The Visual Computer (2022) 38:3775–3789
https://doi.org/10.1007/s00371-021-02220-4

ORIGINAL ARTICLE

Robust HDR video watermarking method based on saliency extraction and T-SVD

Meng Du[1,2] · Ting Luo[1,2] · Haiyong Xu[2] · Yang Song[1] · Chunpeng Wang[3]

Accepted: 18 June 2021 / Published online: 28 June 2021

Abstract
In order to protect the copyright of the high dynamic range (HDR) video, a robust HDR video watermarking method based on saliency extraction and tensor-singular value decomposition (T-SVD) is proposed. Since T-SVD can not only represent high-dimension data, but also remain its intrinsic structure, each frame of the HDR video is considered as the third-order tensor to be transformed by using T-SVD for preserving the main characteristics of the frame. Each frame is divided into non-overlapping blocks, and each block is decomposed by using T-SVD to obtain the orthogonal tensor $\mathcal{U}$, which includes three orthogonal matrices and represents main energies of the frame. Since the second matrix has more correlations of the video frame than other two matrices, it is used to embed watermark for robustness. Moreover, to obtain the trade-off between watermarking robustness and the visual quality, the saliency map of each frame is extracted to predict the most relevant and important areas for determining the watermark embedding strength. The saliency map is computed based on fusing the spatial saliency and the temporal saliency, where the spatial saliency is built by calculating color, intensity and orientation features of the HDR video and the temporal saliency is obtained by using the optical flow. Experiment results show that the proposed watermarking method can resist a variety of tone mapping attacks and video attacks, and is more robust than existing watermarking methods.

Keywords HDR video · Saliency extraction · Tensor-singular value decomposition (T-SVD) · Robust watermarking

1 Introduction

With the development of the Internet technology, the traditional low dynamic range (LDR) video cannot meet people's visual enjoyment. Compared with the traditional LDR video, the high dynamic range (HDR) video provides a wider range of luminance, which can describe a real scene accurately [1, 2]. However, the HDR video cannot be directly displayed by using the current LDR monitor, because of the wide range of luminance [3]. The HDR video is usually converted to the LDR video through tone mapping (TM) without losing too much detail information for the LDR monitor [4–6], and therefore how to protect copyright of the HDR video after TM has become an urgent problem to be solved. The watermarking technology provides a solution to the copyright protection [7–9].

Watermarking technologies can be divided into the spatial domain and transform domain-based methods. The spatial domain-based watermarking method directly modifies pixels values to embed watermark, such as least significant bit (LSB) substitution [10]. Since this type of watermarking method is sensitive to any modification, it is widely used in the image content authentication [11]. However, pixels of the HDR video are floating point values, and it is difficult to build rules for modifying pixels directly, because different HDR videos have different luminance ranges. Recently, some studies convert the floating point format of the HDR image to other encoding formats [12–16]. Cheng et al. [12] and Li et al. [13] embedded watermark into HDR image in the RGBE format and the LogLuv format, respectively, by using LSB, but they do not have robustness.

Compared with the spatial domain-based watermarking method, the transform domain-based watermarking method

✉ Yang Song
songyang@nbu.edu.cn

[1] College of Science and Technology, Ningbo University, Ningbo 315212, China

[2] Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

[3] School of Information, Qilu University of Technology (School Academy of Sciences),, Jinan 250353, China

is more robust and can resist various image attacks [17–19]. Different decompositions were used to design robust watermarking methods, such as discrete cosine transform (DCT) [20], discrete wavelet transform (DWT) [21], singular value decomposition (SVD) [22], Schur decomposition [23], QR decomposition [24] and so on. Bhardwaj et al. utilized the mathematical relationship between the number of video frames and the embedding capacity to select the key frames for embedding watermark [25]. Esfahani et al. embedded watermark into the low entropy part of all three RGB color channels of each frame by combining QR decomposition, SVD, Chirp-Z transform, DWT and entropy [26]. Rasti et al. proposed a video watermarking method based on QR decomposition and entropy analysis, and watermark was embedded into the blocks with small entropy [27]. Above watermarking methods can resist a variety of image attacks, and the increment of the embedding strength will improve watermarking robustness. However, when the embedding strength is increased, it will lead to the visual distortion of the watermarked image. In order to obtain the trade-off between watermarking imperceptibility and robustness, watermarking methods based on human visual perception were studied [28–30].

Hu et al. used contrast sensitivity, luminance adaption and contrast masking to build a just noticeable difference (JND) model for DCT domain, and the JND was utilized to regulate the variation margin of each coefficient for embedding watermark. That method reduced the visual distortion of the watermarked image with high embedding strength [28]. Since visual entropy and edge entropy denote texture and edge information, respectively, Lai et al. utilized them to select the optimal embedding regions, which can decrease visual distortion [29]. Hernandez et al. used the JND estimation, the saliency mapping and the modulation stage to compute the spatiotemporal saliency-modulated JND profile, and it was used to adjust the embedding strength, which can obtain watermarking robustness with the good imperceptibility [30]. However, above visual perception models are designed for the LDR content, and they cannot perceive human visual system (HVS) characteristics for the wide range of luminance in the HDR content. Thus, a special visual perception model should be designed for the HDR content to embed watermark.

Guerrini et al. used luminance, activity and edge perceptions to compute a perceptual mask of the HDR image for DWT domain, and the perceptual mask was utilized to control the watermarking imperceptibility [31]. Since the human eye is sensitive to the high-luminance areas, Yu et al. embedded watermark into the low-luminance areas based on a luminance mask, which is designed based on the modified specular free [32]. Solachidis et al. computed the wavelet transform in the JND-scale space of the HDR image as the embedding domain, and employed the contrast sensitivity function (CSF) to modulate the embedding strength for both of good robustness and imperceptibility [33]. Bai et al. designed the hierarchical embedding intensity and hybrid perceptual mask according to TM, spatial activity and HVS so that the watermark was embedded into different regions with different intensities [34]. Daniel et al. used Luma Variation Tolerance (LVT) curve to design HVS-imperceptibility, which was used to guide watermark embedding into the spatial domain, but the robustness was not high [35]. Above the HDR watermarking methods only consider the visual characteristics of the frame, and the temporal characteristics are ignored for the HDR video.

Compared with the LDR image watermarking method, besides visual perception, the HDR video watermarking method should resist special HDR image processing, such as TM. TM reduces contrast [36, 37] and changes pixel values to a greater extent than traditional image processing, and therefore, it is essential to design the robust HDR image/video watermarking method. Bakhsh et al. applied RGB-to-LogLUV transform on the HDR image to obtain the luminance component, and watermark was embedded into in the DWT domain of the luminance component. [38]. Gholamreza et al. decomposed the HDR image by using DWT, chirp-z transformation and QR decomposition to embed watermark [39]. Xue et al. proposed two watermarking methods based on μ-Law and bilateral filtering, respectively, but the robustness still cannot be improved [40]. Vassilios et al. decomposed the original HDR image into the multiple LDR images through bracketing decomposition to embed watermark [41]. However, above watermarking methods considered each channel independently, and used the two-dimensional transformation to decompose each channel or the luminance component to embed watermark. Thus, those methods cannot make full use of strong correlations of RGB channels to obtain robust features. In order to improve the robustness, the multi-dimensional transformation is required to consider three channels as a whole to remain strong correlations of three channels. Since tensor-singular value decomposition (T-SVD) [42] can combine three channels as the third-order tensor, main characteristics of the HDR video can be preserved for watermarking robustness.

In this paper, a robust HDR video watermarking method by using T-SVD and the saliency map is proposed. Each frame is divided into non-overlapping blocks, and T-SVD is applied on each block to obtain the orthogonal tensor $\mathcal{U}$, which includes the first, second and third orthogonal matrices. Compared with the other two matrices, the second matrix has more correlations of the video frame and is chosen to embed watermark. Moreover, in order to obtain the trade-off between watermarking imperceptibility and robustness, the saliency map of the HDR video is computed based on fusing the spatial saliency and the temporal saliency,

which can be used to determine the embedding strength. The main contributions of the paper are listed as follows.

(1) Different from traditional transformation operated on one signal channel, T-SVD considers each frame of the HDR video as a whole to be transformed so that strong correlations of each frame can be preserved for robust watermark embedding.
(2) The saliency map of the HDR video is extracted based on fusing the spatial saliency and the temporal saliency to balance the imperceptibility and robustness of the HDR video watermarking.
(3) Experimental results show that the proposed method is more robust than the existing watermarking methods.

This paper is organized as follows. Section 2 introduces the background knowledge of related technology. In Sect. 3, we describe the processes of watermarking embedding and extraction. Section 4 discusses and analyzes the experimental results. Finally, Sect. 5 makes a summary.

2 Background

In this section, the related background technologies are introduced. The notation will be introduced in order to be easily described later. Variables are shown in italics, such as a, matrices are shown in bold letters, such as $\mathbf{A}$, and higher-order tensors are shown in calligraphic letters, such as.

2.1 Tensor-singular value decomposition

With the development of the Internet, the multimedia data is developing in a multi-dimensional direction. Tensor is a form of multi-dimensional data, which can store a lot of information. P-order tensor can be written as

$$\mathcal{A}\Theta = \left(a_{i_1 i_2 \cdots_p}\right) \in \mathrm{R}^{l_1 \times l_2 \times \ldots \times l_p}, \tag{1}$$

where $l_1, l_2 \ldots l_p \in \mathrm{Z}$ indicates the number of elements in each dimension. Therefore, the vector can be considered as the first-order tensor, and the matrix can be considered as the second-order tensor.

Higher-order tensors can be represented by a set of matrices. For example, the third-order tensor can be divided into horizontal slice, lateral slice and frontal slice [43], which are represented as $\mathcal{A}_{k::}, \mathcal{A}_{:k:}$ and $\mathcal{A}_{::k}$, respectively, where $k \in \{1, 2, 3\}$.

SVD is an important decomposition method in linear algebra, which is suitable for any matrix. It can decompose a matrix into two orthogonal matrices and a diagonal matrix.

$\mathbf{B} \in R^{m \times n}$ is an image matrix. After SVD decomposition, $\mathbf{B}$ can be decomposed as

$$\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{2}$$

where $\mathbf{U} \in \mathrm{R}^{m \times m}$ and $\mathbf{V} \in \mathrm{R}^{n \times n}$ are the orthogonal matrices, which are left singular vectors and right singular vectors, respectively. $\mathbf{S} \in \mathrm{R}^{m \times n}$ is the diagonal matrix with diagonal values in descending order that are called singular values.

SVD can be used to process two-dimensional data, but when SVD is used to process high-dimensional data, it may destroy the internal structure of high-dimensional data. For high-dimensional data, tensor can be used to process it, which can effectively retain internal structure of high-dimensional data. Thus, for higher-order singular value decomposition, T-SVD can be used. Let $\mathcal{B} \in R^{l_1 \times l_2 \times l_3}$ be the third-order tensor, and $\mathcal{B}$ can be defined as

$$\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{V}^T \tag{3}$$

where $\mathcal{U} \in R^{l_1 \times l_1 \times l_3}$ and $\mathcal{V} \in R^{l_2 \times l_2 \times l_3}$ are orthogonal tensors, respectively, and $\mathcal{S} \in R^{l_1 \times l_2 \times l_3}$ is a diagonal tensor. Equation (3) is called T-SVD.

The third-order tensor $\mathcal{B}$ can also be written as

$$\sum_{k=1}^{l_3} \mathcal{B}_{::k} = \left(\sum_{k=1}^{l_3} \mathcal{U}_{::k}\right)\left(\sum_{k=1}^{l_3} \mathcal{S}_{::k}\right)\left(\sum_{k=1}^{l_3} \mathcal{V}^T_{::k}\right) \tag{4}$$

where $\sum_{k=1}^{l_3} \mathcal{U}_{::k}$ and $\sum_{k=1}^{l_3} \mathcal{V}_{::k}$ are the orthogonal matrices, respectively, and $\sum_{k=1}^{l_3} \mathcal{S}_{::k}$ is the diagonal matrix. Each frame of the HDR video can be considered as the third-order tensor with the size of $n \times m \times 3$, which $l_1 = n$, $l_2 = m$, $l_3 = 3$ and $k \in \{1, 2, 3\}$. Therefore, after operating T-SVD, two orthogonal tensors and one diagonal tensor are obtained according to Eq. (3). The orthogonal tensor $\mathcal{U}_{::k}$ consists of three matrices, and they are named as the first, the second and the third matrices when k is equal to 1, 2 and 3, respectively. Let $\mathbf{U_1}$, $\mathbf{U_2}$ and $\mathbf{U_3}$ be the first, second and third matrices, respectively. In related to the watermarking robustness, $\mathbf{U_2}$ is chosen to embed watermark, and the main discussion will be introduced in Sect. 4.1. To decompose the RGB image, T-SVD preserves the strong correlations of the RGB three channels, when each frame of HDR videos is treated as the third-order tensor.

2.2 Saliency map

The concept of visual saliency was introduced by Itti [44], and the saliency map can predict the most relevant and important areas of the images and videos, which can be used to guide watermark embedding. In the Itti's visual model, First of all, the color, intensity and orientation visual features of the image are extracted, and the Gaussian pyramid of the color, intensity and orientation visual features is formed by subsampling. Then, the center-surround operation is used to compute the color, intensity and orientation saliency maps.

Finally, the color, intensity and orientation saliency maps are merged to obtain the saliency map of the image. Bremond et al. [45] proposed a HDR saliency map method based on the framework of the Itti's model, namely, contrast features (CF) model. Roland et al. considered that the HVS is sensitive to contrasts rather than to absolute differences, which is different from the Itti's model.

In the Itti's visual model, the intensity feature map was computed as

$$I(c,x) = |I(c) \Theta I(x)|. \tag{5}$$

where $c = \{2,3,4\}$,$x = \{c+3, c+4\}$, $I(c)$ and $I(x)$ are the intensity of the cth and xth levels in the pyramid, respectively, and Θ stands for the across-scale subtraction between two maps.

However, in the CF model, the intensity feature map was modified as

$$I'(c,x) = \frac{|I(c) \Theta I(x)|}{I(x)} \tag{6}$$

In the Itti's visual model, the orientation feature map was computed by the differences between Gabor filters at scales c and x.

$$O(c,x,\theta) = |O(c,\theta) \Theta O(x,\theta)|, \tag{7}$$

where $O(c,\theta)$ and $O(x,\theta)$ are the orientation pyramid obtained by convolving the intensity of the cth and xth levels with Gabor filters, respectively, and $\theta \in \{0°, 45°, 90°, 135°\}$.

In the CF model, its orientation feature map was obtained as

$$O'(c,x,\theta) = \frac{O(c,\theta)}{I(x)}, \tag{8}$$

In the CF model, the color feature is obtained by linear combinations of R, G and B, which is the same as that of the Itti's visual model. However, the HDR content has the rich color gamut, and the color perception under different luminance ranges was not considered in the CF model. Therefore, calculating the color feature in the CF model is not suitable for the HDR content. In order to overcome this issue, the color appearance model (CAM) [46] is employed to extract the color feature, which describes how the HVS perceives color information under different luminance ranges.

3 Proposed HDR video watermarking method

In this section, firstly, the saliency map of the HDR video is extracted to guide watermark embedding. Then, processes of watermark embedding are depicted based on T-SVD and saliency map extraction. Finally, processes of watermark extraction are introduced.

3.1 Saliency map extraction of the HDR video

The saliency map **A** of the HDR video is extracted from the spatial saliency map **map1** and temporal saliency map **map2**. Since the human eye is sensitive to the luminance difference of the HDR image, the intensity feature can accurately reflect the luminance difference of the HDR image, and locate the significant areas. The color feature is a global feature, and it describes the distribution of the image information. Moreover, the color feature helps compute the intensity feature. Since the HVS has different perception sensitivity in the different orientation of the HDR image, the orientation feature computes different orientations of the image. For **map2**, the optical flow [47] is used to compute the motion information under different luminance ranges, which can use the change of pixels in the temporal domain and the correlations of adjacent frames to find the relationship between the previous frame and the current frame. To sum up, the color, intensity and orientation features of the HDR video are extracted to compute **map1**, and the optical flow is used to compute **map2** as illustrated in Fig. 1.

Since the HVS is sensitive to contrasts rather than to absolute differences, the intensity and orientation saliency maps are extracted by using the CF model. Compared with the LDR video, the HDR video has richer color levels, and the HVS perceives color differently under different luminance ranges [48]. Since CAM describes how the HVS perceives color information under different luminance ranges, the color feature of the HDR video is extracted by using the CAM. In the human eye, cone cells are responsible for color perception, and there are L-cones, M-cones and S-cones, which are sensitive to long, medium and short wavelengths, respectively. The color feature is computed in detail as follows.

Step. 1 The XYZ tristimulus values are transformed into LMS cone space by using Hunt–Pointer–Estevez (HPE) [46].

Step. 2 The cones' absolute responses are obtained by

$$L' = \frac{L^{n_c}}{L^{n_c} + L_a^{n_c}}, M' = \frac{M^{n_c}}{M^{n_c} + L_a^{n_c}}, S' = \frac{S^{n_c}}{S^{n_c} + L_a^{n_c}} \tag{9}$$

where L_a is the absolute level of adaptation, which is measured in cd/m^2. L^{n_c}, M^{n_c} and S^{n_c} are L-cones, M-cones and S-cones, respectively, and n_c is set to 0.57 [49].

Step. 3 Red–Green channel $C_{R\text{-}G}$ is obtained as

$$C_{R-G} = \frac{1}{11}(11 \cdot L' - 12M' + S') \tag{10}$$

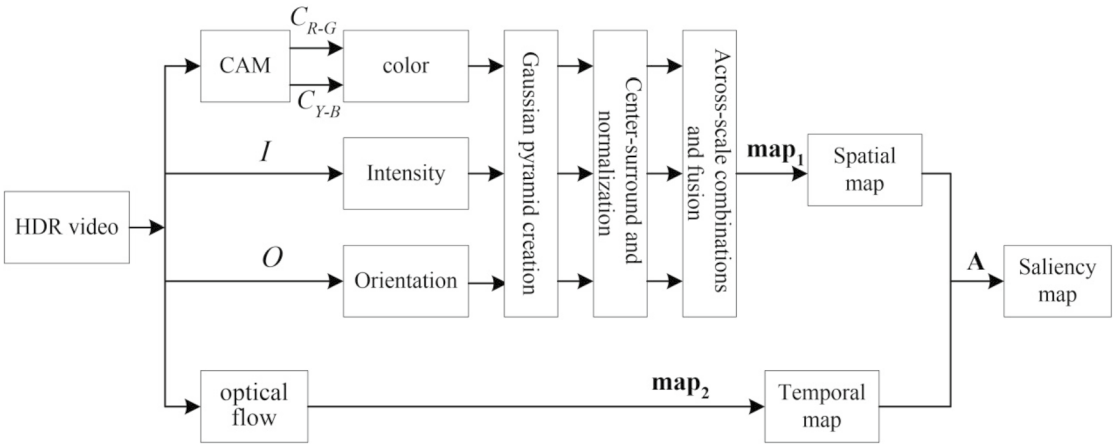


Fig. 1 Proposed saliency map extraction model

Step. 4 Yellow–Blue channel $C_{Y\text{-}B}$ is obtained as

$$C_{Y-B} = \frac{1}{9}(L' + M' - 2 \cdot S') \tag{11}$$

map2 of the HDR video was computed by using the optical flow. The optical flow is the distribution of the surface velocities of movement of object in an image, which can be generated by the relative motion of object and observer. The speed of objects in the HDR video can be measured by estimating the optical flow between video frames. The optical flow is based on two assumptions. The first assumption is that the brightness is constant, which denotes that the brightness of a small region remains persistent despite its position changes. The second assumption is that the space is smooth, which denotes that adjacent points on object have similar velocities and the velocity filed of object is smooth. The process of the saliency map calculation of the HDR video is described as follows

Step. 1: Construct the Gaussian pyramid of the color, intensity and orientation visual features.
Step. 2: The color saliency map **C**, the intensity saliency map I and the orientation saliency map **O** are calculated as

$$\mathbf{C} = NOR(C_{R-G}) + NOR(C_{Y-B}) \tag{12}$$

$$\mathbf{I} = NOR(I'(c,x)) \tag{13}$$

$$\mathbf{O} = NOR(O'(c,x,\theta)) \tag{14}$$

where NOR(·) is the normalization operation.
Step. 3: The spatial saliency map **map1** is computed as

$$\mathbf{map_1} = (\mathbf{C} + \mathbf{I} + \mathbf{O})/3 \tag{15}$$

Step. 4: The temporal saliency map **map2** is computed by using the optical flow.
Step. 5: The saliency map **A** of the HDR video is obtained by fusing **map1** and **map2**.

$$\mathbf{A} = \mathbf{map_1} + \mathbf{map_2} \tag{16}$$

3.2 Watermark embedding

In this section, the process of watermark embedding is presented as illustrated in Fig. 2.

Step. 1 Each frame is regarded as the third-order tensor $\mathcal{A}$, with the size of $M \times N \times 3$. $\mathcal{A}$ is divided into non-overlapping blocks with the size of $n_b \times n_b \times 3$, and each block is denoted as $\mathcal{K}^s$, where s is the index of each block.
Step. 2 Perform T-SVD on each block.

$$\mathcal{K}^s = \mathcal{U}^s \times \mathcal{S}^s \times (\mathcal{V}^s)^T. \tag{17}$$

where $\mathbf{U}_2^s$ is the second matrix of $\mathcal{U}^s$.
Step. 3 The saliency Map **A** is obtained by Eq. (16). The embedding strength matrix **T** is computed according to Eq. (18), and **T** is used to guide watermark embedding.

$$\begin{cases} \mathbf{T}(i,j) = \partial_{\min} & \mathbf{A}(i,j) \geq Med \\ \mathbf{T}(i,j) = \partial_{\max} & \mathbf{A}(i,j) < Med \end{cases}, \tag{18}$$

where $Med = (\max(\mathbf{A}) + \min(\mathbf{A}))/2$. $\partial_{\min}$ and $\partial_{\max}$ are the minimum and maximum embedding strength, respectively, which will be discussed in Sect. 4.2.

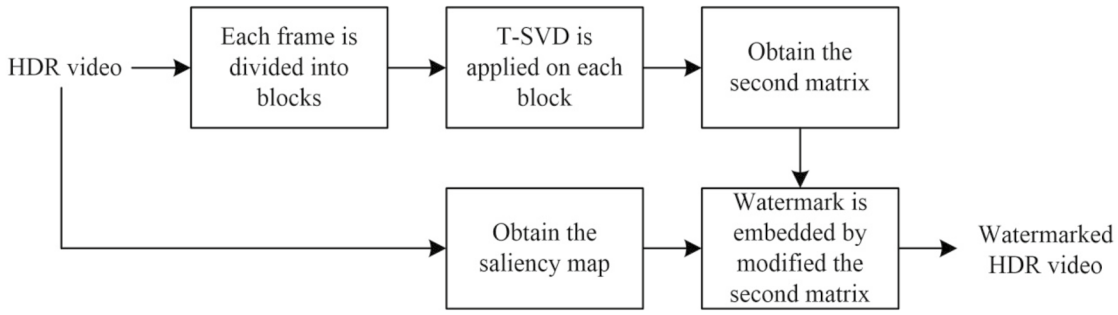


Fig. 2 Watermark embedding

Step. 4. In order to provide the appropriate embedding strength for each block, the embedding strength matrix $\mathbf{T}$ is also divided into non-overlapping blocks with the size of $n_b \times n_b$, which is denoted as $\mathbf{Q}^s$. Each bit of $\mathbf{W}$ is embedded into $\mathbf{U}_2^s$.

$$\begin{cases} \mathbf{U}_2^s(2,1) = sign(\mathbf{U}_2^s(2,1)) \times (u + m/2) \\ \mathbf{U}_2^s(3,1) = sign(\mathbf{U}_2^s(3,1)) \times (u - m/2) \end{cases} \quad if\ \mathbf{W}(i,j) = 1 \tag{19}$$

$$\begin{cases} \mathbf{U}_2^s(2,1) = sign(\mathbf{U}_2^s(2,1)) \times (u - m/2) \\ \mathbf{U}_2^s(3,1) = sign(\mathbf{U}_2^s(3,1)) \times (u + m/2) \end{cases} \quad if\ \mathbf{W}(i,j) = 0 \tag{20}$$

where $u = (abs(\mathbf{U}_2^s(2,1)) + abs(\mathbf{U}_2^s(3,1)))/2$,$m = sum(\mathbf{Q}^s)/16$, and sum$(\cdot)$ returns the sum of the matrix. The modified orthogonal tensor $\mathcal{U}_w^s$ is obtained.
Step. 5 Perform inverse T-SVD on each block.

$$\mathcal{K}_w^s = \mathcal{U}_w^s \times \mathcal{S}^s \times (\mathcal{V}^s)^T \tag{21}$$

Step. 6 Repeat steps 1 to 5 until watermark is embedded into all frames of the HDR video.

If the proposed method is applied in the gray-level video, a group of frames can be represented as a tensor with three dimensions, so that temporal correlation can be mined. A group of frames as the tensor is divided into non-overlapping blocks, and each block is decomposed by using T-SVD to extract $\mathbf{U}_2^s$ for embedding watermark.

3.3 Watermark extraction

Watermark extraction is the reverse process of watermark embedding as illustrated in Fig. 3.

Step. 1 Each watermarked frame is regarded as the third-order tensor $\mathcal{A}^*$, which is divided into non-overlapping blocks with the size of $n_b \times n_b \times 4$, and each block is denoted as $\mathcal{K}^{*s}$.
Step. 2 Perform T-SVD on each block

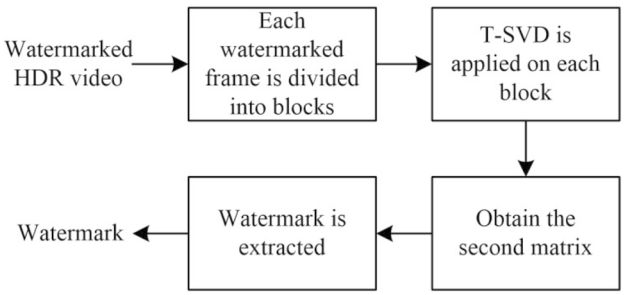


Fig. 3 Watermark extracting

$$\mathcal{K}^* = \mathcal{U}^* \times \mathcal{S}^* \times (\mathcal{V}^*)^T \tag{22}$$

where $\mathbf{U}_2^{*s}$ is the second matrix of $\mathcal{U}^{*s}$.
Step. 3 Each bit of watermark $\mathbf{W}^*$is extracted as

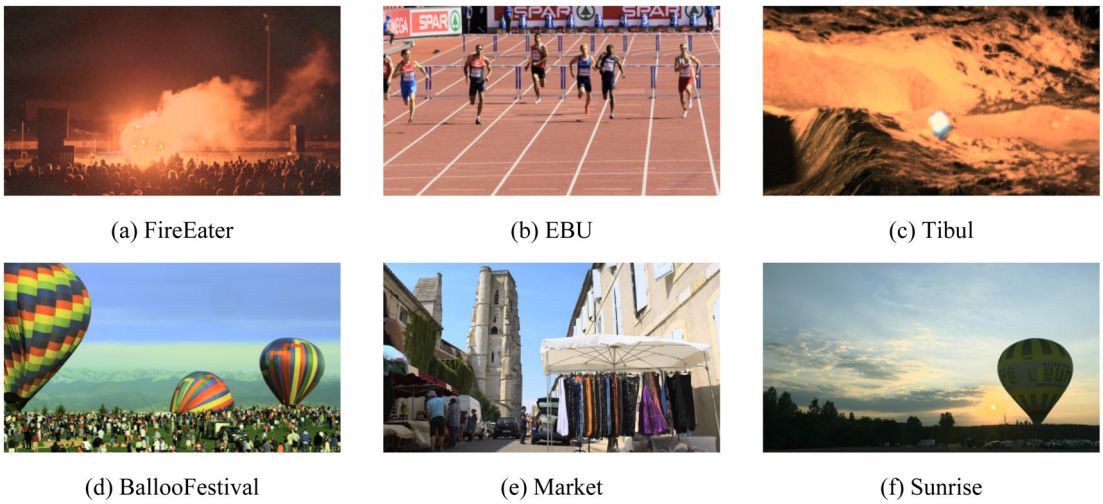
$$\begin{cases} \mathbf{W}^*(i,j) = 1 \ \text{ if } abs(\mathbf{U}_2^{*s}(2,1)) \geq abs(\mathbf{U}_2^{*s}(3,1)) \\ \mathbf{W}^*(i,j) = 0 \ \ if abs(\mathbf{U}_2^{*s}(2,1)) < abs(\mathbf{U}_2^{*s}(3,1)) \end{cases} \tag{23}$$

Step.4 Repeat steps 1 to 3 until watermark is extracted from all frames of the watermarked HDR video.

4 Experimental results and discussion

In this section, FireEater, EBU, Tibul, BallooFestival, Market and Sunrise HDR videos as illustrated in Fig. 4 are used for testing, which contains 150 frames, and the resolution of each frame is the $1920 \times 1080 \times 3$. Watermark is illustrated in Fig. 5, and 18 types of tone mapping (TM) attacks are selected from HDR Toolbox as shown in Table.1. Figure 6 shows the saliency maps of HDR videos, where the most relevant and important areas of the HDR video are extracted.

The HDR-VDP-2 metric is used to evaluate the quality of the watermarked HDR video [50], in which q is the imperceptibility index and is from 0 to 100. HDR-VDP$_{75\%}$ and HDR-VDP$_{95\%}$ are the probabilities of detection in at least 75% and 95% of the images, respectively. High q



(a) FireEater (b) EBU (c) Tibul

(d) BallooFestival (e) Market (f) Sunrise

Fig. 4 HDR videos

Fig. 5 Watermark

Table 1 18 types of TM attacks

TM_s	Name	TM_s	Name
TM_1	ChiuTMO	TM_2	AshikhminTMO
TM_3	BanterleTMO	TM_4	DragoTMO
TM_5	ExponentialTMO	TM_6	KrawczykTMO
TM_7	KimKautzConsistentTMO	TM_8	LischinskiTMO
TM_9	LogarithmicTMO	TM_{10}	MertensTMO
TM_{11}	NormalizeTMO	TM_{12}	PattanaikTMO
TM_{13}	SchlickTMO	TM_{14}	ReinhardTMO
TM_{15}	VanHaterenTMO	TM_{16}	WardHistAdjTMO
TM_{17}	YPTumblinTMO	TM_{18}	TumblinTMO

denotes high visual quality, and high HDR-$VDP_{75\%}$ and HDR-$VDP_{95\%}$ denote the low visual quality of the watermarked HDR video on the contrary. The bit error rate (*BER*) is used to evaluate the correctness of watermark extraction, which can be expressed as the watermarking robustness.

$$BER = \frac{N_w}{N_t} \tag{24}$$

where N_w and N_t are the number of false watermark bits and the number of total watermark bits, respectively. Normalized correlation (*NC*) is used to evaluate the similarity between extracted watermark and original watermark as well.

$$NC = \frac{\sum_{i=1}^{K}\sum_{i=1}^{J}(\mathbf{W}\times\mathbf{W}^*)}{\sqrt{\sum_{i=1}^{K}\sum_{i=1}^{J}(\mathbf{W}\times\mathbf{W})}\sqrt{\sum_{i=1}^{K}\sum_{i=1}^{J}(\mathbf{W}^*\times\mathbf{W}^*)}} \tag{25}$$

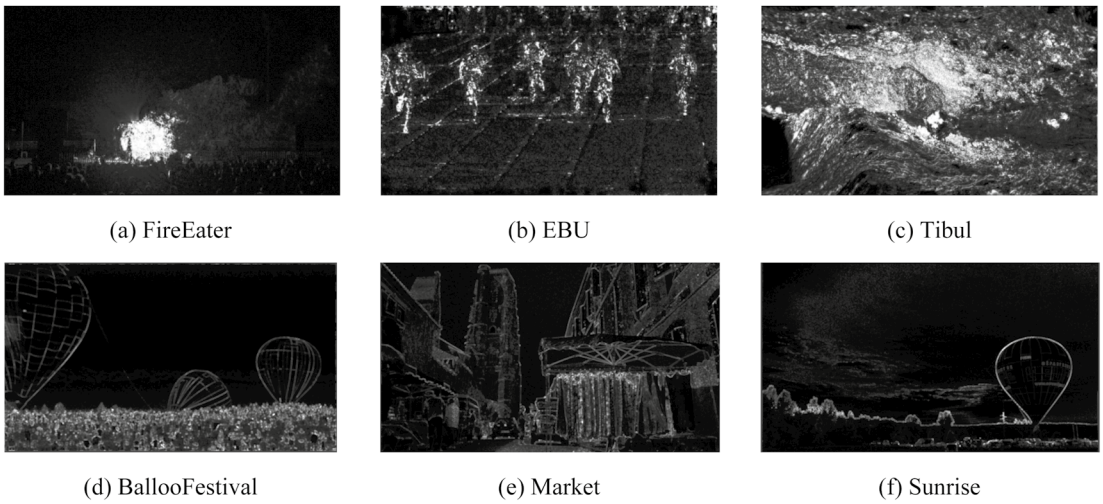
where $\mathbf{W}$ and $\mathbf{W}^*$ are the original watermark and extracted watermark, respectively, and $K\times J$ is the size of watermark.

4.1 Discussion of $\mathbf{U_1}$, $\mathbf{U_2}$ and $\mathbf{U_3}$

In order to select the most suitable matrix to embed watermark, watermark is embedded into three orthogonal matrices $\mathbf{U_1}$, $\mathbf{U_2}$ and $\mathbf{U_3}$ by using the same way in Sect. 3.2, and these three methods are named as Proposed-$\mathbf{U_1}$, Proposed-$\mathbf{U_2}$ and Proposed-$\mathbf{U_3}$, respectively. Watermark is embedded into all six HDR videos, and 3 types of TM attacks are operated on those watermarked HDR videos. Averages *BER* of these HDR video are shown in Table 2, and obviously *BER*s of Proposed-$\mathbf{U_2}$ are lower than those of Proposed-$\mathbf{U_1}$ and Proposed-$\mathbf{U_3}$, which denotes that Proposed-$\mathbf{U_2}$ is more robust and $\mathbf{U_2}$ has more correlations of the video frame. Thus, compared with Proposed-$\mathbf{U_1}$ and Proposed-$\mathbf{U_3}$, Proposed-$\mathbf{U_2}$ is more suitable to be selected to protect the copyright of the HDR video.

4.2 Discussion of the embedding strength

$\partial_{\min}$ and $\partial_{\max}$ are related to the trade-off between the invisibility and robustness of the proposed method. Firstly, the watermarked HDR videos with respect to different embedding strengths are obtained by using the proposed watermarking method. Then, the imperceptibility index and the robustness against several typical TM attacks are computed. The initial value of the intensity $\partial_{\min}$ and $\partial_{\max}$ is



(a) FireEater (b) EBU (c) Tibul

(d) BallooFestival (e) Market (f) Sunrise

Fig. 6 Saliency map of HDR videos

Table 2 Average *BER* of all HDR videos

Attacks	Proposed-$\mathbf{U_1}$	Proposed-$\mathbf{U_2}$	Proposed-$\mathbf{U_3}$
TM_1	0.1163	0.0335	0.0478
TM_2	0.0563	0.0173	0.0213
TM_5	0.1298	0.0236	0.0368

Table 3 The embedding strength of HDR videos

HDR video	$\partial_{\min}$	$\partial_{\max}$
FireEater	0.01	0.02
EBU	0.04	0.06
Tibul	0.02	0.03
BallooFestival	0.02	0.04
Market	0.03	0.05
Sunrise	0.01	0.03

both set from 0.01 to 0.1, and $\partial_{\max}$ is greater than $\partial_{\min}$. The values of $\partial_{\min}$ and $\partial_{\max}$ are gradually increased until the watermarked HDR video is discovered visually. f is computed under different embedding strengths, and the large f represents that $\partial_{\min}$ and $\partial_{\max}$ are most suitable for embedding watermark. Let FireEater be an example, from the subjective perspective, when $\partial_{\min}$ and $\partial_{\max}$ exceed 0.03, the watermarked FireEater can be discovered visually. When $\partial_{\min}$ and $\partial_{\max}$ are assigned to different values, different q and *BER* are computed.

In order to obtain optimal $\partial_{\min}$ and $\partial_{\max}$ for different HDR videos, Eq. (26) is applied [32].

$$f = q_y + \frac{1}{5} \times \left(\sum_{y=1}^{5} (1 - BER_y) \times 100\right), \tag{26}$$

where y is the type of the TM attacks, which are TM_1, TM_2, TM_3, TM_4 and TM_5, respectively. q_y is the imperceptibility index of the watermarked HDR video with different $\partial_{\min}$ and $\partial_{\max}$. BER_y is the bit error rate under the yth TM attack. When $\partial_{\min} = 0.01$ and $\partial_{\max} = 0.02$, f has the maximum value. Similarly, $\partial_{\min}$ and $\partial_{\max}$ of other HDR videos can be obtained as shown in Table3.

Table 4 The results of different n_b

n_b	2	4	6	12
Embedding capacity [.bits]	518,400	129,600	32,400	14,400
Imperceptibility (q) [.dB]	59.72	76.39	81.04	83.25
TM_2	0.0032	0.0078	0.0608	0.0690
TM_4	0.0041	0.0066	0.0395	0.0430
TM_6	0.0091	0.0163	0.0480	0.0807

4.3 Discussion of n_b

In this section, the block size n_b is discussed in the experiment for the performance of the proposed method. As shown in Table 4, when n_b is 2, the embedding capacity is largest and watermarking imperceptibility is the lowest. As n_b is increased, the corresponding embedding capacity is decreased and the imperceptibility is increased. Moreover, for different values of n_b, watermarking robustness is similar. Thus, n_b is only related to the embedding capacity and imperceptibility, and can be set for different requirements.

Table 5 The invisibility of the HDR videos

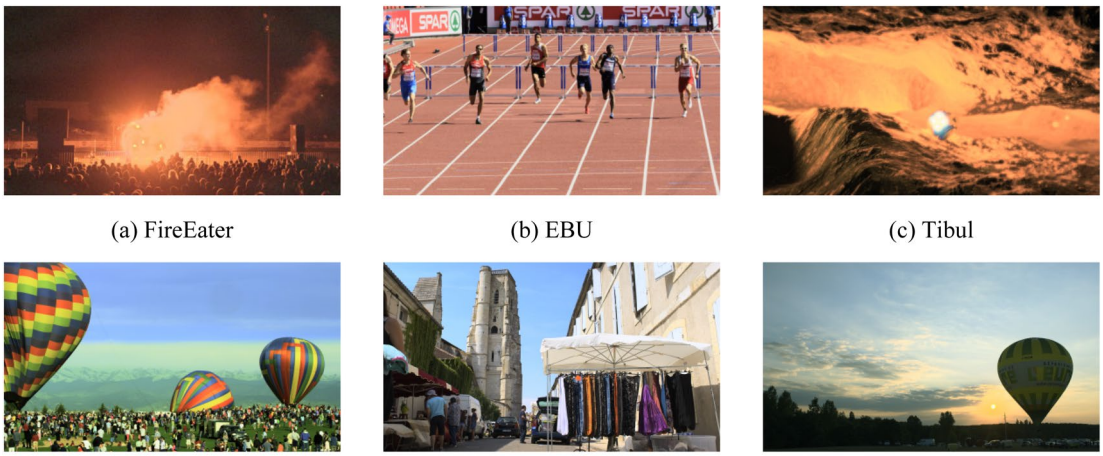
HDR videos	HDR-VDP$_{75\%}$ (100%)	HDR-VDP$_{95\%}$ (100%)	q [.dB]
FireEater	0.04	0.02	79.95
EBU	3.96	2.45	69.58
Tibul	16.96	7.55	76.39
BallooFestival	17.38	12.95	73.18
Market	2.78	1.59	72.24
Sunrise	6.21	3.52	71.83
Average	7.89	4.68	73.86

4.4 Invisibility and robustness

Table5 shows values of HDR-VDP$_{75\%}$, HDR-VDP$_{95\%}$ and q, and it is obviously that averages of HDR-VDP$_{75\%}$, HDR-VDP$_{95\%}$ and q are 7.888%, 4.680% and 73.861, respectively, which denote the watermarked HDR video cannot be observed by human visions as illustrated in Fig. 7, since the saliency map guides watermark embedding effectively. Figure 8 shows nearly 100% of the watermark image can be extracted from the different watermarked HDR videos.

In order to prove the robustness of the proposed HDR video watermarking method on TM attacks, 18 TM attacks are operated on FireEater, EBU, Tibul, BallooFestival, Market and Sunrise, respectively. Figure 9 shows the attacked HDR Sunrise by using the part of TM attacks. From Table6, we can see that *BER*s of FireEater, EBU, Tibul, BallooFestival, Market and Sunrise are lower than 0.15, which denotes that the proposed method can resist different TM attacks.

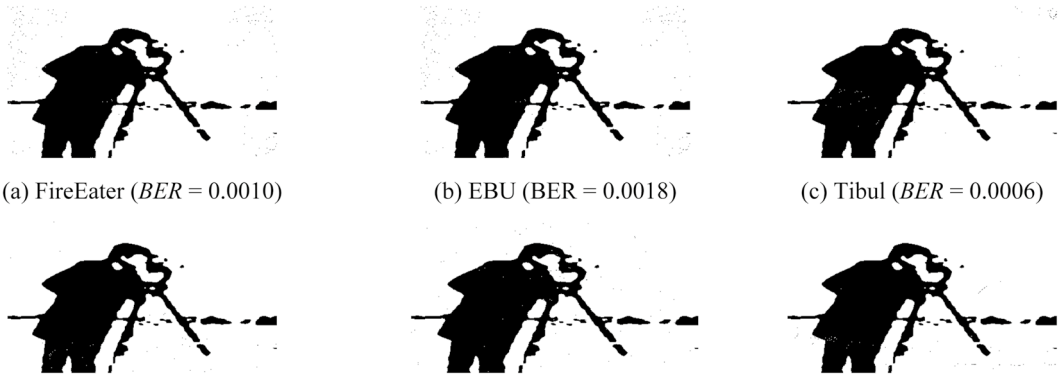
In order to show the robustness on video attacks, a variety of video attacks are operated as shown in Table7, such as frame averaging, and H.265 (encoder_intra_main10, GOP = 1, QP = 22). For frame averaging, we use the average of current frame and its two nearest neighbors to replace the



(a) FireEater (b) EBU (c) Tibul

(d) BallooFestival (e) Market (f) Sunrise

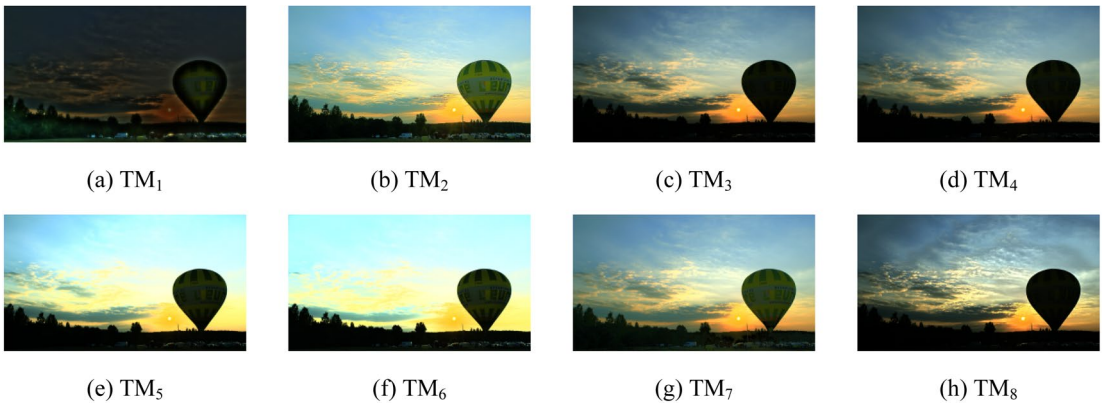
Fig. 7 Watermarked HDR videos



(a) FireEater (*BER* = 0.0010) (b) EBU (BER = 0.0018) (c) Tibul (*BER* = 0.0006)

(d) BallooFestival (*BER* = 0.0003) (e) Market (*BER* = 0.0008) (f) Sunrise (*BER* = 0.0005)

Fig. 8 Extracted watermark



(a) TM_1 (b) TM_2 (c) TM_3 (d) TM_4

(e) TM_5 (f) TM_6 (g) TM_7 (h) TM_8

Fig. 9 Attacked HDR sunrise by using TM attacks

Table 6 *BER* of 18 TM attacks

TM attacks	FireEater	EBU	Tibul	BallooFestival	Market	Sunrise
TM_1	0.0207	0.0301	0.0233	0.0489	0.0564	0.0347
TM_2	0.0050	0.0225	0.0078	0.0229	0.0163	0.0103
TM_3	0.0045	0.0442	0.0123	0.0440	0.0504	0.0112
TM_4	0.0040	0.0191	0.0066	0.0087	0.0222	0.0117
TM_5	0.0056	0.0321	0.0125	0.0160	0.0384	0.0319
TM_6	0.0488	0.0453	0.0163	0.0239	0.0479	0.0505
TM_7	0.0221	0.0288	0.0180	0.0376	0.0343	0.0277
TM_8	0.0017	0.0048	0.0009	0.0038	0.0043	0.0021
TM_9	0.0024	0.0496	0.0090	0.0346	0.0662	0.0424
TM_{10}	0.0578	0.0774	0.0380	0.0626	0.1221	0.0522
TM_{11}	0.0014	0.0033	0.0007	0.0036	0.0028	0.0018
TM_{12}	0.1112	0.0686	0.0166	0.0054	0.0594	0.0748
TM_{13}	0.0027	0.0402	0.0138	0.0601	0.0375	0.0309
TM_{14}	0.0029	0.0179	0.0050	0.0079	0.0179	0.0118
TM_{15}	0.0024	0.0548	0.0095	0.0394	0.0754	0.0509
TM_{16}	0.0549	0.0313	0.0073	0.0307	0.0225	0.0237
TM_{17}	0.0225	0.0276	0.0289	0.0765	0.0654	0.0552
TM_{18}	0.0097	0.0044	0.0044	0.0092	0.0088	0.0076
Average	0.0211	0.0334	0.0128	0.0298	0.0417	0.0295

Table 7 *BER* of video attacks

Attacks	FireEater	EBU	Tibul	BallooFestival	Market	Sunrise
Frame averaging	0.0028	0.0125	0.0013	0.0034	0.0056	0.0117
Frame swapping (10%)	0.0010	0.0018	0.0006	0.0003	0.0008	0.0005
Frame swapping (20%)	0.0010	0.0018	0.0006	0.0003	0.0008	0.0005
Frame dropping (10%)	0.0010	0.0018	0.0006	0.0003	0.0008	0.0005
Frame dropping (20%)	0.0010	0.0018	0.0006	0.0003	0.0008	0.0005
H.265	0.0197	0.1892	0.0297	0.1894	0.0503	0.0478
Average	0.0044	0.0348	0.0056	0.0323	0.0098	0.0103

current frame. Frame averaging is to compute the average of multiple frames, which will change the video much and destroy watermark. However, the proposed method still can extract watermark since corresponding *BER* is higher than 90%, and it shows the robustness of the proposed method. In the proposed method, watermark is embedded into each

Table 8 *BER* of hybrid attacks

Attacks	FireEate	EBU	Tibul	BallooFestival	Market	Sunrise
Gaussian filter(3×3) + TM_{11}	0.0242	0.0485	0.0077	0.0285	0.0336	0.0771
Sharpen(0.5) + TM_{14}	0.0308	0.0436	0.0195	0.0252	0.0381	0.0239
Passion noise + TM_{15}	0.0173	0.0676	0.0171	0.0395	0.0496	0.0408
Salt & Pepper(0.001) + TM_8	0.0058	0.0484	0.0069	0.0251	0.0651	0.0326
Average	0.0195	0.0520	0.0128	0.0296	0.0466	0.0436

Table 9 Comparison of time complexity, embedding capacity and imperceptibility

	Proposed	Kang's [51]	Joshi's [52]
Embedding capacity [.bits]	129,600	14,400	32,400
Imperceptibility (q) [.dB]	79.95	70.45 dB	66.66
Time complexity [.s]	1289.72	485.67	597.84

frame of the video; thus, frame swapping and frame dropping have little influence on the watermark extraction, and can nearly extract 100% of watermark. Averages *BER* of these HDR video are 0.0044, 0.0348, 0.0056, 0.0323, 0.0098 and 0.0056, respectively, which denote the proposed method has the ability of resisting those video attacks.

In order to show that the proposed method can also resist hybrid attacks, six HDR videos are also operated by using a variety of hybrid attacks, such as Gaussian filter(3×3) + TM_{11}, and Salt & Pepper(0.001) + TM_8. As shown in Table8, we can see that *BERs* of the proposed method are lower than 0.1, and it shows that the proposed method performs well on hybrid attacks.

4.5 Comparative

Table9 shows the comparison results of the proposed method, Kang's [51] and Joshi's [52] on embedding capacity, imperceptibility and time complexity. From Table9, when block size n_b is set to 4, the watermarking embedding capacity and watermarking invisibility are much better, but the running time is higher compared with Kang's [51] and Joshi's [52]. The running time of the proposed method is 1289.72 s, where the saliency map computation costs 298.12 s. But if n_b is increased to 12, the running time is decreased to 697.85 s, which is similar as those of Kang's [51] and Joshi's [52].

Bakhsh's [38], Kang's [51] and Joshi's [52] are used to be compared for demonstrating the robustness of the proposed method when all HDR videos are under different attacks as shown in Table10. From Table10, we can see that the proposed method is better than Bakhsh's [38], Kang's [51] and Joshi's [52] methods. For example, for TM_1, averages *BER* of the proposed method are nearly 0.04, 0.02 and 0.06 lower than those of Bakhsh's [38], Kang's [51] and Joshi's [52], respectively. For TM_6, averages *BER* of the proposed method are nearly 0.05, 0.2 and 0.06 lower than those of Bakhsh's [38], Kang's [51] and Joshi's [52], respectively. Compared with Bakhsh's [38], although average *BER* of the proposed method is higher for TM_{12}, lower for other attacks, such as Sharpen (0.5) + TM_{14}, Passion noise + TM_{15} and Salt & Pepper(0.001) + TM_8. Compared with Kang's [51] and Joshi's [52], the proposed method is obviously better. Considering all kinds of attacks, most *BERs* of the proposed method are

Table 10 Comparison of watermark extraction from all HDR video under different attacks

Attacks	Proposed	Bakhsh's [38]	Kang's [51]	Joshi's [52]
TM_1	0.0357	0.0779	0.0549	0.0904
TM_2	0.0141	0.0937	0.1022	0.0657
TM_4	0.0121	0.0752	0.0889	0.0693
TM_6	0.0387	0.0866	0.2288	0.0937
TM_7	0.0281	0.0859	0.1077	0.1014
TM_{10}	0.0684	0.0866	0.1437	0.1445
TM_{12}	0.0560	0.0509	0.0747	0.0714
TM_{16}	0.0784	0.1040	0.0829	0.1028
H.265	0.0877	0.1084	0.4896	0.4712
Gaussian filter (3×3) + TM_{11}	0.0366	0.0639	0.0768	0.1203
Sharpen(0.5) + TM_{14}	0.0301	0.0869	0.1289	0.0832
Passion noise + TM_{15}	0.0387	0.0680	0.0983	0.0788
Salt & Pepper(0.001) + TM_8	0.0307	0.1001	0.2193	0.1598

Table 11 Comparison of watermark extraction from Tibul under different attacks

Attacks	Proposed	Bakhsh's [38]	Kang's [51]	Joshi's [52]
TM_5	0.9962	0.8035	0.8809	0.9668
TM_8	0.9989	0.8583	0.9721	0.9891
TM_9	0.9984	0.8250	0.9755	0.9887
Scaling(1/4)	0.9767	0.7033	0.8740	0.9342
Sharpen(0.5) + TM_{14}	0.9906	0.7686	0.9221	0.9842
Passion noise + TM_{15}	0.9982	0.7582	0.6116	0.5701
Average	0.9672	0.7827	0.8727	0.8905

Table 12 Comparisons of other HDR videos

Attacks	TM_3	TM_8	TM_{11}	TM_{13}	TM_{15}
Proposed	0.0521	0.0712	0.0483	0.0850	0.0614
Bakhsh's [38]	0.0784	0.0981	0.1011	0.1023	0.0857
Kang's [51]	0.1025	0.2085	0.0977	0.2145	0.1257
Joshi's [52]	0.0987	0.1189	0.1278	0.2793	0.0939

lower than those of Bakhsh's, Kang's and Joshi's, which denotes that the proposed method is superior to above three methods. In summary, the proposed method has the capability of protecting the HDR video and its TMO representations, and it is mainly because T-SVD preserves the robust features of the HDR video.

In order to objectively evaluate the robustness of the proposed method, *NC* is also used. The watermarked Tibul is attacked by a variety of attacks as illustrated in Table11. From Table 11, we can see that *NC*s of the proposed method are similar to those of Joshi's for TM_8 and TM_9, but are higher than those of Joshi's [52] for other attacks, especially for Passion + TM_{15} and Scaling (1/4). Compared with Bakhsh's [38] and Kang's [51], the proposed method is obviously better. In all, average *NC* of the proposed method is higher than those of Bakhsh's [38], Kang's [51] and Joshi's [52]. Thus, the proposed method can resist a variety attacks to protect the HDR videos, which shows the effectiveness of the proposed method.

4.6 Robustness on other HDR videos

In order to verify the effectiveness of the proposed method again, one HDR database [53] consisting of 10 HDR videos is also used for testing. The watermarked HDR videos are attacked by different TM attacks, such as TM_3, TM_8 and so on. From Table 12, we can see that *BER*s of the proposed method are obviously lower than those of Bakhsh's [38], Kang's [51] and Joshi's [52] methods, which denotes that the proposed method has strong robustness, and can efficiently protect the copyright of the HDR video again.

5 Conclusion

In this paper, a robust HDR video watermarking method based on T-SVD and the saliency map is proposed. Each frame can be regarded as the third-order tensor to obtain the most robust domain by using T-SVD. After T-SVD, the orthogonal tensor is calculated, which consists of the first, second and third matrices. Compared with the other two matrices, the second matrix is more robust; therefore, it is more suitable to embed watermark. The saliency map is computed, which can represent the most visually important areas of each frame for determining the embedding strength. Experimental results show that the proposed method can effectively protect the copyright of the HDR videos, and is robust on various attacks. However, the proposed method cannot resist lossy compression not well. In the future work, we will further explore visual perception factors of the HDR video to guide watermark embedding for improving watermarking efficiency.

Acknowledgements This work was supported by Natural Science Foundation of China under Grant No. 61971247 and 61501270, Zhejiang Provincial Natural Science Foundation of China under Grant No. LY19F020009 and LQ20F010002. It was also sponsored by the K. C. Wong Magna Fund in Ningbo University.

Declarations

Conflict of interest We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, "Robust HDR video watermarking method based on saliency extraction and T-SVD."

References

1. Karr, B.A., Debattista, K., Chalmers, A.G.: Optical effects on HDR calibration via a multiple exposure noise-based workflow. Vis. Comput. **37**, 895–910 (2021). https://doi.org/10.1007/s00371-020-01841-5

2. Hatchett, J., Debattista, K., Mukherjee, R., Rogers, T.B., Chalmers, A.: An evaluation of power transfer functions for HDR video compression. Vis. Comput. **34**, 167–176 (2018). https://doi.org/10.1007/s00371-016-1322-0
3. Zhang, Y., Naccari, M., Agrafiotis, D., Mark, M., Bull, D.R.: High dynamic rang video compressing exploiting luminance masking. IEEE Trans. Circuits Syst. Video Technol. **26**(5), 950–964 (2016). https://doi.org/10.1109/TCSVT.2015.2426552
4. Cagri, O., Paul, L., Giuseppe, V., Frédéric, D.: Spatio-temporal constrained tone mapping operator for HDR video compression. J. Vis. Commun. Image Represent. **55**, 166–178 (2018). https://doi.org/10.1016/j.jvcir.2018.06.003
5. Khwildi, R., Zalid, A.O.: HDR image retrieval by using color-based descriptor and tone mapping operator. Vis. Comput. **36**, 1111–1126 (2016). https://doi.org/10.1007/s00371-019-01719-1
6. Eilertsen, G., Mantiuk, R.K., Unger, J.: A comparative review of tone-mapping algorithms for high dynamic range video. Comput. Graph. Forum. (2017). https://doi.org/10.1111/cgf.13148
7. Ernawan, F., Kabir, M.N.: A block-based RDWT-SVD image watermarking method using human visual system characteristics. Vis. Comput. **36**, 19–37 (2020). https://doi.org/10.1007/s00371-018-1567-x
8. Wang, X., Hu, K., Hu, J., Du, L., Ho, A.T.S., Qin, H.: Robust and blind image watermarking via circular embedding and bidimensional empirical mode decomposition. Vis. Comput. **36**, 2201–2214 (2020). https://doi.org/10.1007/s00371-020-01909-2
9. Ahmadi, S.B.B., Zhang, G., Wei, S., Boukela, L.: An intelligent and blind image watermarking scheme based on hybrid SVD transforms using human visual system characteristics. Vis. Comput. (2020). https://doi.org/10.1007/s00371-020-01808-6
10. Zhou, R.G., Hu, W., Fan, P., Luo, G.: Quantum color image watermarking based on Arnold transformation and LSB steganography. Int. J. Quantum Inf. (2018). https://doi.org/10.1142/S0219749918500211
11. Qin, C., Ji, P., Zhang, X., Dong, J., Wang, J.: Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy. Signal Process **138**, 280–293 (2017). https://doi.org/10.1016/j.sigpro.2017.03.033
12. Cheng, Y.M., Wang, C.M.: A novel approach to steganography in high-dynamic-range images. IEEE Multimed. **16**(3), 70–80 (2009). https://doi.org/10.1109/MMUL.2009.43
13. Li, M.T., Huang, N.C., Wang, C.M.: A data hiding scheme for high dynamic range images. Int. J. Innov. Comput. Inf. Control **7**(5), 2021–2035 (2011)
14. Yu, C.M., Wu, K.C., Wang, C.M.: A distortion-free data hiding scheme for high dynamic range images. Displays **32**(5), 225–236 (2011). https://doi.org/10.1016/j.displa.2011.02.004
15. Lin, Y.T., Wang, C.M., Chen, W.S., Lin, F.P., Lin, W.: A novel data hiding algorithm for high dynamic range images. IEEE Trans. Multimed. **19**(1), 196–211 (2017). https://doi.org/10.1109/TMM.2016.2605499
16. Chang, C.C., Nguyen, T.S., Lin, C.C.: A new distortion-free data embedding scheme for high-dynamic range images. Multimed. Tools Appl. **75**(1), 145–163 (2016). https://doi.org/10.1007/s11042-014-2279-5
17. Wang, X., Hu, K., Hu, J., Du, L., Ho, A.T.S., Qin, H.: Correction to: Robust and blind image watermarking via circular embedding and bidimensional empirical mode decomposition. Vis. Comput. **37**, 859 (2021). https://doi.org/10.1007/s00371-020-01948-9
18. Yuan, Z., Su, Q., Liu, D., Zhang, X.: A blind image watermarking scheme combining spatial domain and frequency domain. Vis. Comput. (2020). https://doi.org/10.1007/s00371-020-01945-y
19. Shao, Z., Shang, Y., Zeng, R., Coatrieux, G., Wu, J.: Robust watermarking scheme for color image based on quanternion-type moment invariants and visual cryptography. Signal Process. Image Commun. **48**, 12–21 (2016)
20. Fang, H., Zhou, H., Ma, Z., Zhang, W.: A robust image watermarking scheme in DCT domain based on adaptive texture direction quantization. Multimed. Tools Appl. **78**(7), 8075–8089 (2019). https://doi.org/10.1007/s11042-018-6596-y
21. Lee, S.H.: DWT based coding DNA watermarking for DNA copyright protection. Inf. Sci. **273**(20), 263–286 (2014). https://doi.org/10.1016/j.ins.2014.03.039
22. Najafi, E., Loukhaoukha, K.: Hybrid secure and robust image watermarking scheme based on SVD and sharp frequency localized contourlet transform. J. Inf. Sec. Appl. **44**, 144–156 (2019). https://doi.org/10.1016/j.jisa.2018.12.002
23. Hsu, T.Y., Hu, H.T.: A reinforced blind color image watermarking scheme based on Schur decomposition. IEEE Access **7**, 107438–107452 (2019). https://doi.org/10.1109/ACCESS.2019.2932077
24. Jia, S., Zhou, Q., Zhou, H.: A novel color image watermarking scheme based on DWT and QR decomposition. J. Appl. Sci. Eng.. **20**(2), 193–200 (2017). https://doi.org/10.6180/jase.2017.20.2.07
25. Bhardwaj, A., Verma, V.S., KumarJha, R.: Robust video watermarking using significant frame selection based on coefficient difference of lifting wavelet transform. Multimed. Tools Appl. **77**, 19658–19678 (2018). https://doi.org/10.1007/s11042-017-5340-3
26. Esfahani, R., Mohammad, A.A., Norouzi, Z.: A fast video watermarking algorithm using dual tree complex wavelet transform. Multimed. Tools Appl. **78**(12), 16159–16175 (2019). https://doi.org/10.1007/s11042-018-6892-6
27. P. Rasti, S. Samiei, M. Agoyi, S. Escalera, G, Anbarjafari. Robust non-blind color video watermarking using QR decomposition and entropy analysis. Journal of Visual Communication and Image Representation. 38 (2016) 838–847, https://doi.org/10.1016/j.jvcir.2016.05.001.
28. Hu, H.T., Chang, J.R., Hsu, L.Y.: Robust blind image watermarking by modulating the mean of partly sign-altered DCT coefficients guided by human visual perception. AEU-Int. J. Electr. Commun. **70**(10), 1374–1381 (2016). https://doi.org/10.1016/j.aeue.2016.07.011
29. Lai, C.: An improved SVD-based watermarking scheme using human visual characteristics. Opt. Commun. **284**(4), 938–944 (2011). https://doi.org/10.1016/j.optcom.2010.10.047
30. Hernandez, A.C., Hernandez, M.C., Miyatake, M.N., Meana, H.P.: A spatiotemporal saliency-modulated JND profile applied to video watermarking. J. Vis. Commun. Image Represent. **52**, 106–117 (2018). https://doi.org/10.1016/j.jvcir.2018.02.007
31. Guerrini, F., Okuda, M., Adami, N., Leonardi, R.: High dynamic range image watermarking robust against tone-mapping operators. IEEE Trans. Inf. Forensics Sec. **6**(2), 283–295 (2011). https://doi.org/10.1109/TIFS.2011.2109383
32. Yu, M., Wang, Y., Jiang, G., Bai, Y., Luo, T.: High dynamic range image watermarking based on tucker decomposition. IEEE Access **7**, 113053–113064 (2019). https://doi.org/10.1109/ACCESS.2019.2935627
33. Solachidis, V., Maiorana, E., Campisi, P.: HDR image multi-bit watermarking using bilateral- filtering-based masking. Image Process. Algorithms Syst. XI. **8655**, 865505 (2013). https://doi.org/10.1117/12.2005240
34. Bai, Y., Jiang, G., Yu, M., Peng, Z., Chen, F.: Towards a tone mapping robust watermarking algorithm for high dynamic range image based on spatial activity. Signal Process. Image Commun. **65**, 187–200 (2018). https://doi.org/10.1016/j.image.2018.04.005
35. Daniel, P., Ruby, K., Ugalde, F.G., Sanchez, V.: Watermarking of HDR images in the spatial domain with HVS-Imperceptibility. IEEE Access **8**, 156801–156817 (2020). https://doi.org/10.1109/ACCESS.2020.3019517
36. Kuang, J., Johnson, G.M., Fairchild, M.D.: iCAM06: A refined image appearance model for HDR image rendering. J. Vis. Commun. Image Represent. **18**(5), 406–414 (2007). https://doi.org/10.1016/j.jvcir.2007.06.003

37. Debevec, P., Gibson.,S.: A tone mapping algorithm for high contrast images. In: *Proceedings of the 13th Eurographics workshop on Rendering*. (2002) pp. 145–156
38. Bakhsh, F.Y., Moghaddam, M.E.: A robust HDR images watermarking method using artificial bee colony algorithm. J. Inf. Sec. Appl. **41**, 12–27 (2018). https://doi.org/10.1016/j.jisa.2018.05.003
39. Gholamreza, A., Ozcinar, C.: Imperceptible non-blind watermarking and robustness against tone mapping operation attacks for high dynamic range images. Multimed. Tools Appl. **77**(18), 24521–24535 (2018). https://doi.org/10.1007/s11042-018-5759-1
40. Xue, X., Jinno, T., Jin, X., Okuda, M., Goto, S.: Watermarking for HDR image robust to tone mapping. IEICE Trans. Fundament. Electr. Commun. Comput. Sci. **94**(11), 2334–2341 (2011). https://doi.org/10.1587/transfun.E94.A.2334
41. Solachidis, V., Maiorana, E., Campisi, P., Banterle, F.: HDR image watermarking based on bracketing decomposition. IEEE Int. Conf. Dig. Signal Process. (2013). https://doi.org/10.1109/ICDSP.2013.6622687
42. Kilmer, M.E., Martin, C.D.: Factorization strategies for third-order tensors. Linear Algebra Appl. **435**(3), 641–658 (2011). https://doi.org/10.1016/j.laa.2010.09.020
43. Martin, C.D., Shafer, R., LaRue, B.: An order-p tensor factorization with applications in imaging. SIAM J. Sci. Comput. **35**(1), A474–A490 (2013). https://doi.org/10.1137/110841229
44. Itti, L., Dhavale, N., Pighin, F.: Realistic avatar eye and head animation using a neurobiological model of visual attention. Proc. SPIE **2004**(5200), 64–78 (2004). https://doi.org/10.1117/12.512618
45. Br´emond, R., Petit, J., Tarel, J. P.: Saliency maps of high dynamic range images. In: *Proceedings of* 11th *European Conference on Computer Vision, Trends and Topics in Computer Vision* (2012) pp. 118–130, https://doi.org/10.1007/978-3-642-35740-4_10.
46. Kim, M. H., Weyrich, T., Kautz, J.: Modeling human color perception under extended luminance levels. In: *ACM SIGGRAPG 2009 paper*. 28 (3) (2009) 27, https://doi.org/10.1145/1576246.1531333.
47. Horn, B., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**, 185–203 (1981). https://doi.org/10.1117/12.965761
48. Dong, Y., Pourazad, T.M., Nasiopoulos, P.: Human visual system-based saliency detection for high dynamic range content. IEEE Trans. Multimedia **18**(4), 549–562 (2016). https://doi.org/10.1109/TMM.2016.2522639
49. Uscanga, O. E.: On the fundamental data-base of normal and dichromatic color vision. Doctoral dissertation, Verlag nicht ermittelbar. (1979).
50. Mantiuk, R., Kim, K.J., Rempel, A.G., Heidrich, W.: HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph. (TOG) **30**(4), 1–14 (2011). https://doi.org/10.1145/2010324.1964935
51. Kang, X., Zhao, F., Lin, G., Chen, Y.: A novel hybrid of DCT and SVD in DWT domain for robust and invisible blind image watermarking with optimal embedding strength. Multimed. Tools Appl. **77**(11), 13197–13224 (2018). https://doi.org/10.1007/s11042-017-4941-1
52. Joshi, A., Gupta, S., Girdhar, M., Agarwal, P., Sarker, R.: Combined DWT–DCT-based video watermarking algorithm using Arnold transform technique. Proc. Int. Conf. Data Eng. Commun. Technol. **468**, 455–463 (2017). https://doi.org/10.1007/978-981-10-1675-2_45
53. Azimi, M., Banitalebi-Dehkordi, A., Dong, Y., and Nasiopoulos, P.: Evaluating the performance of existing full-reference quality metrics on High Dynamic Range (HDR) video content. ICMSP 2014: XII international conference on multimedia signal processing. (2018).

Meng Du is currently pursuing the master's degree with the Faculty of Information Science and Engineering, Ningbo University. She is currently a student with the Faculty of Information Science and Engineering, Ningbo University. Her research interests include multimedia security and image processing

Ting Luo received the Ph.D. degree with the Faculty of Information Science and Engineering, Ningbo University in 2016. He is currently a Professor with the College of Science and Technology, Ningbo University. His research interests include multimedia security, image processing, data hiding, and pattern recognition

Haiyong Xu is currently pursuing the Ph.D. degree with the Faculty of Information Science and Engineering, Ningbo University. He is currently a Teacher with the College of Science and Technology, Ningbo University. His research interests include multimedia communication, image processing, and machine learning.

Yang Song received the M.S. degree and Ph.D. degree from Ningbo University, in 2015 and 2018. He is currently a teacher with the College of Science and Technology, Ningbo University. His research interests include multimedia communication, image processing, and visual quality assessment.

Chunpeng Wang received the B.E. degree in computer science and technology in 2010 from Shandong Jiaotong University, China, the M.S. degree from the School of Computer and Information Technology, Liaoning Normal University, China, 2013, and the Ph.D. degree in Faculty of Electronic Information & Electrical Engineering, Dalian University of Technology, China, 2017. He is currently a teacher with the School of Information, Qilu University of Technology (Shandong Academy of Sciences), China. His research mainly includes image watermarking and signal processing.