Despite the success of watermarking technique for protecting depth image-based rendering (DIBR) 3-D videos, existing methods still can hardly ensure the robustness against geometric attacks, lossless video quality, and distinguishability between different videos simultaneously. In this article, we propose a novel zero-watermarking scheme to address this challenge. Specifically, we design CT-SVD features to ensure both distinguishability and robustness against signal processing and DIBR conversion attacks. In addition, a logistic–logistic chaotic system is utilized to encrypt features for the enhanced security. Moreover, a rectification mechanism based on salient map detection and SIFT matching is designed to resist geometric attacks. Finally, we establish an attention-based fusion mechanism to explore the complementary robustness of rectified and unrectified features. Experimental results demonstrate that our proposed method outperforms the existing schemes in terms of losslessness, distinguishability, and robustness against geometric attacks.

Legal copyright protection of 3-D depth image-based rendering (DIBR) video content has become important owing to the emerging trend of using advanced immersive technologies in entertainments, business communications, and commercial campaigns. 3-D videos are typically stored in either of the two formats for their online distribution: 1) stereo-scopic and 2) DIBR videos. Compared with the stereo-scopic format, DIBR is the more common format due to its lower storage and transmission bandwidth requirements since depth maps can be compressed more effectively. Unlike protection of 2-D videos, copyright protection of DIBR 3-D videos poses a unique challenge, that is, the robustness against DIBR conversion attacks. Because illegal copies are often distributed by converting the original DIBR video into a stereoscopic format, watermarks need to be extracted not only from the DIBR format but also from its converted format. Traditional copyright protection techniques for 2-D videos can hardly satisfy this requirement since the pixels in the video frames are shifted horizontally, and thus, the synchronization of watermark recovery is lost in the DIBR conversion process.

Digital watermarking is a key technique to protect copyright of the video content, which has wide applications in video coding standard. DIBR watermarking methods can be categorized into three classes based on their embedding strategies: 1) 2-D frame-based watermarking (2-D-W), 2) depth map-based watermarking (DM-W), and 3) zero-watermarking (ZR-W). 2-D-W directly embeds watermarks into 2-D frames of the 3-D videos. This approach is intuitive but the embedding process causes irreversible distortion in terms of video quality. DM-W addresses this problem; however, it is not robust against various attacks, such as blurring, filtering, and noise addition. In contrast to the other approaches, ZR-W utilizes the relationship between features extracted from the 3-D videos and watermarks to losslessly and robustly protect the 3-D video copyright. Although ZR-W schemes outperform 2-D-W and DM-W algorithms in terms of video quality and watermark robustness, it is still
a challenge to ensure the robustness against geometric attacks while achieving the video distinguishability simultaneously.

To address this challenge, we propose a novel ZR-W scheme based on horizontal shift-invariant features and a geometric rectification mechanism to protect DIBR 3-D videos in this article. In this method, contourlet transform and singular value decomposition (CT-SVD) are combined to extract horizontal shift-invariant and discriminative features. In addition, a logistic–logistic chaotic system (LLCS) is utilized to encrypt these features to improve the watermarking security. Furthermore, we design a geometric rectification mechanism by using salient map detection and scale-invariant feature transform (SIFT)-based matching. In this manner, the watermarking robustness against geometric attacks is guaranteed. Finally, an attention-based fusion is proposed to make full use of the complementary robustness of both the rectified and the unrectified CT-SVD features, thus further improving the reliability and accuracy of copyright identification.

This article is an extended version of our previous work, and the following improvements have been made:

1) enhancing the geometric rectification mechanism by combining an RGB-D saliency detection model, to improve rectification efficiency as well as the video distinguishability;
2) exploiting an LLCS to encrypt the designed features, thus improving the watermarking security;
3) comparing to more SOTA zero-watermark methods under more types of attacks with more testing data to demonstrate the superiority of our proposed method.

**PROPOSED SCHEME**

Our proposed method includes two phases: 1) a registration phase and 2) an identification phase. In the following, we describe each phase in detail.

**Registration Phase**

In the registration phase, features of 2-D frames are extracted and a certificate authority (CA) database is set up to store the ownership shares of 2-D frames for copyright identification, as illustrated in Figure 1.

In our scheme, robust and discriminative features are extracted based on CT-SVD. The detailed steps are as follows.

1) Normalize the 2-D frames to \(320 \times 320 \times 100\) based on spatiotemporal smoothing by using a Gaussian filter with a window size of 3 and a standard deviation of 1. In this manner, the watermarking robustness against noise addition attacks is enhanced.
2) Resample the smoothed image to yield the normalized frames. In detail, we use bilinear interpolation and downsampling (to a fixed frame number, 100) for the spatial and temporal resampling, respectively. In this manner, the watermarking robustness against scaling attacks is enhanced.
3) Divide normalized frames into \(m\) groups, and construct temporally informative representative images (TIRI) by averaging each group in the temporal domain. By exploring the temporal properties of 2-D frames sequence, the robustness against noise addition attacks is improved. In our work, \(m = 25\).
4) Partition each TIRI frame into \(N\) nonoverlapping blocks, and perform three-level CT on each block. In our work, \(N = 8\).
5) Select the sixth and seventh directional subbands of the second-level CT domain for feature extraction due to the following two reasons.
   1) The selection of coefficients in the second-level CT domain ensures both the distinguishability and the robustness against signal processing attacks.
   2) The selection of sixth and seventh directional subbands enhance the robustness against DIBR conversion because these two subbands mainly contain horizontal edges and contours, as shown in Figure 2.
6) Apply SVD transform on the selected subbands and use the first singular values in each diagonal matrices for feature extraction. In this manner, the robustness against signal processing attacks is enhanced by deploying the stability of the first singular values of SVD.
7) Binarize the first singular values of different image blocks of TIRIs based on their mean value.
8) Concatenate the binary bits which are generated from all the TIRIs as the final extracted feature vector $F$ of DIBR 3-D video with its dimension as $(100/25) \times 8 \times 8 = 1600$.

To generate ownership shares, which represent the mapping relationship between the video features and watermark information, we employ chaotic mapping based on LLCS\(^{14}\) as follows:

1) Use the logistic maps defined in (1) to generate a chaotic sequence $S = \{s(i + L), 1 \leq i \leq 1600\}$, as shown in the following:

\[
\begin{align*}
\zeta &= u \cdot s(n) \cdot (1 - s(n)) \quad (1) \\
s(n + 1) &= \zeta \cdot 2^n - \lfloor \zeta \cdot 2^n \rfloor \quad (2)
\end{align*}
\]

where $u$ and $v$ are the control parameters, $0 \leq u \leq 10$, $8 \leq v \leq 20$, $s(0)$ is the initial value of chaotic system, $n$ is the iteration number, and $s(n)$ is the output chaotic sequence. $u$, $v$, and $s(0)$ are used together as the secret key, and $\lfloor x \rfloor$ is a floor operation. These parameters are strictly followed by Pak and Huang\(^{14}\) to ensure the effectiveness of the chaotic system. Because the chaotic sequences generated after multiple iterations have better chaotic performance, $L$ is set to random integer value larger than 1000 in our study.

2) Binarize the chaotic sequence $S$ to obtain a binary chaotic sequence $BS = \{bs(i), 1 \leq i \leq 1600\}$ by its mean value

\[
bs(i) = \begin{cases} 
1, & \text{if } s(i + L) > T_S \\
0, & \text{otherwise}
\end{cases} \quad (3)
\]

where $T_S$ is the average value of $S$.

3) Encrypt our designed CT-SVD features $F = \{f(i), 1 \leq i \leq 1600\}$ by applying exclusive-or (XOR) operation with the binary chaotic sequence $BS$ as

\[
cf(i) = f(i) \oplus bs(i) \quad (4)
\]

where $\oplus$ denotes the XOR operation and $CF = \{cf(i), 1 \leq i \leq 1600\}$ is the encrypted feature. In this manner, the watermarking security is enhanced without affecting the robustness and distinguishability of our extracted features.

4) Generate ownership shares $O$ by applying exclusive-or (XOR) operation between encrypted feature $CF$ and watermark information $W$

\[
o(i) = w(i) \oplus cf(i) \quad (5)
\]

where $\oplus$ denotes the XOR operation, $o(i)$ and $w(i)$ are the $i$th bit of $O$ and $W$, respectively.

5) Store the $O$ with the corresponding secret key of LLCS into the CA database for watermark recovery.

\section*{Identification Phase}

In this phase, a geometric rectification mechanism based on salient map detection and SIFT feature matching is designed to resist geometric attacks. Two rectified features and one unrectified feature are extracted from a query DIBR video. Then, LLCS is performed to encrypt these features, and three watermarks are recovered by the XOR operations between these encrypted features and the corresponding ownership share. Finally, the copyright ownership is identified based on our designed attention-based fusion model to utilize the complementary robustness of the rectified and unrectified features. The detailed processes are shown in Figure 3.

Our designed geometric rectification comprises of two parallel channels, as shown in Figure 3: one channel is a connection between a rotation rectification and a translation rectification, while the other one is a shearing rectification. As the rotation attacks will lead to incorrect translation rectification results while the cyclic-translation attacks have no effect on rotation rectification, the rotation rectification should be performed first. In addition, shearing rectification should be separated from the other two rectifications because incorrect rectification will be caused by the interactions of shearing with rotation or cyclic-translation attacks. The processes of three rectifications are as follows:

\begin{itemize}
\item The normalized 2-D video frames are rectified from rotation attacks according to the factors
\end{itemize}
expressed as follows:

\[
\Delta_\theta = \frac{1}{N(N-1)} \sum_{i=1}^{N(N-1)} \arccos \left( \frac{\mathbf{v}_q(i) \cdot \mathbf{v}_s(i)}{|\mathbf{v}_q(i)| \cdot |\mathbf{v}_s(i)|} \right)
\]

(6)

where \( \Delta_\theta \) is the factor of rotation rectification, \( \mathbf{v}_q(i) \) and \( \mathbf{v}_s(i) \) are the vectors obtained by the \( i \)th pair of matched SIFT points in salient regions of the queried and stored videos, respectively, and \( N \) is the number of the matched SIFT points.

Then, the pixels of rotation rectified frames are moved based on translation rectification according to the factors expressed as follows:

\[
\Delta x = \begin{cases} 
\sum_{k=1}^{N} (x_q(k) - x_s(k)) / N, & \text{if } x_q(k) > x_s(k) \\
\sum_{k=1}^{N} (x_q(k) - x_s(k)) + W, & \text{otherwise}
\end{cases}
\]

(7)

\[
\Delta y = \begin{cases} 
\sum_{k=1}^{N} (y_q(k) - y_s(k)) / N, & \text{if } y_q(k) > y_s(k) \\
\sum_{k=1}^{N} (y_q(k) - y_s(k)) + H, & \text{otherwise}
\end{cases}
\]

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

Meanwhile, the shearing rectification is performed on normalized 2-D video frames. We define rectification factors \( \Delta a \) and \( \Delta b \) of shearing attacks as follows:

\[
\Delta a = \frac{\sum_{k=1}^{N} (x_q(k) - x_s(k))}{\sum_{k=1}^{N} y_q(k)}
\]

(8)

\[
\Delta b = \frac{\sum_{k=1}^{N} (y_q(k) - y_s(k))}{\sum_{k=1}^{N} x_q(k)}
\]

Then, the normalized 2-D frames is rectified by substituting these obtained rectification factors into the following equation:

\[
\begin{cases} 
x_q(k) = x_s(k) + \Delta a \cdot y_s(k) \\
y_q(k) = \Delta b \cdot x_s(k) + y_s(k)
\end{cases}
\]

(9)

Rectified feature extraction steps are shown as follows:

1. Normalize the 2-D frames of query and stored DIBR videos.
2. Define a salient mask for geometric rectification.
   Here, we propose to use a pretrained RGB-D salient map detection model, namely UC-Net, to enhance both the efficiency of geometric rectification and the distinguishability of the rectified features. In specific, we generate a single-channel saliency map by applying the pretrained UC-Net on the first frame of each DIBR video. This saliency map is then binarized to generate a mask so that the geometric rectification is applied only to its salient regions.
3. Extract the SIFT points in salient regions of both the query and stored DIBR videos.
4. Match the SIFT points in salient regions. In this manner, the rectification efficiency is improved since the number of SIFT points in a salient region is much fewer than that in a whole video frame.
In addition, the distinguishability of rectified features is also enhanced by further exploring the salient information.

5) Rectify the normalized 2-D frames based on the rules described in (6)–(9).

6) Extract a rotation–translation-rectified feature, namely \( f_{\text{rt}} \), and a shearing rectified feature, namely \( f_{\text{sh}} \), following the same extraction steps as in our registration phase.

Because the signal processing attacks, such as noise addition and DIBR conversion, could affect the SIFT matching results, we also extract CT-SVD features from the original query frames, namely \( f_{\text{ct}} \), following the same steps as in the registration phase to further enhance the robustness.

In our scheme, an attention-based fusion is designed inspired by Hua and Zhang to further enhance the performance of copyright identification by exploiting the complement robustness of the rectified and unrectified features. The detailed steps of copyright identification are as follows:

1) Encrypt the rectified and unrectified features by LLCS following the same steps as in the registration phase.

2) Recover three watermarks by XORing these encrypted features with the corresponding ownership share of the stored video.

3) Calculate the three bit error rates (BERs) between the original watermark and these recovered watermarks. The definition of BER can be found in Liu et al.’s work, and the range of its values is [0, 1].

4) Fuse these BER values based on an attention-based fusion model, as shown in (10) and (11), to simultaneously satisfy the heterogeneity and monotonicity defined by Hua and Zhang.

\[
x_1 = (1 - \text{BER}_1) + (1 - \text{BER}_2)
\]

\[
x_2 = |\text{BER}_1 - \text{BER}_2|
\]

\[
U(\text{BER}_1, \text{BER}_2) = 1 - \frac{1}{2} \left( x_1 + \frac{1}{\text{U}} x_2 \right)
\]

\[
\text{BER}_{\text{fus}} = U \left( \text{BER}_{f_1}, \text{BER}_{f_2}, \text{BER}_{f_3} \right)
\]

where \( U \) is the attention-based fusion function, \( \text{BER}_1 \) and \( \text{BER}_2 \) are the BERs to be fused, \( \text{BER}_{f_1} \) and \( \text{BER}_{f_2} \) are obtained from the rectified features, the \( \text{BER}_{f_3} \) is obtained from the unrectified feature, and \( \text{BER}_{\text{fus}} \) is the fusion result. \( \lambda \) is a constant and is set to 0.01 empirically in our study. If any fused BER is smaller than our heuristic threshold, the query DIBR video is treated as an illegal copy.

**EXPERIMENTAL RESULTS AND DISCUSSION**

**Experimental Settings**

The testing database in our experiments comprises 200 different DIBR 3-D video clips from open datasets provided by the MEPG 3Dav group Interactive Visual Media Group of Microsoft Research, and Shenzhen Institute of Advanced Technology, as well as DIBR videos generated by using the method described by Rzeszutek et al. The watermark image is of size 40 \( \times \) 40. To verify the robustness of our system, we apply the attacks listed in Table 1 and attack examples are illustrated in Figure 4. We perform an extensive set of experiments, and compare the obtained results to those yielded by five other state-of-the-art ZR-W algorithms to evaluate the effectiveness of our proposed method. All the experiment results are tested on an Intel Core i7-7700HQ CPU Processor, and a GeForce RTX 2080 Ti GPU Processor in Python implementation.

**Comparison of Video Distinguishability**

We first use the inter-BER metric to compare the distinguishability of our proposed method with the other five schemes. The inter-BER is defined as the BER
value between the genuine and fake watermarks. Here, the genuine watermark is generated by stacking the ownership share and the master share of the same video while fake watermarks are generated by stacking the ownership share with the master share of other videos. A higher inter-BER indicates higher distinguishability. The obtained results are given in Table 2.

As we can see from Table 2, the average inter-BER of our proposed method by fusing the three features is 0.487 and larger than those of Cui et al.’s work, and Wang et al.’s work, and comparative with those of Liu et al.’s work. Moreover, our minimum inter-BER is 0.207 and much larger than those of the other five techniques. These results demonstrate that our proposed method achieves higher distinguishability than other benchmarks. The reason for this superiority is twofold: the selection of CT subbands, containing discriminative horizontal edges and contours, ensures the feature distinguishability, and the deployment of RGB-D salient map detection further enhances the distinguishability of the two rectified features.

Comparison of Watermarking Robustness

Next, we subject the videos to various attacks on all the 200 videos and use the intra-BER metric to evaluate the robustness against the attacks. The intra-BER is defined as the BER value between the watermarks recovered from the original and attacked videos. A smaller mean intra-BER indicates stronger robustness. The comparison results are listed in Table 3. Here, the bold font means the best performance, and the underline means the second-best performance.

As shown in Table 3, when fusing the three features, our proposed method achieves remarkable mean intra-BER values. It can be easily found that our proposed scheme is comparable to the best robustness performance under each attack, which is not achieved by any other benchmark methods. Moreover, our average value of the mean intra-BER is 0.041 and smaller than those by utilizing any single feature and other benchmark methods. These results demonstrate that our proposed scheme is superior to other five schemes in terms of robustness, especially against geometric and DIBR conversion attacks. The reason for this superiority is threefold.

1) Our deployed unrectified CT-SVD feature resists the signal processing and DIBR conversion attacks.
2) Our well-designed rectified feature based on geometric rectification enhances the robustness against geometric attacks.
3) Our proposed attention-based fusion strategy further exploits the complementary robustness of the rectified and unrectified features.

Comparison of Copyright Identification Performance

We then compare the different methods in terms of overall performance for copyright identification, measured by the false positive rate $P_{fp}$ and the false negative rate $P_{fn}$ defined in Liu et al.’s work. A smaller $P_{fn}$ indicates better performance when $P_{fp}$ is fixed.

We set $P_{fp}$ to 0.5% and compare the resulting $P_{fn}$ for all attacks. The results are given in Table 4. Here, the bold font means the best performance and the underline means the second-best performance. We can find that fusing the features in our proposed method gives better results than based on either of

| ZR-W methods | Inter-BER |
|-------------|-----------|
| Average     | Minimum   |
| 7           | 0.490     | 0.031 |
| 8           | 0.431     | 0.078 |
| 9           | 0.177     | 0.013 |
| 10          | 0.497     | 0.142 |
| 11          | 0.485     | 0.096 |
| Proposed method $fr_1$ | 0.498 | 0.241 |
| $fr_2$      | 0.498     | 0.243 |
| $fu$        | 0.492     | 0.207 |
| Fused       | 0.487     | 0.207 |
TABLE 3. Comparison of mean intra-BER values.

| Attack types | 7    | 8    | 9    | 10   | 11   | fr_1 | fr_2 | fu  | Fused |
|--------------|------|------|------|------|------|------|------|-----|-------|
| SN 0.02      | 0.004| 0.055| 0.035| 0.050| 0.095| 0.051| 0.207| 0.014| 0.015 |
| SN 0.05      | 0.013| 0.075| 0.038| 0.067| 0.106| 0.062| 0.252| 0.021| 0.022 |
| GN 0.02      | 0.013| 0.088| 0.061| 0.082| 0.115| 0.096| 0.191| 0.025| 0.025 |
| GN 0.05      | 0.025| 0.135| 0.071| 0.119| 0.123| 0.133| 0.209| 0.036| 0.036 |
| GB 3 x 3     | 0.012| 0.026| 0.011| 0.047| 0.017| 0.045| 0.087| 0.034| 0.033 |
| GB 5 x 5     | 0.016| 0.034| 0.016| 0.052| 0.019| 0.058| 0.111| 0.051| 0.048 |
| AF 3 x 3     | 0.015| 0.032| 0.013| 0.012| 0.005| 0.073| 0.091| 0.039| 0.039 |
| AF 5 x 5     | 0.025| 0.066| 0.031| 0.031| 0.011| 0.093| 0.155| 0.075| 0.072 |
| MF 3 x 3     | 0.007| 0.029| 0.014| 0.022| 0.007| 0.081| 0.171| 0.026| 0.025 |
| MF 5 x 5     | 0.015| 0.083| 0.029| 0.041| 0.013| 0.103| 0.191| 0.053| 0.053 |
| BC +20%      | 0.023| 0.032| 0.015| 0.039| 0.025| 0.031| 0.045| 0.027| 0.027 |
| BC -20%      | 0.019| 0.024| 0.013| 0.036| 0.021| 0.029| 0.049| 0.023| 0.023 |
| CC +20%      | 0.024| 0.029| 0.014| 0.044| 0.021| 0.019| 0.031| 0.018| 0.018 |
| CC -20%      | 0.002| 0.003| 0.003| 0.010| 0.002| 0.003| 0.022| 0.002| 0.001 |
| JC 30        | 0.004| 0.018| 0.015| 0.029| 0.013| 0.051| 0.054| 0.026| 0.023 |
| JC 50        | 0.002| 0.014| 0.009| 0.020| 0.010| 0.045| 0.043| 0.019| 0.017 |
| DBL 3%       | 0.078| 0.037| 0.113| 0.134| 0.032| 0.055| 0.054| 0.051| 0.034 |
| DBL 5%       | 0.112| 0.050| 0.142| 0.202| 0.053| 0.069| 0.075| 0.074| 0.053 |
| DBL 7%       | 0.137| 0.066| 0.152| 0.255| 0.073| 0.096| 0.097| 0.092| 0.070 |
| DSR 3%       | 0.062| 0.033| 0.108| 0.136| 0.031| 0.054| 0.056| 0.049| 0.033 |
| DSR 5%       | 0.091| 0.051| 0.135| 0.204| 0.050| 0.083| 0.075| 0.076| 0.052 |
| DSR 7%       | 0.118| 0.064| 0.143| 0.256| 0.068| 0.099| 0.091| 0.096| 0.069 |
| CR 5%        | 0.169| 0.070| 0.085| 0.001| 0.025| 0.018| 0.269| 0.017| 0.016 |
| CR 10%       | 0.251| 0.093| 0.115| 0.048| 0.047| 0.087| 0.315| 0.057| 0.057 |
| SC 50%       | 0.000| 0.015| 0.009| 0.007| 0.022| 0.024| 0.045| 0.024| 0.022 |
| SC 200%      | 0.000| 0.004| 0.002| 0.003| 0.011| 0.006| 0.023| 0.005| 0.004 |
| RT 45°       | 0.469| 0.407| 0.013| 0.058| 0.130| 0.065| 0.468| 0.461| 0.068 |
| RT 60°       | 0.479| 0.410| 0.009| 0.071| 0.153| 0.079| 0.471| 0.475| 0.082 |
| RT 90°       | 0.493| 0.404| 0.000| 0.092| 0.057| 0.113| 0.491| 0.487| 0.117 |
| RT 180°      | 0.497| 0.440| 0.000| 0.124| 0.055| 0.138| 0.493| 0.491| 0.140 |
| CTR right 30%| 0.351| 0.349| 0.179| 0.398| 0.373| 0.006| 0.391| 0.437| 0.009 |
| CTR down 30% | 0.453| 0.475| 0.185| 0.412| 0.362| 0.010| 0.425| 0.493| 0.013 |
| SH (0,0.5)   | 0.319| 0.254| 0.197| 0.354| 0.171| 0.388| 0.035| 0.306| 0.038 |
| SH (0,0.5)   | 0.422| 0.353| 0.193| 0.413| 0.233| 0.473| 0.046| 0.431| 0.049 |
| SH (0,2,0.2) | 0.324| 0.265| 0.211| 0.391| 0.195| 0.433| 0.043| 0.381| 0.045 |
| Average      | 0.144| 0.131| 0.068| 0.121| 0.082| 0.093| 0.168| 0.142| 0.041 |
TABLE 4. Comparison of $p_{fn}$ ($p_{fp} = 0.5\%$).

| Attack types | 7   | 8   | 9   | 10  | 11  | Proposed scheme |
|--------------|-----|-----|-----|-----|-----|-----------------|
|              |     |     |     |     |     | $fr_1$ | $fr_2$ | $fu$ | Fused |
| SN 0.02      | 0.0%| 6.5%| 16.5%| 0.0%| 1.5%| 0.0% | 11.5%| 0.0%| 0.0% |
| SN 0.05      | 0.0%| 8.0%| 18.0%| 0.0%| 2.5%| 0.0% | 19.5%| 0.0%| 0.0% |
| GN 0.02      | 0.0%| 8.5%| 41.0%| 0.5%| 5.0%| 0.0% | 8.5% | 0.0%| 0.0% |
| GN 0.05      | 0.0%| 13.5%| 43.5%| 1.0%| 6.5%| 1.5% | 13.5%| 0.0%| 0.0% |
| GB 3 × 3     | 0.0%| 0.5%| 1.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| GB 5 × 5     | 0.0%| 1.0%| 1.0% | 0.0%| 0.0%| 0.0% | 2.0% | 0.0%| 0.0% |
| AF 3 × 3     | 0.0%| 1.0%| 1.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| AF 5 × 5     | 0.0%| 3.5%| 11.0%| 0.0%| 0.0%| 0.0% | 5.0% | 0.0%| 0.0% |
| MF 3 × 3     | 0.0%| 0.5%| 1.0% | 0.0%| 0.0%| 0.0% | 7.5% | 0.0%| 0.0% |
| MF 5 × 5     | 0.0%| 2.0%| 11.0%| 0.0%| 0.0%| 0.5% | 8.0% | 0.0%| 0.0% |
| BC +20%      | 0.0%| 0.0%| 0.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| BC -20%      | 0.0%| 0.0%| 0.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| CC +20%      | 0.0%| 0.5%| 0.5% | 0.0%| 1.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| CC -20%      | 0.0%| 0.0%| 0.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| JC 30        | 0.0%| 0.0%| 1.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| JC 50        | 0.0%| 0.0%| 0.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| DBL 3%       | 0.0%| 0.0%| 77.5%| 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| DBL 5%       | 0.5%| 0.0%| 90.5%| 1.0%| 0.0%| 5.0% | 0.0% | 0.0%| 0.0% |
| DBL 7%       | 0.5%| 0.5%| 94.5%| 7.5%| 1.5%| 0.0% | 0.0% | 0.0%| 0.0% |
| DBR 3%       | 0.0%| 0.0%| 78.0%| 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| DBR 5%       | 0.5%| 0.0%| 90.5%| 1.5%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| DBR 7%       | 0.5%| 0.5%| 92.0%| 7.0%| 0.5%| 0.0% | 0.0% | 0.0%| 0.0% |
| CR 5%        | 2.0%| 3.5%| 74.5%| 0.0%| 0.0%| 0.0% | 36.5%| 0.0%| 0.0% |
| CR 10%       | 15.5%| 4.5%| 96.0%| 0.0%| 1.0%| 0.0% | 52.5%| 0.0%| 0.0% |
| SC 50%       | 0.0%| 0.0%| 0.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| SC 200%      | 0.0%| 0.0%| 0.0% | 0.0%| 0.0%| 0.0% | 0.0% | 0.0%| 0.0% |
| RT 45°       | 97.5%| 100%| 0.5%| 0.0%| 14.0%| 0.0% | 93.5% | 97.0%| 0.0% |
| RT 60°       | 99.5%| 100%| 0.0%| 0.0%| 21.5%| 0.0% | 95.0% | 97.5%| 0.0% |
| RT 90°       | 99.0%| 98%| 0.0%| 0.0%| 1.0%| 0.5% | 96.5% | 99.5%| 1.0% |
| RT 180°      | 100%| 97.5%| 0.0%| 0.0%| 0.5%| 1.0% | 100% | 100%| 1.5% |
| CTR right 30%| 91.5%| 95.0%| 100%| 75.0%| 97.5%| 0.0% | 85.5% | 93.5%| 0.0% |
| CTR down 30% | 95.5%| 100%| 100%| 90.5%| 95.5%| 0.0% | 90.5% | 100%| 0.0% |
| SH (0,0.5)   | 85.5%| 84.5%| 100%| 68.0%| 38.5%| 87.5%| 0.0% | 75.5%| 0.0% |
| SH (0.5,0)   | 92.5%| 95.5%| 100%| 92.5%| 73.5%| 93.5%| 0.0% | 92.0%| 0.0% |
| SH (0.2,0.2) | 87.0%| 85.5%| 100%| 72.5%| 55.0%| 90% | 0.0% | 89.5%| 0.0% |
| Average      | 24.8%| 26.0%| 38.3%| 11.9%| 11.9%| 7.8% | 20.7% | 24.1%| 0.1% |
the other three features. More interestingly, our proposed method significantly outperforms the other five algorithms, yielding near-perfect identification, with the average $P_{fp}$ as 0.1%. On the contrary, the other five benchmarks all fail for a number of attacks.

Evaluation of the Improvement by Using Salient Map Detection
Because the geometric rectification is the most time-consuming process in our proposed method, it is important to improve the rectification efficiency for the real-world applications. In this section, we evaluate the improvement of our scheme, which applies the salient map detection compared to our previous work.\textsuperscript{12} As shown in Table 5, the average processing time of geometric rectifications without the salient map detection is 0.41 s per video frame, while the value with the salient map detection is merely 0.14 s, including 0.08 s for the salient map detection, and 0.06 s for rectification. This result demonstrates the improvement of the rectification efficiency (nearly three times faster) by using the salient map detection. In addition, it can be found that the performance of robustness is comparable to the scheme without salient map detection, as in Liu et al.’s work.\textsuperscript{12} At the same time, the distinguishability is improved by 1.8% in terms of its average value and 2.9% in terms of its minimum value by adding the salient map detection.

Evaluation of Watermarking Security
We evaluate the security performance by using the LLCS in terms of key-sensitive-BER and encryption-BER values. The former is calculated between different binarized chaotic sequences under different initial values (from 0 to 1 at an increment of $1/10^{16}$) in our study, which indicates the sensitivity to tiny key differences. The latter is the BER values between the extracted features and their encrypted format (using 200 random $s(0)$ in our study). An encryption function is considered to be more secure when these two values are close to 0.5. The test results are listed in Table 6.

As shown in Table 6, The average, maximum and minimum values of the two BER metrics are all approximately 0.5, confirming the remarkable security performance by deploying LLCs: an attacker cannot forge an invalid watermark without the secret key of the LLCs even if he/she has the watermarking information and ownership shares, which cannot be achieved by our previous work.\textsuperscript{12}

**CONCLUSION**
In this article, a novel ZR-W scheme based on horizontal shift-invariant feature and geometric rectification is proposed for the copyright identification of the DIBR 3-D videos. The advantages of our proposed scheme include in the following:

1) By using our well-designed CT-SVD feature, the robustness against signal processing attacks, DIBR attacks, and the video distinguishability is ensured simultaneously.
2) By establishing the geometric rectification, the geometric attacks are resisted.
3) By introducing the RGB-D salient map detection model, the rectification efficiency is improved and the distinguishability of rectified feature are enhanced.
4) 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.
5) By deploying the LLCs to encrypt the features, the watermarking security is ensured.

The experimental results demonstrate that our scheme outperforms the existing ZR-W schemes for

### Table 5. Evaluation of the improvements by using the salient map detection.

|                | Average Inter-BER | Average Intra-BER | Average $P_{fp}$ ($P_{fp}$ ≤ 0.5%) | Average processing time of rectification/frame |
|----------------|-------------------|-------------------|-------------------------------------|-----------------------------------------------|
| Our previous work\textsuperscript{12} | 0.478             | 0.201             | 0.039                               | 0.2%                                          | 0.41 s                                         |
| Proposed method | 0.487             | 0.207             | 0.041                               | 0.1%                                          | 0.14 s                                         |

### Table 6. Security analysis.

|                | Maximum | Average | Minimum |
|----------------|---------|---------|---------|
| Key-sensitive-BER | 0.5505  | 0.5001  | 0.4336  |
| Encryption-BER   | 0.5332  | 0.4999  | 0.4485  |
3-D videos in terms of distinguishability and robustness against geometric attacks simultaneously.

ACKNOWLEDGMENTS
This work was supported in part by the National Natural Science Foundation of China under Grant 61602527 and Grant U1734208, in part by the Natural Science Foundation of Hunan Province, China under Grant 2020JJ4746, and in part by the Guangxi Key Laboratory of Trusted Software under Grant KX202032.

REFERENCES
1. A. Redert et al., “Advanced three-dimensional television system technologies,” in Proc. IEEE 1st Int. Symp. 3-D Data Process. Visual. Transmiss., 2002, pp. 313–319.
2. H.-D. Kim, J.-W. Lee, T.-W. Oh, and H.-K. Lee, “Robust DT-CWT watermarking for DIBR 3-D images,” IEEE Trans. Broadcast., vol. 58, no. 4, pp. 533–543, Dec. 2012.
3. T. Luo, L. Zuo, G. Jiang, W. Gao, H. Xu, and Q. Jiang, “Security of MVD-based 3-D video in 3-D-HEVC using data hiding and encryption,” J. Real-Time Image Process., vol. 17, no. 4, pp. 773–785, 2020.
4. M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, “Robust DT CWT-based DIBR 3-D video watermarking using chrominance embedding,” IEEE Trans. Multimedia, vol. 18, no. 9, pp. 1733–1748, Sep. 2016.
5. S.-C. Pei and Y.-Y. Wang, “Auxiliary metadata delivery in view synthesis using depth no synthesis error model,” IEEE Trans. Multimedia, vol. 17, no. 1, pp. 128–133, Jan. 2015.
6. X. Liu, F. Li, J. Du, Y. Guan, Y. Zhu, and B. Zou, “A robust and synthesized-unseen watermarking for the DRM of DIBR-based 3-D video,” Neurocomputing, vol. 222, pp. 155–169, 2017.
7. X. Liu, R. Zhao, F. Li, S. Liao, Y. Ding, and B. Zou, “Novel robust zero-watermarking scheme for digital rights management of 3-D videos,” Signal Processing: Image Commun., vol. 54, pp. 140–151, 2017.
8. C. Cui, H. Mao, X. Niu, L. Zhang, T. Hayat, and A. Alsaedi, “A novel hashing scheme for depth-image-based-rendering 3-D images,” Neurocomputing, vol. 191, pp. 1–11, 2016.
9. C. Wang, X. Wang, Z. Xia, and C. Zhang, “Ternary radial harmonic Fourier moments based robust stereo image zero-watermarking algorithm,” Inf. Sci., vol. 470, pp. 109–120, 2019.
10. X. Liu et al., “A novel zero-watermarking scheme with enhanced distinguishability and robustness for volumetric medical imaging,” Signal Processing: Image Commun., vol. 92, 2021, Art. no. 116124.
11. X. Liu et al., “Robust and discriminative zero-watermark scheme based on invariant features and similarity-based retrieval to protect large-scale DIBR 3-D videos,” Inf. Sci., vol. 542, pp. 263–285, 2021.
12. X. Liu, Y. Zhang, S. Du, J. Zhang, M. Jiang, and H. Fang, “Discriminative and geometrically robust zero-watermarking scheme for protecting DIBR 3-D videos,” in Proc. IEEE Int. Conf. Multimedia Expo., 2021, pp. 1–6.
13. Minh N. Do and M. Vetterli, “The contourlet transform: An efficient directional multiresolution image representation,” IEEE Trans. Image Process., vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
14. C. Pak and L. Huang, “A new color image encryption using combination of the 1D chaotic map,” Signal Process., vol. 138, pp. 129–137, 2017.
15. J. Zhang et al., “UC-Net: Uncertainty inspired RGB-D saliency detection via conditional variational autoencoders,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 8582–8591.
16. X.-S. Hua and H.-J. Zhang, “An attention-based decision fusion scheme for multimedia information retrieval,” in Proc. Pacific-Rim Conf. Multimedia, 2004, pp. 1001–1010.
17. C. Fehn, K. Schüür, I. Feldmann, P. Kauff, and A. Smolcic, “Distribution of attest test sequences for E4 in MPEG 3DAV,” in Proc. MPEG Meeting-ISO/IEC JTC1/SC29/ WG11, MPEG02/M9219, 2002, pp. 1–3.
18. C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-quality video view interpolation using a layered representation,” in ACM Trans. Graph., vol. 23, no. 3, pp. 600–608, 2004.
19. X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. Kuo, and Q. Peng, “Subjective and objective video quality assessment of 3-D synthesized views with texture/depth compression distortion,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 4847–4861, Dec. 2015.
20. R. Rzeszutek, R. Phan, and D. Androutsos, “Semi-automatic synthetic depth map generation for video using random walks,” in Proc. IEEE Int. Conf. Multimedia Expo., 2011, pp. 1–6.

XIYAO LIU is currently an associate professor with the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include information security, multimedia technology, and deep learning. Liu received the Ph.D. degree from Peking University, Beijing, China. He is a member of the IEEE. Contact him at lxyzoewx@csu.edu.cn.

YAYUN ZHANG’s research interests include information security, multimedia processing, computational vision, and computational intelligence. Zhang received the master’s degree from Central South University, Changsha, China. Contact her at jassiezhang@tencent.com.
SIBO DU is currently an undergraduate in Central South University, Changsha, China. His research interests include information security and multimedia processing. Contact him at 8208180323@csu.edu.cn.

JIAN ZHANG is currently an associate professor with the School of Computer Science and Engineering, Central South University, Changsha, China. His research interests include cyberspace security, cloud computing, and multimedia technology. Zhang received the Ph.D. degree from Central South University. He is the corresponding author of this article. Contact him at jianzhang@csu.edu.cn.

MING JIANG is currently a professor with the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China. His research interests include image processing, machine learning, and information security. Jiang received the Ph.D. degree from New York University, New York, NY, USA. Contact him at mjiang@guet.edu.cn.

HUI FANG is currently with the computer science department, Loughborough University, U.K. His research interests include image/video processing, deep neural networks, and information security. Fang received the Ph.D. degree from the University of Bradford, Bradford, U.K. Contact him at H.Fang@lboro.ac.uk.