A New Copyright Protection Scheme for Depth Map in 3D Video

Zhaotian Li, Yuesheng Zhu, Guibo Luo and Biao Guo
Communication & Information Security Lab, Institute of Big Data Technologies
Shenzhen Graduate School, Peking University
Shenzhen, Guangdong 518055 – China
[e-mail: lizhaotian@sz.pku.edu.cn, zhuys@pkusz.edu.cn, luoguibo@sz.pku.edu.cn and guob@sz.pku.edu.cn]
*Corresponding author: *Yuesheng Zhu*

Received December 6, 2016; revised April 19, 2017; accepted April 24, 2017; published July 31, 2017

Abstract

In 2D-to-3D video conversion process, the virtual left and right view can be generated from 2D video and its corresponding depth map by depth image based rendering (DIBR). The depth map plays an important role in conversion system, so the copyright protection for depth map is necessary. However, the provided virtual views may be distributed illegally and the depth map does not directly expose to viewers. In previous works, the copyright information embedded into the depth map cannot be extracted from virtual views after the DIBR process. In this paper, a new copyright protection scheme for the depth map is proposed, in which the copyright information can be detected from the virtual views even without the depth map. The experimental results have shown that the proposed method has a good robustness against JPEG attacks, filtering and noise.

Keywords: depth image based rendering, depth map, virtual views, copyright protection, morphological dilation.

https://doi.org/10.3837/tiis.2017.07.014 **ISSN : 1976-7277**

1. Introduction

The popularity of 3D videos is on the rise in recent years, therefore the demand for 3D content has increased gradually. There are mainly two major ways to produce 3D videos. One is directly capturing left view and right view by two cameras simultaneously. The other method is using 2D-to-3D conversion system. Compared with the first method, the second one can directly use 2D video and corresponding depth map to synthesis left-eye and right-eye view which can enrich the existing 3D resources. The framework of 2D-to-3D conversion system is shown in **Fig. 1**.

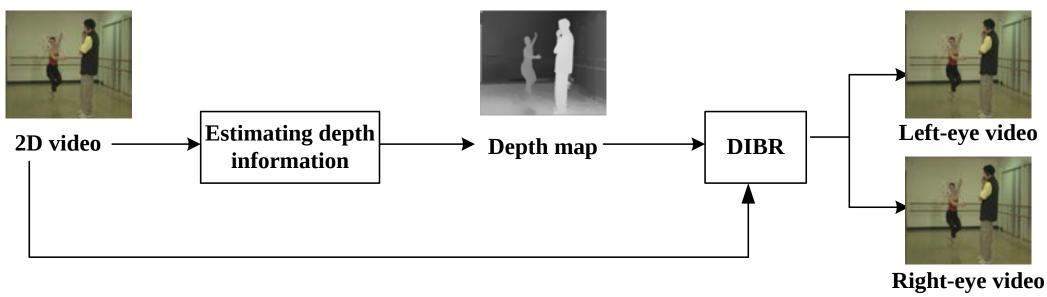


Fig.1. The framework of 2D-to-3D conversion system

The system contains two parts: (1) the estimation process of depth information. This part is to estimate depth information from 2D videos by using depth cue. (2) the depth-image-based rendering (DIBR) process. The left-eye and right-eye view can be generated by DIBR method.

It is easy for human to have a perception for complicated scene structure in the 2D video, but for computer vision, it is hard to understand the stereo structure of plain image. Estimating the depth information from 2D video is a challenge in the field of computer vision. So the depth map plays a unique role in the process of 2D-to-3D conversion system and the problem of how to protect depth map should be taken into account for 3D video synthesis.

2. Related work

As for information security and content protection for 3D videos (2D video plus depth map format), digital watermarking is widely used to avoid the illegal distribution of depth map and 3D video. There are some digital watermarking schemes [1]-[4] proposed for 3D video protection in which the copyright information is embedded into 2D video. The protection framework can be illustrated in **Fig. 2**. The copyright information can be embedded into the 2D videos, then the watermarked 2D video and corresponding depth map are transmitted to the receiver side. Then the receiver side performs DIBR operation to generate the left-eye and right-eye images. The copyright information can be extracted from the virtual view to prove the ownership of the 3D contents.

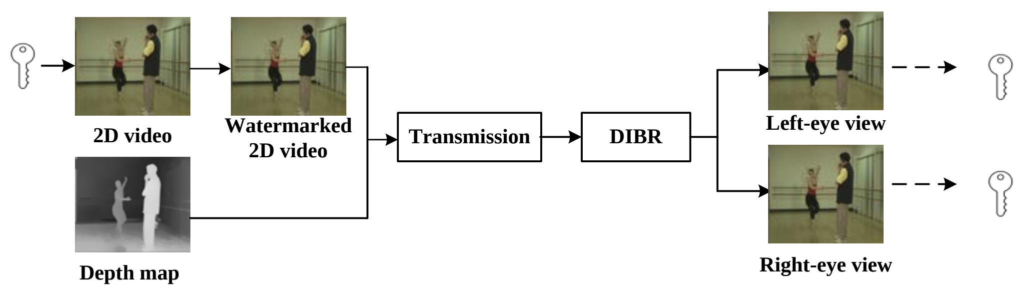


Fig. 2. Protection process for 2D videos, where the copyright information is embedded into the 2D video

As we known, the quality of depth map determines the three dimensional effect of 3D video. So the depth map plays an important role in the 2D-to-3D conversion system and the protection for depth map cannot be ignored. The above methods aim to the protection for 2D video and 3D video. However, they cannot protect the copyright of the depth map.

There are a few watermarking studies [5]-[7] proposed for the protection of depth map. As illustrated by the Unseen Visible Watermarking (UVW) scheme [5] [6], hidden information can be embedded and imperceptible under the normal rendering conditions and recognized by human visual system when changing the rendering conditions. The embedding regions in method [5] are the farthest regions in depth map which can avoid the situation that the disparity value is overflowed. The method [6] proposed a D-nose model to modify the suitable regions mentioned in method [5]. Both scheme [5] and [6] embed watermarking into depth map in the spatial domain , so it shows weak poor robustness to various attacks. The method [7] proposed for the protection for depth map with the scheme based on Quantization Index Modulation (QIM) algorithm. In this method, the copyright information is embedded in the DCT coefficients of depth map and the watermark can be extracted from depth map with IDCT. Meanwhile, it shows a good performance in perceptibility and robustness compared with method [5].

The framework of above methods [5]-[7] can be simplified shown in **Fig. 3**, they have two common features: (1) the depth value is changed when the watermarking information is embedded, but the variation is limited to a small range to guarantee the virtual views have zero distortion after the DIBR process. (2) the embedding and extraction of the copyright information are all conducted within the depth map.

Generally, a copyright protection technique should consider the situations not only in which the depth maps are illegally distributed, as shown in **Fig. 4** (Illegal distribution 1), but also the circumstance that virtual views are leaked to the public without permission, as shown in **Fig. 4** (Illegal distribution 2). In practical, the Illegal distribution 2 is more common than the Illegal distribution 2 as the depth map is the intermediate product in 2D-to-3D conversion system and only the virtual views are the final products directly exposed to users. The unlawful distribution process is shown in **Fig. 4**.

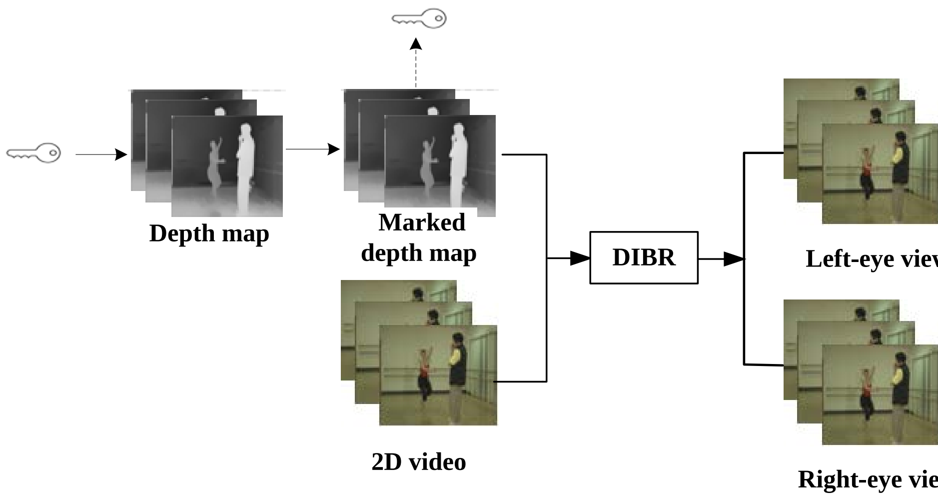


Fig. 3. Protection process for depth map, where the embedding and extraction process are all conducted in depth map.

The methods [5]-[7] only consider the situation of "Illegal distribution 1", but they cannot protect the depth map in the situation of "Illegal distribution 2" as the copyright information does not preserve in the virtual view. To our knowledge, there is no study for the protection of depth map in the situation of Illegal distribution 2. This is the problem we concern in this paper.

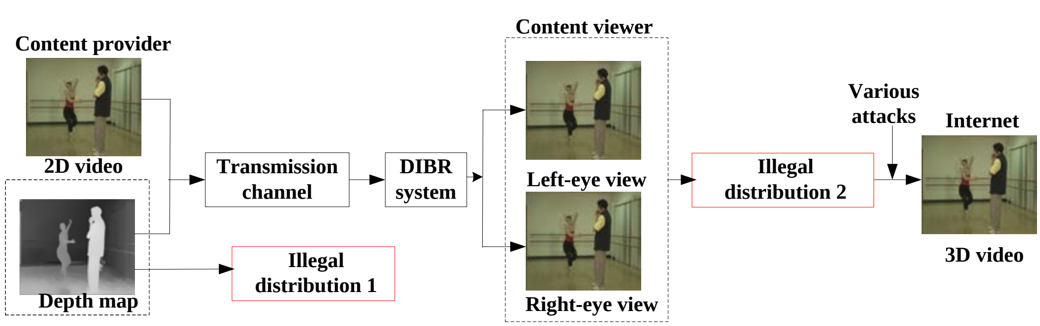


Fig. 4. Illegal distribution by content viewers

In this paper, a new copyright protection method for the depth map is proposed, in which the copyright information can be detected in the virtual views as shown in **Fig. 5.**

The rest of this paper is organized as follows. The difficulty and key point for the proposed method are shown in section 2. The proposed scheme is described in Section 3. Then, the experimental results are presented in Section 4. Finally, the conclusions are given in Section 5.

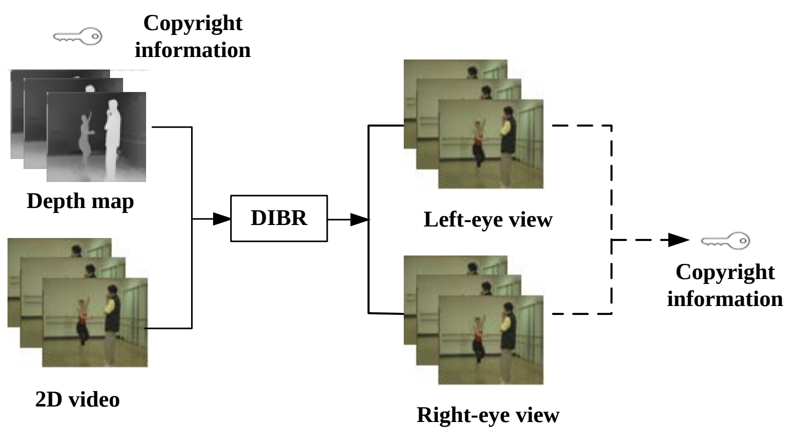


Fig. 5. The goal of the proposed method.

3. Proposed method

For the copyright information embedding into the depth map but extracting from virtual views, the following difficulties and key points are necessary to consider:

(1) The uniqueness of embedding and extraction process: the copyright embedding process is conducted on the depth map while the extraction process is conducted on the virtual views. This is completely different from the traditional watermark's method which is conducted on the same carrier. Meanwhile, the DIBR process would produce a great loss to the watermark, so it increases the difficulty to extract watermark effectively.

(2) The invisibility of the watermark: the watermark embedded in the depth map must not cause the visual perception of the human eye. If there is too much variation of depth map, it will destroy the original depth information.

(3) The fidelity of virtual views: the watermark embedded in the depth map will affect the virtual views. In the embedding process, the distortion of virtual view must be small.

(4) Reliable extraction of watermark information: the watermark information is hidden in the left and right-eye view. It is not possible to extract the watermark only from a single view. It is necessary to restore the depth information according to both the left and right eye view, then it is possible to get the watermark information embedded in the depth information.

However, the above key points are contradictory: according to the first key point, the copyright information can be extracted from virtual views only if the variation of depth map is sufficient enough. But too much variation in depth map is contradictory with the invisibility of the watermark in key point 2 and small distortion of virtual views in key point 3. If the variation of depth map is too small, it cannot guarantee the accuracy of the watermark extraction result in key point 4.

Aiming to this problem, there has no effective research as the above constraints ahead of the proposed method.

The framework of the proposed scheme consists of two parts, including the embedding module and the extraction module, as shown in **Fig. 6**. Meanwhile, Theoretical analysis for the embedding module in DIBR process is given in this part.

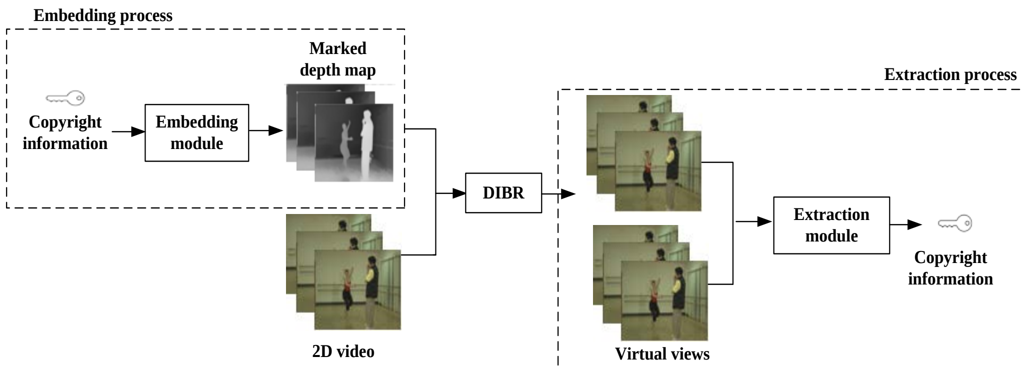


Fig. 6. The framework of the proposed scheme

3.1 Embedding module

This part focuses on how to embed the copyright information into the depth map to ensure the copyright information still reserved in the virtual views after DIBR process.

A. The background of zero distortion theory

In order to reach this target, the relationship between depth map and virtual views is analyzed in this part.

The left-eye and right-eye image can be synthesized with 2D image and depth information [8] by using rendering operation as follows:

$$x_l = x_c + \frac{f * t_c}{2Z} \tag{1}$$

$$x_r = x_c - \frac{f * t_c}{2Z} \tag{2}$$

where x_c , x_l , x_r are the x-coordinate of a pixel in the 2D view, left-eye and right-eye view, respectively. f is the focal length of the camera, t_c is the baseline distance, Z represents the depth value of current pixel in the real-world. The horizontal shifting distance d can be calculated as:

$$d = x_l - x_r = \frac{f * t_c}{Z} \tag{3}$$

The real-world depth values Z can be converted to frame depth value I according to [9]:

$$I = 255 * \left(\frac{1}{Z} - \frac{1}{Z_{far}}\right) / \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) \tag{4}$$

where Z_{near} is the nearest clipping plane, Z_{far} is the farthest clipping plane in 3D scene. Using the equation (3) and equation (4), the relationship between I and d can be represented as follows:

$$d = f * t_c * \frac{I}{255} * [(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}) + \frac{1}{Z_{far}}] \tag{5}$$

Studies [5]-[7] show the disparity d depends linearly on depth value I when Z_{far} is considered ∞. According to the equation (5), the relationship between depth value I and disparity value d can be simplified as follows:

$$d = \frac{f * t_c * I}{255 * Z_{near}} \tag{6}$$

When the copyright information is embedded into the depth map, assume the variation of depth value is $\triangle I$, then the variation of disparity is calculated as:

$$\triangle d = \frac{f * t_c * \triangle I}{255 * Z_{near}} \tag{7}$$

When making $|\triangle d| < 1$, the disparity variation of x_l and x_r is $\triangle d / 2$,they are less than 1. Therefore, the change of depth map will has no effect on the synthesized views.

Let $\triangle d = 1$, then

$$\triangle I = \frac{255 * Z_{near}}{f * t_c} \tag{8}$$

So when the variation of depth value does not exceed $\triangle I$, there will has no effect on the virtual views as $\triangle d$ is smaller than 1.The studies [5]-[7] use this feature to embed watermark into depth map.

On the contrary, if the depth value changes more than $\triangle I$, the virtual views might also have some changes and preserve the copyright information. The proposed embedding module adopts this feature to transfer the copyright information from the depth map to the virtual views.

B. Embedding process with dilation operation

As the variation of depth value $\triangle I$ is not a small value any more, where it is suitable to embed the copyright information without visual perceptual is another problem. For human perception, it seems obvious if the flat regions of the depth map have slight change, while it is not easy to aware the change even if the sharp edge regions of the depth map have some changes. So the copyright information should be embedded in the sharp edge regions in the proposed embedding module. Here we apply dilation morphological operation to process the depth map. The dilation operation only extends a few pixels of foreground objects along the boundary and this change is not easy to raise the awareness of human perception.

The proposed embedding module is shown in **Fig. 7**. The copyright information is chosen arbitrarily, which can be converted to a bit sequence by an encoding method (ASCII coding is used in our experiment). Each frame of depth map is only embedded one bit. The depth map is processed with dilation operation when is embedded bit ‘1’, otherwise nothing to do on the depth map.

If let v denote the bit sequence, v_i denote the i^{th} bit value in the bit sequence v, I_i denotes the i^{th} frame depth map, then the depth map is processed as follow:

$$I_i = \begin{cases} I_i \oplus B, & if\ v_i = 1 \\ I_i, & otherwize \end{cases} \tag{9}$$

where B is the structuring element with size $L \times L$. In our experiment, the parameters L is selected as 5.

After the above process, the original depth map sequence turns to the marked depth map sequence which contains copyright information.

In the following DIBR process, the dilated depth map is treated as marked depth map which is used to synthesis the marked virtual views. The copyright information can be transferred from the depth map to the virtual views. So the virtual views contain the watermarking information which is originally embedded in the depth map. At last, with the proposed extraction method, copyright information can be detected from virtual views in the extraction process.

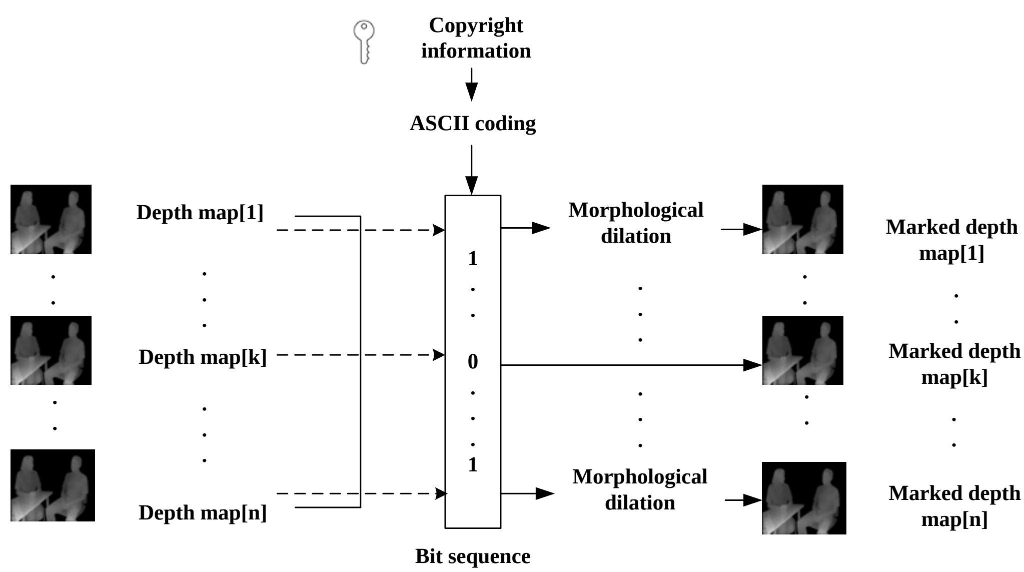


Fig. 7. The embedding process with dilation operation

3.2 Theoretical analysis for the embedding module in DIBR process

This part will present a theoretical analysis about the effect when the depth map is processed with morphological dilation. Meanwhile, it can explain how the copyright information transfer from the depth map to the virtual view, and the distortion of virtual views is little in the DIBR process.

The DIBR process includes two parts: 3D warping and hole filling process. In order to simplify the demonstration, a row of pixels is selected to illustrate the 3D warping and hole filling process by using the original depth map and the dilated depth map, as shown in **Fig. 8(a)** and **Fig. 8(b)**, respectively.

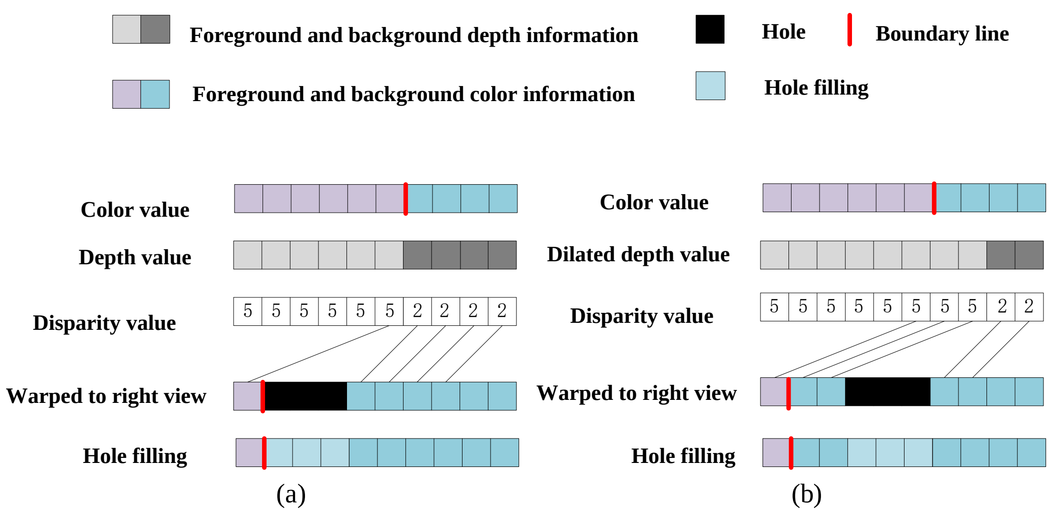


Fig. 8. The process of 3D warping and hole filling. (a)The process with original depth map. (b) The process with dilated depth map.

In **Fig. 8**(a), assume that the disparity difference between foreground pixels and background pixels is 3, and the foreground pixels shift to left by 5 pixels and the background pixels shift to left by 2 pixels after 3D warping. Consequently, a 3 pixels width hole will be produced in virtual view. This hole will be filled with depth-based hole-filling method [10] according to the neighboring background pixels information.

In **Fig. 8**(b), when a depth map is processed with dilation operation, the region of bigger depth value pixel will be extended by several pixels. The disparity value only depends on the depth value, so the boundary of bigger disparity object is also extended by several pixels. The holes region pixels as shown in the 4th row of **Fig. 8**(b), shifting several pixels near the boundary of foreground and background compared with 4th row of **Fig. 8**(a).

The sufficient change of depth value causes the variation of disparity value, the result of 3D warping also has some changes. In the hole filling process, the hole is filled by neighboring background pixels information, and background pixels usually have similar texture, so after the hole filling process, the marked virtual views can still have similar texture compared with original virtual views, as shown in the 5th row of **Fig. 8**. One example of the comparison result is shown in **Fig. 9**.

Fig. 9. (a) original depth map; (b) the dilated depth map; (c)3D warping with the original depth map; (d) 3D warping with the dilated depth map; (e) hole filling for (c); (f) hole filling for (d).

3.3 Extraction module

The extraction module is to extract the copyright information from virtual views without provided depth map. Each mark value of the depth map is detected from each corresponding pair of virtual views. The framework of the extraction module outlined in **Fig. 10**.

In the extraction module, the virtual views A is generated from 2D video and the original depth map, and the virtual views B is generated from 2D video and the marked depth map.

Using stereo matching method [11], disparity map A and disparity map B is recovered from the virtual views A and B, respectively. Then "XOR" operation is used to get the difference area of disparity map A and disparity map B. If the difference area percentage p_i is bigger than the threshold β, for the i^{th} frame, the mark value can be gotten as follow:

$$\hat{v}_i = \begin{cases} 1 & if \;\; p_i \geq \beta \\ 0 & otherwise \end{cases} \tag{10}$$

where $\beta = average(p)$, p is the set of p_i.

All virtual views frames in the sequence are processed with the above operation to obtain the mark value of all depth maps, so a bit sequence is obtained. Finally, the copyright information can be gotten from the bit sequence by a decode method that corresponds to the encode method (ASCII coding is used in our experiment).

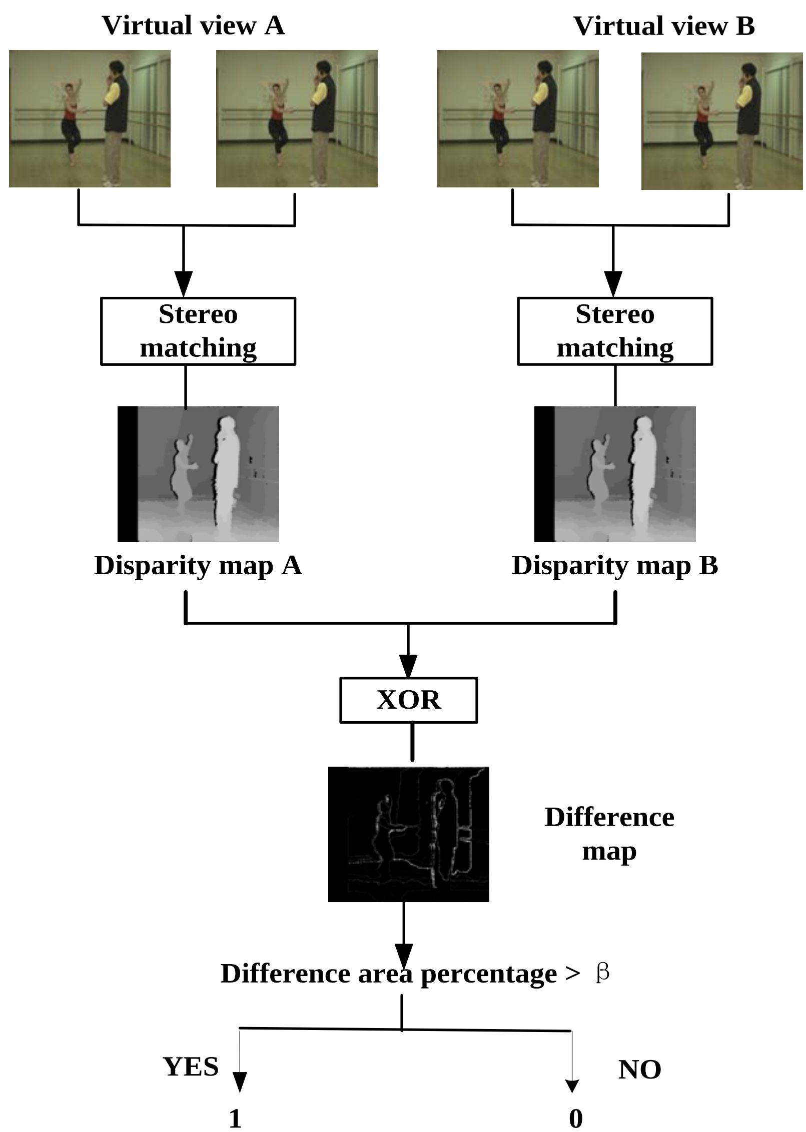


Fig. 10. The extraction process for each pair of virtual views

4. Experimental Results and Analysis

In our experiment, four Video-plus-Depth (MVD) sequences ("Ballet", "Break dancers" [12], "PoznanHall2", "PoznanStreet", "Dancer" [13]) are used to evaluate the performance of the proposed method. The parameters of the dataset are shown in **Table 1**. The imperceptibility of the dilated depth map and marked virtual view, and robustness of the copyright information are evaluated. To our knowledge, there is no study to detect the copyright information from virtual views for the protection of depth map, so the comparison among other methods is not evaluated.

Table 1. Parameters of Datasets

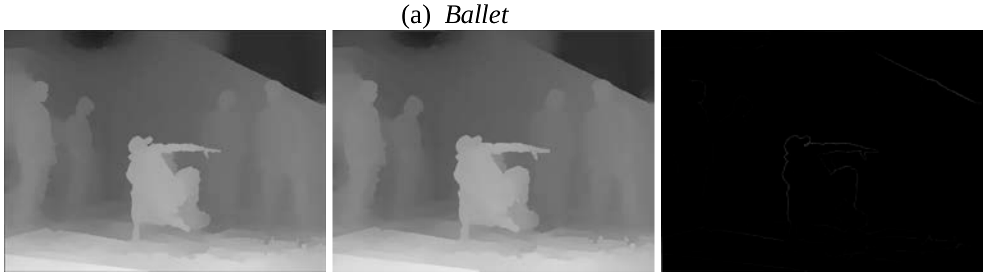
Name	Resolution	Frames	Scene
Ballet	1024 × 768	1–100	Stationary
Breakdancers	1024 × 768	1–100	Stationary
PoznanHall2	1920 × 1088	1–100	Dynamic
PoznanStreet	1920 × 1088	1–100	Stationary
Dancer	1920 × 1088	1–100	Dynamic

Scene: the camera is motion or not.

4.1 Imperceptibility

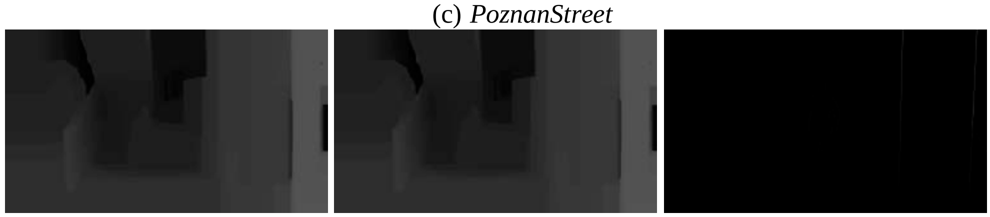
Fig. 11 shows the difference between the original and the dilated depth map. The first column and second column in **Fig 11** shows the original and the dilated depth map, respectively. In the third column, the white area representing the differences between the original depth map and dilated depth map. The change of the dilated depth map is almost invisible as only a few pixels are extended along the foreground objects. The PSNR of the dilated depth map is about 23 dB. It is the virtual views that are directly provided to viewers, so it is tolerable that the depth map does not have high quality when it still cannot cause awareness of visual perception.

(a) *Ballet*



(b) *Breakdance*

(c) *PoznanStreet*



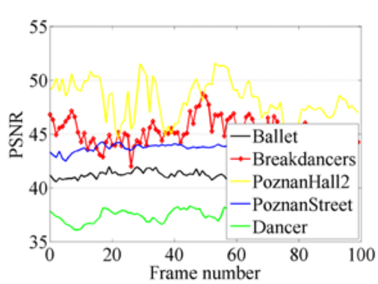
(d) *PoznanHall2*

Fig. 11. Imperceptibility of the dilated depth map.

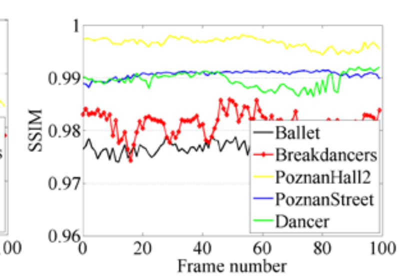
The non-hole area PSNR (Peak Signal-to-Noise Ratio) [14] is used to measure the squared intensity differences of the original and marked virtual view non-hole pixels, and SSIM (structural similarity) [15] is used to measure the structural similarity between the original and marked virtual view pixels. The PSNR and SSIM values for each synthesized frame of the tested sequences are shown in **Fig. 12**. The PSNR and SSIM values of the marked virtual view surpass 40 dB and 0.97. However, "Dancer" sequence has relative low PSNR as it has very complex background. Compared with the original virtual view, the marked virtual view has nearly no synthesis error, as shown in **Fig. 13**.

It is important to know how the viewer feels about the synthesized videos which contain the copyright information from a perceptual perspective, so the subjective evaluation is also performed with these video data sets by ten participants. The participants have watched the original synthesized views and marked synthesized views independently of each other and were asked to give their own scores. The score is from 1 to 5 which reflects the degree of similarity, where 1 stands for the largest difference while 5 stands for the same.

The average scores are obtained and used as a measure of the subjective evaluation. The average scores of the dataset are shown in Table II, respectively. The results show the proposed method does not affect the image quality.



(a)



(b)

Fig. 12. The synthesis error for each synthesized frame if the depth map is conducted by dilation operation. (a) PSNR values; (b) SSIM values.

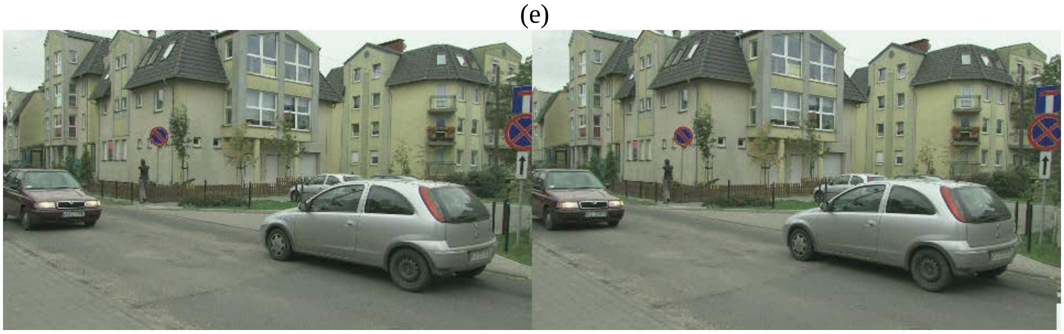

(a)

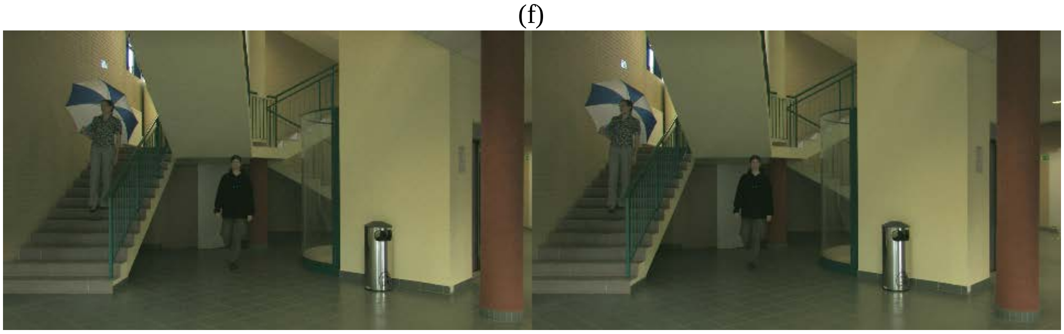

(b)

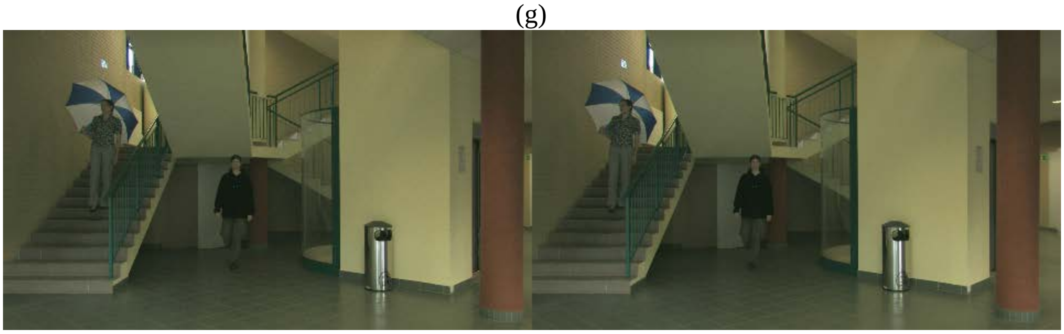

(c)


(d)

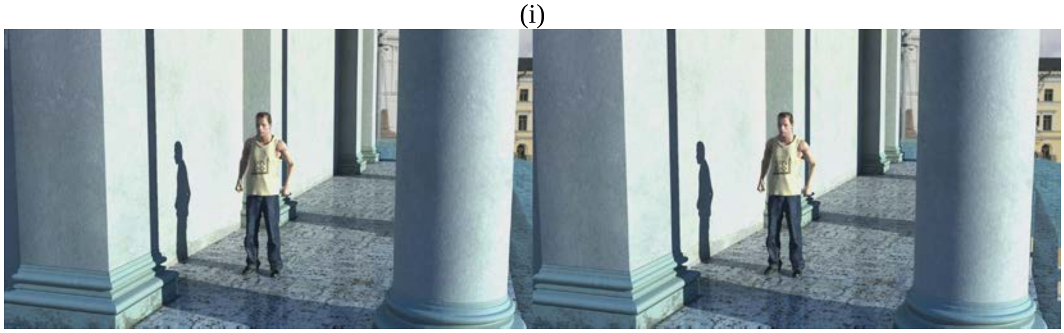

(e)


(f)


(g)


(h)

(i)



(j)

Fig. 13. Visual comparison for original virtual view and marked synthesized view. (a) (c) (e) (g) (i) original virtual view of "Ballet" , "Breakdancers", "PoznanStreet", "PoznanHall2" , and “Dnacer”, respectively; (b) (d) (f) (h) (j) marked virtual view of "Ballet" , "Breakdancers", "PoznanStreet", "PoznanHall2" , and “Dnacer”, respectively.

Table 2. Subjective evaluation of datasets

Dataset	Ballet	Breakdance	Dancer	PoznanHall2	PoznanStreet
Average score	4.5	4.8	4.8	4.9	4.9

4.2 Robustness

In this part, the robustness of the proposed method is evaluated. As the depth map is not directly exposed to viewers and only the virtual views are distributed illegally, the signal distortions are applied to the virtual views to evaluate the robustness of extraction accuracy. The signal distortions consist of salt noise, filtering, scaling, and JPEG compression. For salt noise, noise density 0.04 and 0.08 are tested respectively. For filtering, median filter, mean filter and Gaussian filter are used for 3*3 filtering windows. For scaling, scaling factor 0.8 and 0.5 are tested respectively. For JPEG compression, the compression ratio is up to 30. **Table 3** and **Fig. 14** illustrate the extraction accuracy rate.

According to the experimental results, our proposed method has a good accuracy rate against the signal distortions. The extraction accuracy rate of "Ballet" and "PoznanHall2" sequence reaches 100%. The extraction accuracy rate of "Breakdancers" and "PoznanStreet" sequence does not reach 100% as it has relatively complex background texture or some inaccuracies in the original depth map. As for JPEG compression, the accuracy rate of "Ballet", "Breakdancers" and "PoznanHall2" reach almost 100% when the compression ratio is less than 20.

As method [5] [7] adopt zero-synthesis-error theory to embed watermarking into depth map, the watermarking cannot be detected from the synthesized view. **Table 4** shows the extraction accuracy of method [5] [7] are all 0.

Table 3. The extraction accuracy rate of proposed method based on noise, filtering and scaling

	Ballet	Breakdancers	PoznanHall2	PoznanStreet
Salt noise 0.04	100%	100%	100%	99%
Salt noise 0.08	100%	97%	100%	86%
mean filter (3*3)	100%	100%	100%	99%
median filter (3*3)	100%	100%	100%	100%
Gaussian filter (3*3)	100%	100%	100%	100%
Scaling(0.8)	100%	100%	100%	100%
Scaling(0.5)	100%	97%	99%	88%

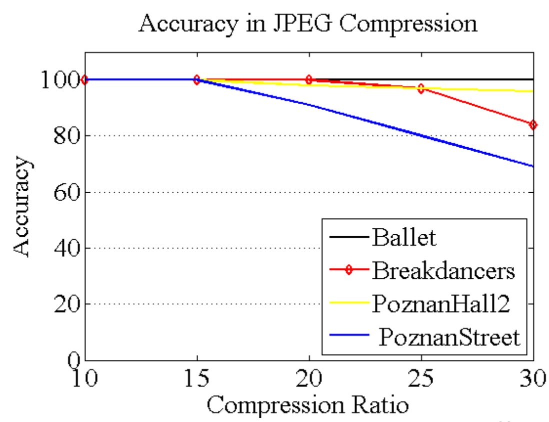


Fig. 14. The extraction accuracy in JPEG compression with different compression ratios

Table 4. The extraction accuracy rate of method [5] [7] based on noise, filtering and scaling

	Ballet		Breakdancers		PoznanHall2		PoznanStreet	
	[5]	[7]	[5]	[7]	[5]	[7]	[5]	[7]
Salt noise 0.04	0	0	0	0	0	0	0	0
Salt noise 0.08	0	0	0	0	0	0	0	0
mean filter (3*3)	0	0	0	0	0	0	0	0
median filter (3*3)	0	0	0	0	0	0	0	0
Gaussian filter (3*3)	0	0	0	0	0	0	0	0
Scaling(0.8)	0	0	0	0	0	0	0	0
Scaling(0.5)	0	0	0	0	0	0	0	0
JPEG compression	0	0	0	0	0	0	0	0

5. Conclusion

In this paper, we analyzed how to obtain the copyright information of depth map from virtual views and proposed a new method to protect the depth map and 3D video. Due to the property of DIBR, the challenges are not only to make the copyright information transfer from depth map to virtual views, but also to keep the good visual quality of marked depth map and virtual views. We analyzed the property of DIBR process and applied it to protect the depth map. In the embedding process, one bit information is embedded into the sharp regions of depth map. In the extraction process, the difference map is generated from the disparities, and then the mark value of each pair virtual views is determined by the percentage of difference area. The experimental results have shown the proposed method can detect copyright information in the virtual views without the depth map, and it has a good robustness against signal distortions.

Acknowledgments

This work is supported by the Shenzhen Municipal Development and Reform Commission (Disciplinary Development Program for Data Science and Intelligent Computing), and the Shenzhen Engineering Laboratory of Broadband Wireless Network Security.

References

[1] Asikuzzaman M, Alam M J, Lambert A J, et al., “A blind watermarking scheme for depth-image-based rendered 3D video using the dual-tree complex wavelet transform,” in *Proc. of IEEE International Conference on Image Processing.* pp. 5497-5501, 2014. Article (CrossRef Link)

[2] Lin Y H, Wu J L. “A digital blind watermarking for depth-image-based rendering 3D images,” *IEEE Transactions on Broadcasting*, 57(2): 602-611, 2011. Article (CrossRef Link)

[3] Kim H D, Lee J W, Oh T W, et al., “Robust DT-CWT watermarking for DIBR 3D images,” *IEEE Transactions on Broadcasting*, 58(4): 533-543, 2012. Article (CrossRef Link)

[4] Lee M J, Lee J W, Lee H K., “Perceptual watermarking for 3D stereoscopic video using depth information,” in *Proc. of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 81-84, 2011. Article (CrossRef Link)

[5] Lin Y H, Wu J L. “Unseen visible watermarking for color plus depth map 3D images,” in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp, 1801-1804, 2012. Article (CrossRef Link)

[6] Pei S C, Wang Y Y, “Auxiliary Metadata Delivery in View Synthesis Using Depth No Synthesis Error Model,” *IEEE Transactions on Multimedia*, 17(1): 128-133, 2015. Article (CrossRef Link)

[7] Guan Y, Zhu Y, Liu X, et al., “A digital blind watermarking scheme based on quantization index modulation in depth map for 3D video,” in *Proc. of International Conference on Control Automation Robotics & Vision*, pp.346-351, 2014. Article (CrossRef Link)

[8] Liang Zhang, and Wa james Tam, “Stereoscopic image generation based on depth images for 3D TV,” *IEEE Transaction on Broadcasting*, 51.2, pp.191-199, 2005. Article (CrossRef Link)

[9] Karsten muller, Philipp merkle, and thomas wiegand, "3D video representation using depth maps," in *Proc. of the IEEE*, 99(4), pp.643-656, 2011. Article (CrossRef Link)

[10] I. Ahn, and C. Kim, "A novel depth-based virtual view synthesis method for free viewpoint video," *IEEE Transaction on Broadcasting*, vol. 59, no. 4, pp. 614-626, Dec. 2013. Article (CrossRef Link)

[11] Hirschmuller H., "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), pp.328-341, 2008. Article (CrossRef Link)

[12] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics*, volume 23, pp. 600–608. 2004. Article (CrossRef Link)

[13] H. Schwarz, D. Marpe, and T. Wiegand, "Description of exploration experiments in 3d video coding," ISO/IEC JTC1/SC29/WG11 MPEG2010 N, 11274, 2010.

[14] Chen W Y, Chang Y L, Lin S F, et al., "Efficient depth image based rendering with edge dependent depth filter and interpolation," in *Proc. of IEEE International Conference on Multimedia and Expo*. pp. 1314-1317, 2005. Article (CrossRef Link)

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. Article (CrossRef Link)

Zhaotian Li is currently studying for a Master degree of computer application technology at the Lab of Communication and Information Security, Shenzhen Graduate School, Peking University. Her research interest is the computer vision and multimedia technology.

Yuesheng Zhu received his B.Eng. degree in radio engineering, M. Eng. degree in circuits and systems and Ph.D. degree in electronics engineering in 1982, 1989 and 1996, respectively. He is currently working as a professor at the Lab of Communication and Information Security, Shenzhen Graduate School, Peking University. He is a senior member of IEEE, fellow of China Institute of Electronics, and senior member of China Institute of Communications. His interests include digital signal processing, multimedia technology, communication and information security.

Guibo Luo received his MSc degree in 2013 from Peking University. He is currently working as an engineer at the Lab of Communication and Information Security, Shenzhen Graduate School, Peking University. His research interests are computer vision, machine learning, and multimedia technology.

Biao Guo is currently studying for a PhD degree at the Lab of Communication and Information Security, Shenzhen Graduate School, Peking University. His research interests are computer vision, machine learning, and bioimage processing.