical DIBR method and our proposed method are both based on sub-pixel matching [21]. If we assume the resolution of the image is  $m \times n$  pixels (24 bits) and the number of viewpoints are  $k$ ,  $m \times n \times (k-1) \times 3$  sub-pixel operations are required for the typical DIBR method; however, our algorithm just requires  $m \times n \times 3 \times S_k$  sub-pixel operations, because  $S_k$  is usually smaller than  $k$ .

## 3. Reverse search and matching

### 3.1. Positive warping between the original and virtual images

In general, the parallax value of each pixel can be estimated using a depth map [22].

In Fig. 6,  $C$  represents the original image column coordinates [23]. The original image is warped from  $C$  to  $C + p$ , where  $p$  is the warping pixel. Hence, it is possible to derive a reverse warping relationship between the original image and the virtual image based on this positive warping estimation [24].



**Fig. 6.** Schematic diagram of original image warping based on estimated parallax value.

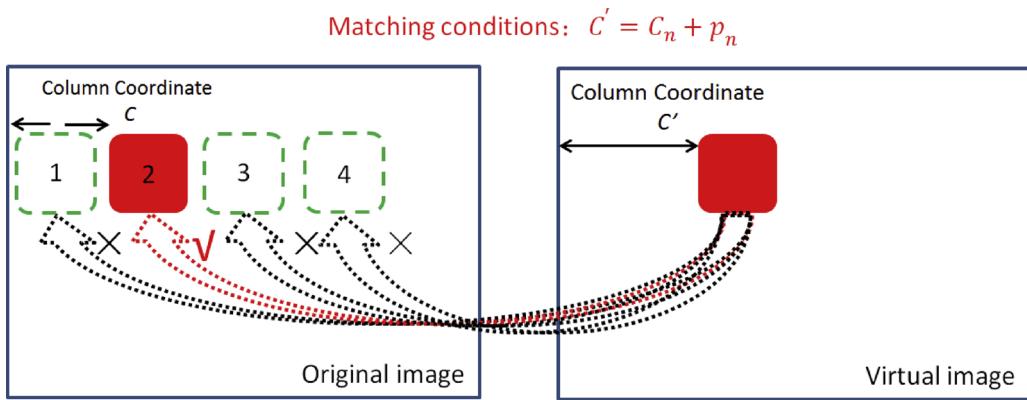


Fig. 7. Schematic diagram of reverse warping.

### 3.2. Reverse warping relationship between the original and virtual images

Fig. 7 illustrates the reverse warping between the original and virtual images. The coordinates of the virtual image column are estimated as  $C'$ . There are four possible objects warping  $c_n + p_n$  toward the virtual image with  $c_n$  ( $n = 1\text{--}4$ ) as the column coordinate and  $p_n$  ( $n = 1\text{--}4$ ) as parallax may generate [25]. As can be seen in Fig. 7, object 2 complies with the estimate condition and is therefore warped [26] from the original image to the virtual image. This reverse matching method provides an effective pixel-by-pixel operation in the multi-view image [27].

### 3.3. Principles of reverse search and matching

Fig. 8 shows the schematic diagram of reverse search and matching to obtain a single multi-view image. Using Eqs. (8) and (9) [19], it can be predicted that  $p$  and  $p'$  are 3 and -4, respectively, and the match range  $S_k$  is [-4,3]. Thus, when performing a search for the 13th column, the pixel searching range can be calculated using  $13 - S_k$ , i.e., for the 13th column, it ranges from the 10th to the 17th. From Fig. 8, the parallax value of 16th column is predicted to be -3, and this value will be used to warp from the original image to the virtual image during image warping [28,29].

Accordingly, the reverse search matching equation can be derived as follows:

$$a - S_k + p_{(i,a-S_k)} = a \quad (10)$$

where,  $a$  represents the column coordinate for searching,  $S_k$  represents the matching range as a one-dimensional vector array.  $p_{(i,a-S_k)}$  denotes the parallax value of the pixel in  $i$ -th row and  $a-S_k$ -th column.

## 4. Experimental processing

Fig. 9 shows the flow chart of the experiments. Fig. 9(A) is the flow chart of typical DIBR method, whereas Fig. 9(B) is the flow chart of our proposed reverse search method. We define those two methods as Approach-A and the Approach-B, respectively. Different methodologies were followed in these two approaches after parallax matrix calculation. In the proposed method, viewpoint position allocation is performed after parallax matrix calculation, followed by reverse matching multi-view hole filling. Nevertheless, it is noteworthy that, for both experiments, the operation time is recorded after the parallax matrix calculation step.

## 5. Results and discussion

### 5.1. Tools of simulation

The Software Platform: Matlab2010a, 64-bit.

The Hardware Platform:

Model: Apple Macbook pro A1398. CPU: Intel i7 3630QM. Memory: 8 GB DDR3 1600 MHz. Hard disk: 256 GB SSD. GPU: NVIDIA GeForce GT 650M + Intel GMA HD4000, 1 GB GDDR5.

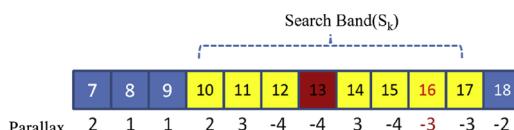
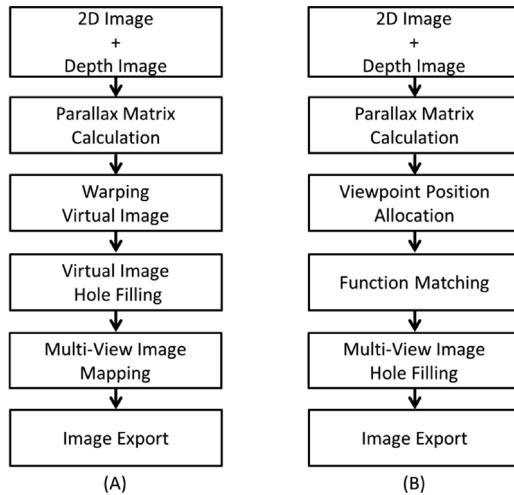


Fig. 8. Schematic diagram of reverse search and matching.



**Fig. 9.** Flow charts of Experiments. (A) Flow chart of typical DIBR method. (B) Flow chart of our Proposed Reverse Search Method (Approach-A and Approach-B, respectively).

Display Equipment: 8 Views Lenticular Lens Naked-eye 3D Display.

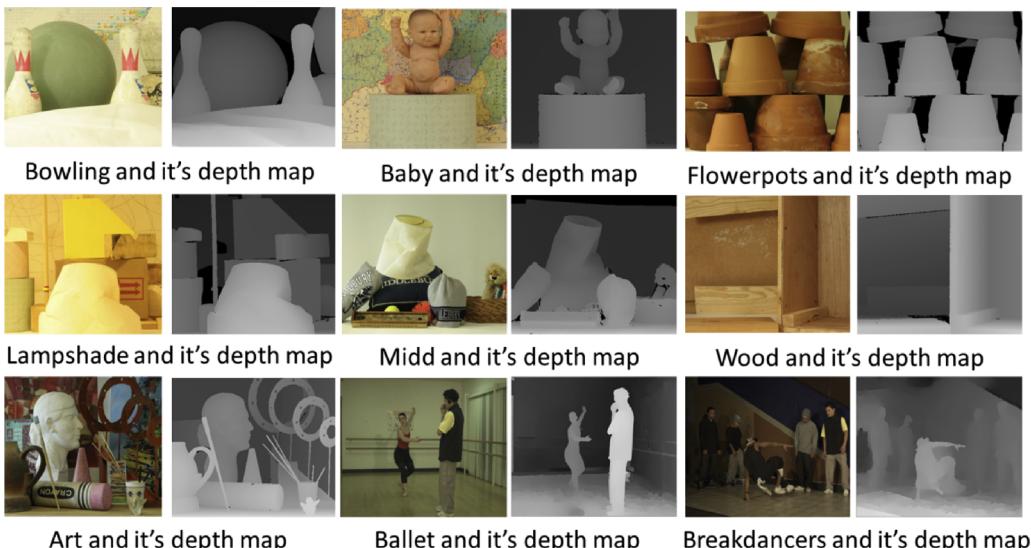
Experimental Scenes: Experimental scenes are shown in Fig. 10; these include Bowling, Baby, Flowerpots, Lampshade, Midd, Wood, ArtfromVISION, Breakdancers, and Ballet from Microsoft Research. The image resolution of these scenes is  $2160 \times 3840$  s.

## 5.2. Experimental results

The 8-view naked-eye 3D display is used for the video/image player in this experiment. Table 1 records the obtained experimental results ignoring the hole filling process in Approach-A and Approach-B, because the hole filling mechanisms used in both approaches are dissimilar and this difference might affect the accuracy of operation time comparison. Fig. 11 shows experimental results for the scene Baby.

In this scene, we use Approach-A and Approach-B to generate multi-view image, respectively. As is evident from Fig. 11 we see that the images generated by Approach-A and Approach-B are the same.

It can be seen from Table 1, the operation time to generate an 8-view image for Approach-A is 3.4 s and Approach-B just requires 1.4 s, which is 2.5 times faster. The memory usage reduced from 185 Mb to 30 Mb, which is a reduction of about 83.78%, and the MSSIM value was 1 indicating that the two pictures of 8-view images generated by both approaches is the same. Further, Approach-A required about 16.5 million warped operations to generate an 8-view image, whereas Approach-B just required about 2.3 million matching operations to create the same image, which is a reduction of about 86.06% in the number of operations.

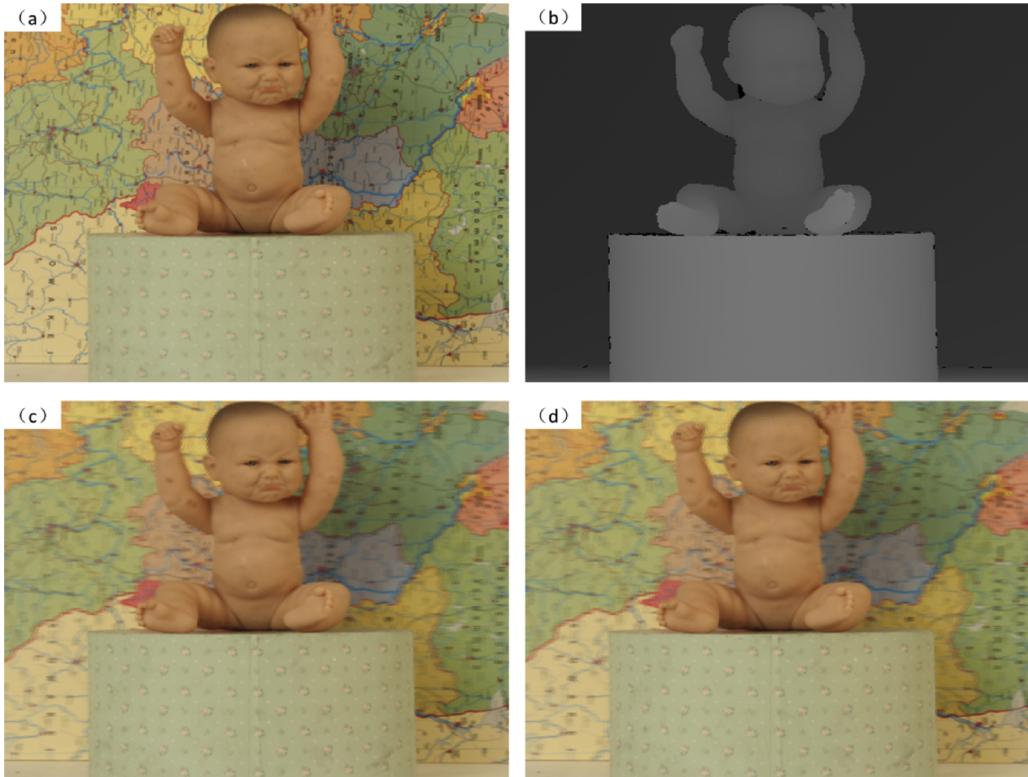


**Fig. 10.** Experimental Scenes.

**Table 1**

Eight-view Image Experimental Data for Approach-A and Approach-B.

Scene	Operation Time (Approach-A) (S)	Operation Time (Approach-B) (S)	Occupied Memory (Approach-A) (Mb)	Occupied Memory (Approach-B) (Mb)	Number of Operations (Approach-A)	Number of Operations (Approach-B)	MSSIM
Bowling	3.39	1.43	174	29	16458711	2329646	1
Baby	3.37	1.39	177	31	16446300	2337269	1
Flowerpots	3.35	1.34	188	29	16486377	2317089	1
Lampshade	3.36	1.41	192	29	16482348	2330912	1
Midd	3.41	1.44	181	29	16462473	2334175	1
Wood	3.38	1.38	181	30	16492734	2347262	1
Art	3.36	1.42	182	29	16474140	2309391	1
Ballet	3.35	1.45	180	30	16482822	2321791	1
Breaksancer	3.34	1.44	180	29	16484127	2338907	1



**Fig. 11.** Scene Baby Experiment Result. (a) Scene Image; (b) Depth Image, (c) Multi-view Image generated by Approach-A, and (d) Multi-view Image generated by Approach-B.

It can be concluded that by using Approach-B, the operation time and memory usage reduced significantly by 58.82% and 83.78%, respectively, which is in good agreement with our theoretical analysis.

## 6. Conclusion

We demonstrated a simplified reverse search approach to generate fast multi-views image. The proposed method significantly reduced the computing time and memory usage to obtain the same multi-view image. Our reverse search approach requires 41.2% lesser computing time as compared with the typical DIBR approach to generate the same 8-view image. Furthermore, our algorithm requires lesser memory of about 30 Mb compared with the typical DIBR, which requires 180 Mb to generate the same 8-view image. Moreover, the typical DIBR requires 16.5 million additional warping operations for sub-pixels to generate an 8-view image, whereas our proposed method requires only 2.3 million comparison operations for sub-pixels to generate the same image. The experimental data indicates that the proposed algorithm is has a higher efficiency than the typical DIBR algorithm, and provides a promising approach to generate multi-view images for real-world applications.

## Acknowledgements

This work is supported by National Natural Science Foundation of China (Grant No. 61405037); Fujian Science and Technology Key Project (2018H6011); Training Program of Fujian Excellent Talents in University (FETU).

## References

- [1] J. Konrad, M. Wang, P. Ishwar, 2D-to-3D image conversion by learning depth from examples, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRW), (2012), pp. 16–22.
- [2] Y. Feng, J. Ren, J. Jiang, Object-based 2D-to-3D video conversion for effective stereoscopic content generation in 3D-TV applications, *IEEE Trans. Broadcast.* 57 (2011) 500–509.
- [3] Y. Morvan, D. Farin, P.H.N. de With, Multiview depth-image compression using an extended H.264 encoder, *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems Vol. 4678, Lecture Notes in Computer Sciences*, (2007), pp. 675–686.
- [4] Y.C. Fan, B.L. Lin, S.Y. Chou, Hardware structure of 2D to 3D image conversion system for digital archives, *IEEE Int. Symp. on Consumer Electronics*, (2013), pp. 111–112 Jun. 3–6.
- [5] Y.K. Lai, Y.F. Lai, Y.C. Chen, An effective hybrid depth-generation algorithm for 2D-to-3D conversion in 3D displays, *J. Disp. Technol.* 9 (2013) 154–161.
- [6] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, C. Zhang, Multiview imaging and 3DTV, *IEEE Signal Process. Mag.* 24 (2007) 10–21.
- [7] P. Merkle, A. Smolic, K. Mueller, T. Wiegand, Multi-view video plus depth representation and coding, *Proceedings of the IEEE International Conference on Image Processing*, (2007), pp. I-201–I-204.
- [8] S.F. Tsai, C.C. Cheng, C.T. Li, L.G. Chen, A real-time 1080p 2D-to-3D video conversion system, *Proceeding of the International Conference on Consumer Electronics* 57 (2011) 915–922.
- [9] M.H. Lee, I.K. Park, Accelerating Depth Image-Based Rendering Using GPU. *Lecture Notes in Computer Science. MRCS 2006, LNCS 4105*, (2006), pp. 562–569.
- [10] L. Do, G. Bravo, S. Zinger, P.H. De With, GPU-accelerated real-time free-viewpoint DIBR for 3DTV, *IEEE Trans. Consum. Electr.* 58 (2012) 633–640.
- [11] M. Jin, T. Maruyama, A fast and high quality stereo matching algorithm on FPGA, *Field Programmable Logic and Applications (FPL)*, 22nd International Conference, (2012), pp. 507–510.
- [12] C.C. Cheng, C.T. Li, L.G. Chen, An ultra-low-cost 2D-to-3D conversion system, *SID Symp. Dig. Tech. Pap.* 41 (2010) 766–769.
- [13] A. Smolic, K. Muller, K. Dix, P. Merkle, P. Kauff, T. Wiegand, Intermediate view interpolation based on multiview video plus depth for advanced 3D video systems, *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP)* (2008) 2448–2451.
- [14] H.T. Nguyen, M.N. Do, Error analysis for image-based rendering with depth information, *IEEE Trans. Image Process.* 18 (2009) 703–716.
- [15] Y.C. Fan, Y.T. Kung, B.L. Lin, Three-dimensional auto-stereoscopic image recording, mapping and synthesis system for multi-view 3D display, *IEEE Trans. Magn.* 47 (2011) 683–686.
- [16] Y. Gao, X. Xiu, J. Liang, W. Lin, Perceptual multiview video coding using synthesized just noticeable distortion maps, *Proceedings of the 2011 IEEE International Symposium on Circuits and Systems (ISCAS)*, (2011), pp. 2153–2156.
- [17] W. Tao, H. Jin, Y. Zhang, Mean shift-A robust approach toward feature space analysis, *IEEE Trans. Syst. Man Cybern.* 37 (2007) 603–619.
- [18] D. Wang, J. Liu, J. Sun, W. Liu, Y. Li, A novel key-frame extraction method for semi-automatic 2D-to-3D video conversion, *Proceeding of the 2012 IEEE International Symposium Broadband Multimedia Systems and Broadcasting (BMSB)*, (2012), pp. 1–5.
- [19] L. Chai, Y. Sheng, Optimal design of multichannel equalizers for the structural similarity index, *IEEE Trans. Image Process.* 23 (2014) 5626–5637.
- [20] Y.C. Fan, P.W. Chen, Y.C. Chiu, A 2D-to-3D image conversion system using block slope pattern based vanishing point detection technology, *Proceedings of the 2012 IEEE International Symposium on In Computer, Consumer and Control (IS3C)*, (2012), pp. 321–324.
- [21] P. Merkle, A. Smolic, K. Muller, T. Wiegand, Efficient prediction structures for multiview video coding, *IEEE Trans. Circuits Syst. Video Technol.* 17 (2007) 1461–1473.
- [22] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, M. Tanimoto, View generation with 3D warping using depth information for FTV, *Image Commun.* 24 (1–2) (2001) 65–72.
- [23] L. Do, S. Zinger, Y. Morvan, P.H.N. de With, Quality improving techniques in DIBR for free-viewpoint video, *Proceedings of the 3DTV Conference 2009: The True Vision-Capture, Transmission and Display of 3D Video* (2009) 1–4.
- [24] S. Zinger, D. Ruijters, P.H.N. de With, iGLANCE project: free-viewpoint 3Dvideo, *Proceedings of the 17th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG)* (2009) 35–38.
- [25] D. Ruijters, S. Zinger, iGLANCE: transmission to medical high definition auto stereoscopic displays, *Proceedings of the 2009 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video* (2009) 1–4.
- [26] S. Zinger, L. Do, P.H.N. de With, Free-viewpoint depth image based rendering, *J. Vis. Commun. Image Represent.* 21 (2010) 533–541.
- [27] L. Do, S. Zinger, P.H.N. de With, Quality improving techniques for free-viewpoint DIBR, *IS&T International Electronic Imaging Symposium 7524* (2010) 10.
- [28] H. Liao, T. Inomata, I. Sakuma, T. Dohi, 3-D augmented reality for MRI-guided surgery using integral videography auto stereoscopic image overlay, *IEEE Trans. Biomed. Eng.* 57 (2010) 1476–1486.
- [29] H. Liao, H. Ishihara, H.H. Tran, K. Masamune, I. Sakuma, T. Dohi, Fusion of laser guidance and 3-D auto stereoscopic image overlay for precision-guided surgery, *Proceedings of Medical Imaging and Augmented Reality*, LNCS 5128, (2008), pp. 367–376.