Discriminative and geometrically robust zero-watermarking scheme for protecting DIBR 3D videos

DISCRIMINATIVE AND GEOMETRICALLY ROBUST ZERO-WATERMARKING SCHEME FOR PROTECTING DIBR 3D VIDEOS

Xiyao Liu§, Yayun Zhang§, Sibo Du§, Jian Zhang§*, Ming Jiang† and Hui Fang‡*

§School of Computer Science and Engineering, Central South University, Changsha, China
†School of Computer Science and Information Security,
Guilin University of Electronic Technology, Guilin, China.
‡Department of Computer Science, Loughborough University, Loughborough, U.K.

ABSTRACT

Copyright protection of depth image-based rendering (DIBR) 3D videos is crucial due to the popularity of these videos. Despite the success of recent watermarking schemes, it is still challenging to ensure the robustness against strong geometric attacks when both lossless quality and distinguishability of protected videos are required. In this paper, we propose a novel zero-watermarking scheme to improve the performance under strong geometric attacks when satisfying the other two requirements. In our scheme, CT-SVD-based features are extracted to ensure both distinguishability and robustness against signal processing and DIBR conversion attacks, while a SIFT-based rectification mechanism is designed to resist geometric attacks. Further, an attention-based fusion strategy is proposed to complement the robustness of rectified and unrectified CT-SVD features. Experimental results demonstrate that our scheme outperforms the existing zero-watermarking schemes in terms of distinguishability and robustness against strong geometric attacks such as rotation, cyclic translation and shearing.

Index Terms— Zero-watermarking, DIBR 3D videos, CT-SVD, SIFT-based rectification, attention-based fusion

1. INTRODUCTION

Three-dimensional (3D) videos provide better immersive experiences to viewers than traditional 2D videos, thus they have been becoming more and more popular in entertainments [1, 2, 3]. One typical format for storage and online distribution of 3D videos is the stereoscopic format, which contains two different 2D views captured by two parallel cameras for each frame. Another common used format is depth image-based rendering (DIBR) format which consists of a 2D view with its depth map for each frame. Compared with the stereoscopic format, the DIBR format saves storage and transmission bandwidth by efficiently compressing the depth map and thus is more widely adopted [2, 3].

Copyright protection of DIBR 3D videos is important. Compared with the protection of traditional 2D videos, it requires a unique characteristic to resist DIBR conversions. Simply put, watermarks need to be extracted from both the DIBR format and its converted stereoscopic format because the attackers can easily convert a DIBR format into a stereoscopic format for illegal distribution. Recently, many digital watermarking schemes are exclusively designed for DIBR 3D videos, which can be categorized into three main classes based on their embedding strategies: (1) 2D-frame-based watermarking schemes exploit the DIBR invariant characteristics of 2D frames for watermark embedding [4, 5, 6, 7, 8]; (2) depth-map-based watermarking schemes embed watermark into depth maps based on unseen visible or reversible watermarking [9, 10, 11]; and (3) zero-watermarking schemes utilize the relationship between 3D video features and the watermark information for the copyright identification without direct watermark embedding [12, 13, 14, 15].

Despite the success of these above mentioned watermarking schemes, it is still challenging to ensure the robustness, especially against geometric attacks, when both lossless quality and distinguishability of protected videos are required. The 2D-frame-based watermarking achieves sufficient robustness and distinguishability but causes irreversible distortion on image quality. The depth-map-based watermarking keeps the synthesized 3D videos distortion-free but is not sufficiently robust. Although the lossless zero-watermarking outperforms the other two categories of schemes in terms of both video quality and watermark robustness, none of the state-of-the-art zero-watermarking schemes ensures sufficient robustness against strong geometric attacks, such as rotation, cyclic-translation and shearing while achieving the distinguishability at the same time.

To address this challenge, in this paper, we propose an innovative zero-watermarking scheme based on contourlet transform (CT)-Singular Value Decomposition (SVD) fea-

tures and scale-invariant feature transform (SIFT)-based rectification for protecting DIBR 3D videos. In our proposed scheme, CT-SVD are performed on 2D frames of DIBR 3D videos to extract discriminative and robust features and a SIFT-based rectification is used to guarantee the watermarking robustness against strong geometric attacks. In addition, an attention-based fusion is designed to exploit the complementary robustness of both the rectified and the unrectified CT-SVD features, which further improves the performance of copyright identification.

Our contributions are highlighted as below:

- A novel zero-watermarking scheme is proposed to offer a lossless copyright protection of DIBR 3D videos.
- CT-SVD features are designed to ensure not only the distinguishability but also the robustness against signal processing and DIBR 3D conversion attacks.
- A SIFT-based rectification mechanism is established to resist strong geometric attacks, such as large-degree rotation, large-scale cyclic-translation and shearing.
- An attention-based fusion is designed to offer an optimal copyright protecting solution by exploiting the complementary robustness of rectified and unrectified CT-SVD features.
- Comprehensive experimental results demonstrate the superiority of the proposed scheme against the state-of-the-art zero-watermarking schemes.

2. PROPOSED SCHEME

Our proposed zero-watermarking scheme includes two phases, which are a copyright registration phase and a copyright identification phase, as shown in Fig. 1.

2.1. Copyright registration phase

In this phase, 2D frame features of DIBR 3D videos are extracted and their ownership shares, representing the relationship between these features and corresponding watermarks, are created and stored in a certificate authority (CA) database for copyright identification.

2.1.1. Extraction of 2D frame feature

In our scheme, robust and discriminative features of 2D frames are extracted based on CT-SVD. The detailed steps are as follows. Firstly, 2D frames are normalized based on spatio-temporal smoothing and resampling. Secondly, the normalized frames are averaged in temporal domain to construct the temporally informative representative images (TIRIs). Thirdly, three level contourlet transform (CT) [16] is applied on non-overlapping blocks of TIRIs and the 6th and 7th directional subbands of the 2nd level CT domain are selected. There are two reasons for the selection of these two subbands: (1) the coefficients in the 2nd level CT domain are

selected to ensure both the distinguishability and the robustness against signal processing attacks, and (2) the 6th and 7th directional subbands are chosen because they mainly contain horizontal edges and contours, as shown in Fig. 2, which are more robust against DIBR conversion. Fourthly, SVD transform is performed on the selected subbands and the first singular values in each diagonal matrices are selected. Finally, the feature of a DIBR 3D video is generated by binarizing these singular values of different image blocks of TIRIs based on their mean value.

2.1.2. Generation of ownership share

After the feature extraction, an ownership share is generated by XORing the extracted feature and the binary watermark. The generated ownership share is then stored into the CA database for copyright identification.

2.2. Copyright identification phase

In this phase, two rectified features and one unrectified feature are extracted from the 2D frames of a queried DIBR video. XOR operations are performed between these features and the corresponding ownership share to recover three watermarks. Finally, an attention-based fusion is utilized to exploit the complementary robustness of the rectified and unrectified features for copyright identification. The procedure details are listed as below.

2.2.1. Extraction of rectified feature

To resist geometric attacks, a SIFT-based rectification mechanism, including rotation, translation, and shearing rectifications, is designed. Firstly, we normalize the 2D frames of the queried DIBR video and those of the DIBR videos stored in database. Then, we extract the SIFT points in the 1st normalized 2D frames of both the queried and stored DIBR videos. Next, we match the SIFT points of the queried DIBR video to those of the stored ones. After that, we perform geometric rectifications in two parallel channels, as shown in Fig. 1(b). The former channel is a sequential connection between a rotation rectification and a translation rectification, while the latter is a single shearing rectification. The three rectifications are described as below.

The rotation rectification rotates the normalized 2D frames according to the factor expressed in eq.(1).

$$\Delta_r = \frac{1}{N(N-1)} \sum_{i=1}^{N(N-1)} \arccos \frac{\vec{V}_s(i) \times \vec{V}_q(i)}{\left|\vec{V}_s(i)\right| \cdot \left|\vec{V}_q(i)\right|} \quad (1)$$

where Δ_r is the factor of rotation rectification, $\vec{V}_q(i)$ and $\vec{V}_s(i)$ are the vectors obtained by the i-th pair of matched SIFT points in 2D frames of the queried and stored videos, respectively, and N is the total number of the matched SIFT points.
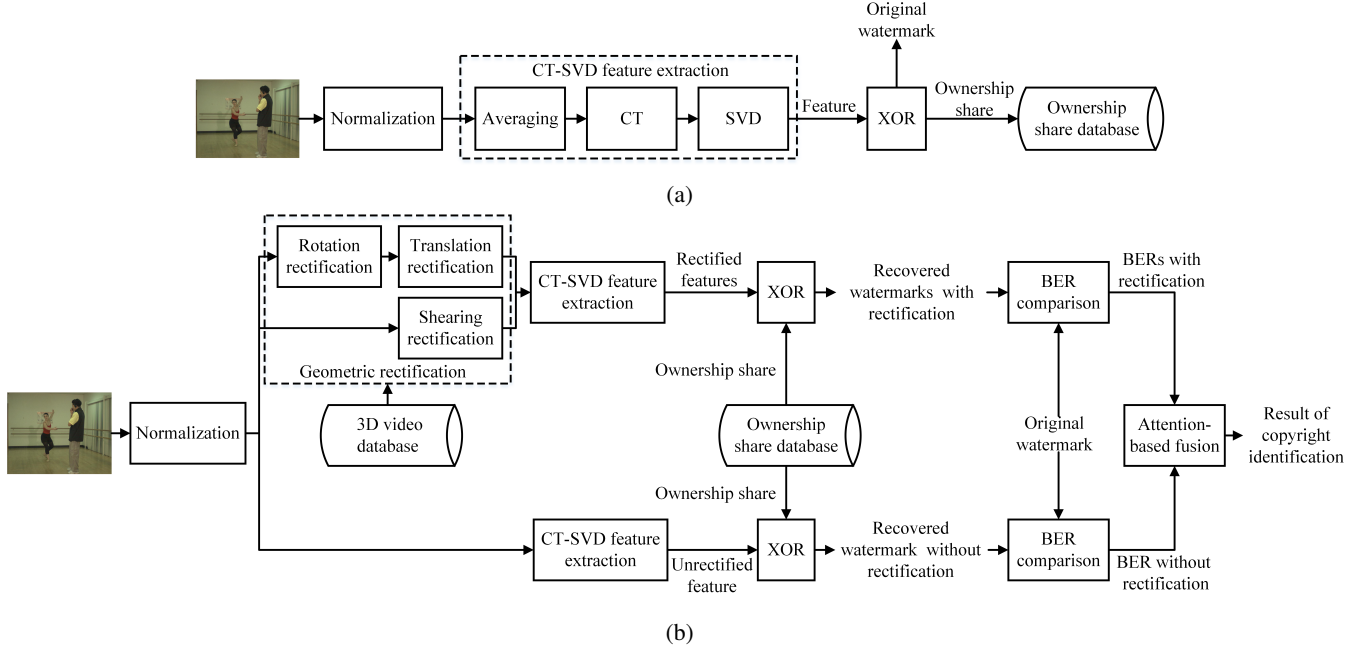
Fig. 1: Framework of our proposed zero-watermarking: (a) copyright registration phase, (b) copyright identification phase
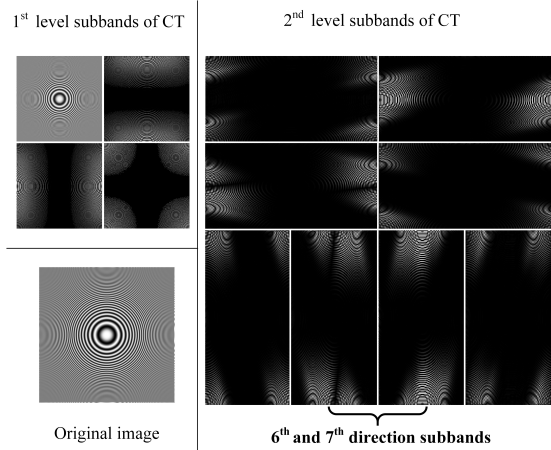

Fig. 2: Illustration of contourlet transform subbands

The translation rectification moves the pixels in normalized 2D frames according to the factors expressed in eq.(2)

$$\Delta x = \begin{cases} \frac{\sum_{k=1}^{N}\left(x_s(k)-x_q(k)\right)}{N}, & \text{if } x_s(k) > x_q(k) \\ \frac{\sum_{k=1}^{N}\left(x_s(k)-x_q(k)\right)}{N} + W, & \text{otherwise} \end{cases}$$

$$\Delta y = \begin{cases} \frac{\sum_{k=1}^{N}\left(y_s(k)-y_q(k)\right)}{N}, & \text{if } y_s(k) > y_q(k) \\ \frac{\sum_{k=1}^{N}\left(y_s(k)-y_q(k)\right)}{N} + H, & \text{otherwise} \end{cases} \quad (2)$$

where Δ_x, Δ_y are the translation rectification factors, $(x_q(k), y_q(k))$ are the coordinates of k-th matched SIFT point in the 2D frame of queried DIBR video, $(x_s(k), y_s(k))$ are those of stored videos, and W, H are the width and the height of normalized 2D frames, respectively.

Shearing are represented as shown in eq.(3).

$$\begin{cases} x_q(k) = x_s(k) + \Delta_a \cdot y_s(k) \\ y_q(k) = \Delta_b \cdot x_s(k) + y_s(k) \end{cases} \quad (3)$$

According to eq.(4), we calculate the rectification factors (Δ_a and Δ_b) of shearing transformation. Then, we can rectify the normalized 2D frames by substituting our obtained rectification factors into eq.(3).

$$\Delta_a = \frac{\sum_{k=1}^{N}(x_s(k) - x_q(k))}{\sum_{k=1}^{N} y_q(k)}$$
$$\Delta_b = \frac{\sum_{k=1}^{N}(y_s(k) - y_q(k))}{\sum_{k=1}^{N} x_q(k)} \quad (4)$$

In specific, translation rectification is performed after the rotation rectification because cyclic-translation attacks will not affect the results of rotation rectification, but conversely, rotation attacks will lead to an incorrect translation rectification. Moreover, shearing rectification is separated from the other two rectifications to avoid incorrect rectification caused by the interactions of shearing with the other two attacks.

After geometric rectifications, a rotation-translation rectified feature, namely f_{r1}, and a shearing rectified feature, namely f_{r2}, are extracted following the same extraction steps in our copyright registration phase.

Table 1: InterBERs of different schemes.

InterBER	[13]	[14]	[15]	Proposed Scheme			
				f_{r1}	f_{r2}	f_u	Fused
Minimum	0.156	0.073	0.124	0.225	0.227	0.225	0.225
Average	0.422	0.187	0.488	0.495	0.495	0.494	0.488

Table 2: Attack types and their parameters

Attack	Parameter
Salt & pepper noise (SN)	noise density = 0.05
Gaussian noise (GN)	mean = 0, variance = 0.05
Gaussian blurring (GB)	variance = 1, window = 5×5
Average filtering (AF)	window = 5×5
Contrast change (CC)	{ +20%, -20%}
DIBR baseline distance	left view {5%, 7%},
adjustment (BA)	right view { 5%, 7%}
Cropping (CR)	5%
Scaling (SC)	{50%, 200%}
Rotation (RT)	{45°, 90°}
Cyclic-translation (CTR)	right 30%, down 30%
Shearing (SH)	(Δ_a, Δ_b)={(0, 0.5), (0.5, 0), (0.2, 0.2)}

2.2.2. Extraction of unrectified feature

The CT-SVD feature of unrectified 2D frames, namely f_u, is also extracted following the same steps in copyright registration phase. In this manner, the watermarking robustness can be further improved because some of the attacks, such as noise addition and DIBR conversion, may affect the results of SIFT point matching or rectification factor calculation.

2.2.3. Attention-based fusion

In our scheme, an attention-based fusion is designed to further enhance the performance of copyright identification by exploiting the complement robustness of all the rectified and unrectified features. Three watermarks are recovered by XORing these features with the corresponding ownership share of each stored video. Three bit error rates (BERs) between the original watermark and them are calculated. These BERs are fused by eq.(5) and eq.(6) to simultaneously satisfy the heterogeneity and monotonicity expressed in eq.(7). If any fused BER is smaller than a heuristic threshold, the queried DIBR 3D video is treated as an illegal copy of the video corresponding to this fused BER.

$$
\begin{aligned}
F(BER_1, BER_2) &= 1 - \frac{1}{2}\left(x_1 + \frac{1}{1+\lambda}x_2\right) \\
x_1 &= (1 - BER_1) + (1 - BER_2) \\
x_2 &= |BER_1 - BER_2|
\end{aligned} \tag{5}
$$

$$
BER_{fused} = F\left(F\left(BER_{f_{r_1}}, BER_{f_{r_2}}\right), BER_{f_u}\right) \tag{6}
$$

where F is the attention-based fusion function, BER_1 and BER_2 are the BERs to be fused, λ is a constant and is set to 0.01 empirically in our study. $BER_{f_{r_1}}$ and $BER_{f_{r_2}}$ are obtained from the rectified features, the BER_{f_u} is obtained from

the unrectified feature, and BER_{fused} is the fusion result.

$$
\begin{aligned}
F(BER_1, BER_2) &> F(BER_1 - \xi, BER_2 + \xi) \\
F(BER_1, BER_2) &< F(BER_1, BER_2 + \xi)
\end{aligned} \tag{7}
$$

where $0 < \xi \le BER_1 \le BER_2$

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental settings

The testing database of our study contains 50 DIBR 3D video clips with different frame numbers and frame sizes, including videos from datasets of the MEPG 3DAV group [17], Interactive Visual Media Group of Microsoft Research [18] and Shenzhen Institutes of Advanced Technology (SIAT) [19] as well as 2D frames selected from existing movies, with corresponding depth maps calibrated based on the scheme in [20].

When implementing our proposed scheme, all the frames are normalized to $320 \times 320 \times 100$ (height × width × frame-number). And a total of 25 TIRIs are generated and each of them are divided into 8×8 non-overlapping blocks for CT-SVD feature extraction. As a result, the dimension of our extracted CT-SVD feature is 1600, which is the same as the size of our utilized watermark.

Three state-of-the-art zero watermarking schemes [13, 14, 15], are compared with our proposed scheme in terms of distinguishability, robustness, and over-all performances in sections 3.2, 3.3 and 3.4 respectively.

3.2. Comparison of distinguishability

In our study, interBER is used to compare the distinguishability of different schemes. The interBER is defined as the BER between the genuine watermark and the fake watermark. The genuine watermark is generated by XORing the ownership share and the master share of the same video while the fake watermark is generated by XORing the shares of different videos. The watermarking distinguishability is higher if the interBER value of a scheme is larger. The comparison results are listed in Table 1. Here, f_{r1}, f_{r2} and f_u indicates using the single CT-SVD feature with rotation-translation rectification, shearing rectification, or without any rectification, respectively.

As shown in Table 1, the average interBER of our proposed scheme by fusing the three features is 0.488 and much larger than those of [13, 14] and comparative with that of [15]. Moreover, our minimum interBER is 0.225 and much larger than those of [13, 14, 15], which are 0.156, 0.073 and 0.124 respectively. These results demonstrate that our proposed scheme superiors to the other three schemes in terms of watermarking distinguishability. The reason for these results is that we select the 6th and 7th of subbands of 2nd level CT domain, which contain discriminative horizontal edges and

Table 3: Mean intraBERs and P_{fn} values ($P_{fp} = 0$) of different schemes.

| Attack Types | [13] | | [14] | | [15] | | Proposed Scheme | | | | | | | |
| | | | | | | | f_{r1} | | f_{r2} | | f_u | | Fused | |
	IntraBER	P_{fn}	IntraBER	P_{fn}	IntraBER	P_{fn}	IntraBER	P_{fn}	IntraBER	P_{fn}	IntraBER	P_{fn}	IntraBER	P_{fn}
SN 0.05	0.073	0.100	0.039	0.080	0.054	0.100	0.064	0.040	0.258	0.740	0.022	0.000	0.023	0.000
GN 0.05	0.143	0.300	0.067	0.380	0.127	0.320	0.139	0.160	0.216	0.540	0.039	0.000	0.039	0.000
GB 5×5	0.020	0.000	0.014	0.000	0.074	0.160	0.056	0.000	0.109	0.020	0.055	0.000	0.053	0.000
AF 5×5	0.045	0.000	0.025	0.020	0.059	0.120	0.096	0.020	0.162	0.100	0.089	0.020	0.087	0.020
CC +20%	0.020	0.000	0.015	0.000	0.019	0.000	0.028	0.000	0.051	0.000	0.028	0.000	0.028	0.000
CC -20%	0.003	0.000	0.003	0.000	0.001	0.000	0.005	0.000	0.038	0.000	0.005	0.000	0.005	0.000
BA left,5%	0.042	0.000	0.121	0.700	0.048	0.040	0.067	0.000	0.077	0.000	0.072	0.000	0.046	0.000
BA left,7%	0.052	0.020	0.130	0.760	0.063	0.080	0.093	0.020	0.095	0.000	0.094	0.000	0.066	0.000
BA right,5%	0.046	0.020	0.113	0.700	0.048	0.060	0.080	0.000	0.074	0.000	0.070	0.000	0.047	0.000
BA right,7%	0.058	0.000	0.126	0.760	0.065	0.140	0.096	0.000	0.089	0.000	0.092	0.000	0.063	0.000
CR 5%	0.049	0.000	0.137	0.940	0.021	0.040	0.001	0.000	0.266	0.720	0.001	0.000	0.001	0.000
SC 50%	0.013	0.000	0.007	0.000	0.034	0.040	0.044	0.000	0.050	0.000	0.044	0.000	0.036	0.000
SC 200%	0.003	0.000	0.002	0.000	0.012	0.000	0.009	0.000	0.037	0.000	0.009	0.000	0.009	0.000
RT 45°	0.403	1.000	0.011	0.000	0.150	0.560	0.074	0.000	0.465	1.000	0.464	1.000	0.078	0.000
RT 90°	0.388	1.000	0.000	0.000	0.187	0.700	0.023	0.000	0.488	1.000	0.488	1.000	0.027	0.000
CTR right,30%	0.349	0.920	0.175	1.000	0.369	0.980	0.004	0.000	0.388	0.840	0.439	1.000	0.008	0.000
CTR down,30%	0.451	1.000	0.179	1.000	0.361	0.980	0.008	0.000	0.429	0.980	0.495	1.000	0.012	0.000
SH (0,0.5)	0.216	0.740	0.193	1.000	0.160	0.720	0.378	1.000	0.037	0.000	0.291	0.840	0.040	0.000
SH (0.5,0)	0.319	0.960	0.185	1.000	0.225	0.840	0.466	1.000	0.037	0.000	0.426	1.000	0.041	0.000
SH (0.2,0.2)	0.242	0.760	0.202	1.000	0.199	0.800	0.417	1.000	0.065	0.000	0.393	0.980	0.069	0.000
Average	0.147	0.341	0.087	0.467	0.114	0.334	0.107	0.162	0.172	0.297	0.181	0.342	0.039	0.001

contours, for the feature extraction. Although watermarking distinguishability by fusing the three features is naturally and slightly worse than those by using any single feature, the robustness of the fusing one are much stronger as shown in section 3.3, leading to its better overall watermarking performance as shown in section 3.4.

3.3. Comparison of robustness

Then, we perform different attacks on all the 50 videos and compare the robustness of different schemes in terms of mean value of intraBER, which is defined as the BER between the watermarks recovered from original and attacked videos. The attacks with the detailed parameters are shown in Table 2 and the comparison results are listed in Table 3.

As shown in Table 3, our proposed scheme by fusing the three features achieved an impressive performance in terms of the mean intraBER. In specific, the average value of our mean intraBER is 0.039. This value is much lower than those by utilizing any single feature. These results indicate that the attention-based fusion further enhances the watermarking robustness. The reason is because that the use of rectified features ensures the strong robustness against rotation, cyclic-translation and shearing attacks due to the invariance of SIFT, while the use of unrectified feature enhances the robustness under the noise addition and DIBR conversion attacks. Moreover, the average value of our mean intraBER is much lower compared with those of the other schemes, which are 0.147, 0.087 and 0.114. Specifically, for cyclic-translation and shearing attacks, our largest mean intraBER is 0.069 and it is much better than those of the other three schemes, of which the smallest value is 0.160. For rotation attacks, our largest mean intraBER is 0.078 while the mean intraBERs

of [13, 15] are close to 0.4 and 0.2. For DIBR attacks, our largest mean intraBER is 0.066 while the smallest value of [14] is 0.113. These results demonstrate that our proposed scheme is more robust than the other three schemes. The reason for these results is threefold: (1) Our well-designed CT-SVD feature ensures the robustness against signal processing and DIBR conversion attacks. (2) The SIFT-based rectification resists strong geometric attacks. (3) The attention-based fusion further exploits the complementary robustness of the rectified and unrectified features.

3.4. Comparison of over-all performance

Finally, the overall performance of our proposed scheme is compared with other scheme in terms of false negative rates P_{fn} with a fixed false positive rate P_{fp} and smaller P_{fn} values represent better overall performance. In our experiments, the P_{fp} values of different schemes are set to 0 by defining their identification thresholds as the same with their minimum interBERs, respectively. The results are listed in Table 3.

As shown in Table 3, most of the P_{fn} values of our proposed scheme by fusing the rectified and unrectified features equal to 0 with an insignificant average value as 0.001. These P_{fn} values are much lower than those by using any single feature of our proposed scheme, which demonstrate the effectiveness of our designed attention-based fusion. Furthermore, our P_{fn} values are much lower than those of the other benchmark schemes [13, 14, 15] with their average values as high as 0.341, 0.467 and 0.334 respectively. Especially, our P_{fn} values are still 0 under cyclic translation and shearing attacks while those of the three benchmark schemes are close to 1. These results indicates that our proposed scheme outperforms the other three schemes in terms of the overall performance.

4. CONCLUSION

In this paper, a novel zero-watermarking scheme is proposed for the copyright identification of the DIBR 3D videos. The advantages of our proposed scheme include: (1) By using our well designed CT-SVD feature, the watermarking robustness against signal attacks, DIBR attacks, and the watermarking distinguishability are ensured simultaneously. (2) By establishing the SIFT-based rectification, the strong geometric attacks are resisted. (3) By designing the attention-based fusion, the complementary robustness of the rectified and unrectified CT-SVD features are exploited, which further improves the performances for copyright identification. The experimental results demonstrate that our scheme superiors to the existing zero watermarking schemes for 3D videos in terms of distinguishability and robustness against strong geometric attacks. Our future work is to explore this technique to protect medical volume images and other types of multimedia data.

5. REFERENCES

[1] Aljoscha Smolic, Karsten Mueller, Philipp Merkle, Christoph Fehn, Peter Kauff, Peter Eisert, and Thomas Wiegand, "3d video and free viewpoint video-technologies, applications and mpeg standards," in *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 2006, pp. 2161–2164.

[2] Peter Kauff, Nicole Atzpadin, Christoph Fehn, Marcus Müller, Oliver Schreer, Aljoscha Smolic, and Ralf Tanger, "Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217–234, 2007.

[3] André Redert, Marc Op De Beeck, Christoph Fehn, Wijnand IJsselsteijn, Marc Pollefeys, Luc Van Gool, Eyal Ofek, Ian Sexton, and Philip Surman, "Advanced three-dimensional television system technologies," in *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*. IEEE, 2002, pp. 313–319.

[4] Hee-Dong Kim, Ji-Won Lee, Tae-Woo Oh, and Heung-Kyu Lee, "Robust dt-cwt watermarking for dibr 3d images," *IEEE Transactions on Broadcasting*, vol. 58, no. 4, pp. 533–543, 2012.

[5] Md Asikuzzaman, Md Jahangir Alam, Andrew J Lambert, and Mark R Pickering, "Robust dt cwt-based dibr 3d video watermarking using chrominance embedding," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1733–1748, 2016.

[6] Yong-Seok Lee, Young-Ho Seo, and Dong-Wook Kim, "Digital blind watermarking based on depth variation prediction map and dwt for dibr free-viewpoint image," *Signal Processing: Image Communication*, vol. 70, pp. 104–113, 2019.

[7] Chen Cui, Shen Wang, and Xiamu Niu, "A novel watermarking for dibr 3d images with geometric rectification based on feature points," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 649–677, 2017.

[8] Seung-Hun Nam, Wook-Hyoung Kim, Seung-Min Mun, Jong-Uk Hou, Sunghee Choi, and Heung-Kyu Lee, "A sift features based blind watermarking for dibr 3d images," *Multimedia Tools and Applications*, vol. 77, no. 7, pp. 7811–7850, 2018.

[9] Yu-Hsun Lin and Ja-Ling Wu, "Unseen visible watermarking for color plus depth map 3d images," in *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2012, pp. 1801–1804.

[10] Soo-Chang Pei and Yu-Ying Wang, "Auxiliary metadata delivery in view synthesis using depth no synthesis error model," *IEEE Transactions on Multimedia*, vol. 17, no. 1, pp. 128–133, 2014.

[11] Xianglian Shi, Bo Ou, and Zheng Qin, "Tailoring reversible data hiding for 3d synthetic images," *Signal Processing: Image Communication*, vol. 64, pp. 46–58, 2018.

[12] Xiyao Liu, Rongchang Zhao, Fangfang Li, Shenghui Liao, Yipeng Ding, and Beiji Zou, "Novel robust zero-watermarking scheme for digital rights management of 3d videos," *Signal processing: Image communication*, vol. 54, pp. 140–151, 2017.

[13] Chen Cui, Haokun Mao, Xiamu Niu, Lixian Zhang, Tasawar Hayat, and Ahmed Alsaedi, "A novel hashing scheme for depth-image-based-rendering 3d images," *Neurocomputing*, vol. 191, pp. 1–11, 2016.

[14] Chunpeng Wang, Xingyuan Wang, Zhiqiu Xia, and Chuan Zhang, "Ternary radial harmonic fourier moments based robust stereo image zero-watermarking algorithm," *Information Sciences*, vol. 470, pp. 109–120, 2019.

[15] Xiyao Liu, Yifan Wang, Ziqiang Sun, Lei Wang, Rongchang Zhao, Yuesheng Zhu, Beiji Zou, Yuqian Zhao, and Hui Fang, "Robust and discriminative zero-watermark scheme based on invariant features and similarity-based retrieval to protect large-scale dibr 3d videos," *Information Sciences*, vol. 542, pp. 263–285, 2021.

[16] Minh N Do and Martin Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[17] Christoph Fehn, Klaas Schüür, Ingo Feldmann, Peter Kauff, and Aljoscha Smolic, "Distribution of attest test sequences for ee4 in mpeg 3dav," in *MPEG Meeting-ISO/IEC JTC1/SC29/WG11, MPEG02/M9219*, 2002.

[18] C Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski, "High-quality video view interpolation using a layered representation," in *ACM transactions on graphics (TOG)*. ACM, 2004, vol. 23, pp. 600–608.

[19] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C.J. Kuo, and Q. Peng, "Subjective and objective video quality assessment of 3d synthesized views with texture/depth compression distortion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4847–4861, Dec 2015.

[20] Richard Rzeszutek, Raymond Phan, and Dimitrios Androutsos, "Semi-automatic synthetic depth map generation for video using random walks," in *2011 IEEE International Conference on Multimedia and Expo*. IEEE, 2011, pp. 1–6.