360-degree VR Video Watermarking based on Spherical Wavelet Transform

YANWEI LIU, Institute of Information Engineering, Chinese Academy of Sciences, China
JINXIA LIU*, Zhejiang Wanli University, China
ANTONIOS ARGYRIOU, University of Thessaly, Greece
SIWEI MA, Peking University, China
LIMING WANG, Institute of Information Engineering, Chinese Academy of Sciences, China
ZHEN XU, Institute of Information Engineering, Chinese Academy of Sciences, China

Similar to conventional video, the increasingly popular 360° virtual reality (VR) video requires copyright protection mechanisms. The classic approach for copyright protection is the introduction of a digital watermark into the video sequence. Due to the nature of spherical panorama, traditional watermarking schemes that are dedicated to planar media cannot work efficiently for 360° VR video. In this paper, we propose a spherical wavelet watermarking scheme to accommodate 360° VR video. With our scheme, the watermark is first embedded into the spherical wavelet transform domain of the 360° VR video. The spherical geometry of the 360° VR video is used as the host space for the watermark so that the proposed watermarking scheme is compatible with the multiple projection formats of 360° VR video. Second, the just noticeable difference model, suitable for head-mounted displays (HMDs), is used to control the imperceptibility of the watermark on the viewport. Third, besides detecting the watermark from the spherical projection, the proposed watermarking scheme also supports detecting watermarks robustly from the viewport projection. The watermark in the spherical domain can protect not only the 360° VR video but also its corresponding viewports. The experimental results show that the embedded watermarks are reliably extracted both from the spherical and the viewport projections of the 360° VR video, and the robustness of the proposed scheme to various copyright attacks is significantly better than that of the competing planar-domain approaches when detecting the watermark from viewport projection.

CCS Concepts: • **Security and privacy**; • **Digital rights management**;

Additional Key Words and Phrases: 360° VR video, watermarking, spherical wavelet, just noticeable difference

ACM Reference Format:
Yanwei Liu, Jinxia Liu, Antonios Argyriou, Siwei Ma, Liming Wang, and Zhen Xu. 2020. 360-degree VR Video Watermarking based on Spherical Wavelet Transform. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, 1, Article 1 (January 2020), 23 pages. https://doi.org/10.1145/3425605

*corresponding author.

Authors' addresses: Yanwei Liu, liuyanwei@iie.ac.cn, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China, 10093; Jinxia Liu, Zhejiang Wanli University, Ningbo, China, liujinxia1969@126.com; Antonios Argyriou, University of Thessaly, Volos, Greece; Siwei Ma, Peking University, Beijing, China; Liming Wang, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China; Zhen Xu, Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China.

1551-6857/2020/1-ART1 $15.00
https://doi.org/10.1145/3425605

1 INTRODUCTION

In recent years, advancements in virtual reality (VR) and computer vision technologies have increased the popularity of 360° VR video. Due to its exciting immersion and interactive experience, 360° VR video has been used widely in a variety of applications, such as gaming, virtual exhibition, education, concerts and films [1][2].

At the same time, the available Internet capacity is gradually increasing allowing 360° VR video to be delivered over modern networks [7]. But similar to traditional visual media, different types of illegal techniques can allow access to the delivered 360° VR video, resulting thus in unauthorized copies of the content that can be modified and distributed online. Hence, copyright protection of 360° VR video is an important issue [3] that needs immediate attention. One of the common technical solutions to video copyright is digital watermarking which hides watermarks in digital objects to control copying devices, prove copyright violation and trace unauthorised copies [4]. In the past years, the watermarking technologies mostly aimed at traditional planar images/videos [5].

Traditionally, digital watermarking [4] is the go-to method for copyright protection of visual media. As the visual media forms evolved, multimedia watermarking techniques also evolved to accommodate them. Classical watermarking technology for traditional planar image/video has been studied for many years, leading to several watermarking schemes in the spatial domain [13][14], transform domain [15][16][17], and compression domain [18][19]. Of course these schemes are only efficient for conventional planar image/video. One of the biggest challenges in watermarking is the trade-off between watermarking robustness and imperceptibility [20]. Besides selecting the appropriate host domain to guarantee robustness, watermarking schemes also optimize the perceptual quality of the watermarked video. These watermarking schemes use perceptual models [21] [22] to select the appropriate host positions for the watermark. Obviously, these planar perceptual models that are used in traditional watermarking approaches are no longer suitable for the HMD-based 360° VR video viewing environment.

As visual media evolved towards 3D video and free-view video, watermarking has adapted to 3D video representations. For 3D model representation, the 3D object and 3D mesh watermarking approaches have been proposed. By hiding information in the texture of the object, watermarking for 3D video object was proposed in [43]. This approach can efficiently protect 2D view representations of a 3D object, depending on the accuracy of projective registration of the 2D view. In a pioneering work on 3D mesh watermarking [45], the spherical wavelet transform was used to decompose the original 3D mesh into a series of details at different scales for watermark embedding. Utilizing the disparity-coherence, the blind detection approach was proposed for stereo video watermarking [42]. Watermarking on the depth-image-based rendered 3D image that has been proposed is based on the watermark pattern warping feature [23]. To enhance the robustness of watermarking for depth-image rendered 3D image, a blind multiple watermarking scheme was proposed in [24], and the double-tree complex wavelet transform (DT-CWT) domain of the 3D data has also been used for hosting the watermark [25]. Considering the relationship between the multiple views in geometry, watermark synchronization in view dimension was also exploited for free-view 3D image watermarking [26]. One common characteristic of the aforementioned watermarking schemes is that they are based on emerging features of the 3D data representation to enhance watermarking robustness.

More recently by further extending the viewing dimension spatially, 360° VR video has emerged. 360° VR video usually presents users the panoramic views of the scene, and it has

several key differences when compared to the planar video and 3D video. First, the data delivery format of 360° VR video is not invariable. Different projection methods from the spherical data to the planar data result in a variety of candidate 360° VR video forms that are available for distribution. Second, 360° VR video is displayed in an HMD screen by interactively selecting the rendered viewport. Viewport rendering introduces warping distortions to the viewport video data. Third, 360° VR video delivery supports viewport-dependent partial data transmission [2][6] where only the necessary viewport data are delivered to the user end at one moment. From the above discussion it is easy to see that watermarking technologies for planar video or 3D video cannot adapt to the particular features of 360° VR video and consequently they cannot work efficiently for the emerging 360° VR video. Clearly, a new watermarking technique that is suitable for 360° VR video needs to be proposed.

The interesting fact is that there are several options for embedding watermarks in 360° VR video. The equirectangular projection (ERP) (or the other sphere-to-plane projections in Fig. 1 [8]), the sphere projection and viewport projection, can all be used to hide the watermark. Among them, the planar ERP image is compatible with many traditional video watermarking approaches. Hence, the watermark can be directly embedded in this data domain. By considering the large spatial resolution of 360° VR image, Miura et al. in [27] proposed a DCT-based data hiding technique. This approach is similar to the traditional one that is used for planar media. By utilizing the Scale Invariant Feature Transform (SIFT), Kang et al. in [28] and [48] proposed a viewport image watermark detection approach in the DCT domain for a panoramic image. To form the viewport, the ERP image needs to be projected first to the sphere and then to the viewport. This process includes several projections that introduce image warping distortion which subsequently affects the robustness of the embedded watermark when it is detected from the viewport. Baldoni et al. in [44] proposed a ERP image watermarking approach with a simple pre-processing procedure to partly avoid the distorted area for watermark embedding. Though the above approaches aimed for 360° image watermarking, they actually embedded the watermarks in the planar ERP image and neglected the projection distortion effect on the watermarking robustness during the conversion between multiple projection formats.

As another alternative, the viewport can also host the watermark. However, the viewport is generally rendered from the spherical 360° VR video depending on the human's head movements. Hence, the viewport movement trajectory is difficult to be captured in advance and the direct watermarking in the viewport is extremely challenging.

Thus, the last data domain that can be used for effective watermarking is the spherical data domain. It provides the originally captured high-fidelity data that naturally forms an intermediate data format for different geometry projections as well as the viewport rendering. Correspondingly, spherical domain watermarking can alleviate the impact of several geometry projections on watermarking robustness when the watermark is detected from the viewport. Hence, we believe it is the most suitable watermarking domain.

Our proposed scheme makes use of the developments in wavelet theory on sphere settings [9], a research area that has been developed for processing the spherical data in geographical information systems [10], and computer vision [11]. Since the wavelet transform on the sphere is dedicated to fitting for the spherical geometry, it is natural to be used to process band-limited image signals on the sphere. In particular the directional scale-discretized wavelet transform [12] can exactly capture all the information of the signal and subsequently reconstruct the initial signal at floating-point precision. This observation motivated us *to propose hiding the watermark in the spherical transform domain of the 360° VR video data.*

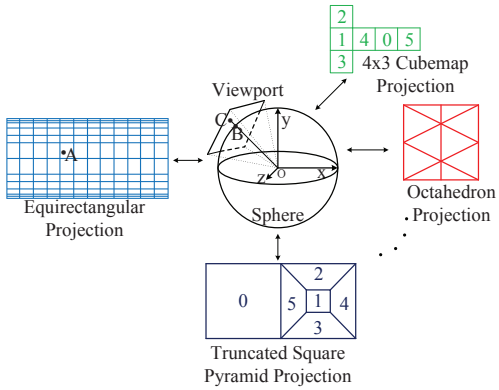


Fig. 1. Multiple projection formats of 360° VR video (In some projections, the number denotes the index of faces in the frame packing)

As with traditional watermarking schemes in the transform domain, this approach is robust, stable, and imperceptible. Hence, *in this paper we propose a novel scheme for watermarking on the spherical data domain of 360° VR video, that to the best of our knowledge, is the first work in this area.* The contributions are summarized as follows.

First, we propose to embed the watermark in the spherical domain for the 360° VR video to fully exploit the invisible nature of the spherical geometry. Spherical projection takes place during format conversion so that the proposed spherical watermarking does not depend on the specific representation of the 360° VR video. Furthermore, we select the spherical wavelet transform (SWT) domain of 360° VR video as the watermark carrier to exploit the robustness of the transform domain watermarking to common attacks.

Second, we design a viewport-oriented just noticeable difference (JND) model in the spherical wavelet domain to characterize the spatial frequency sensitivity of the human visual system (HVS) to the specific HMD. For the luminance channel, the spatial JND of the spherical video after magnifying display in the HMD is first obtained by using the wavelet subband decomposition to mimic the multiple channel models of the HVS. After that, the JND in the spherical wavelet transform domain is modeled by estimating the size of the wavelet coefficient that produced the detected spatial JND, and next, it is used to control the imperceptibility of the watermark on the viewport.

Third, we propose to use the normalized spherical cross-correlation as the robustness metric to detect the watermark in spherical domain. Based on this, a non-blind robust watermark detection approach for the viewport video is also proposed. Hence, the proposed scheme protects not only the original VR source video but also its corresponding viewport data. Consequently, the proposed watermarking scheme is robust to the viewport-dependent 360° VR video transmission.

The rest of the paper is organized as follows. Spherical wavelet transform is introduced in Section 2. The detailed spherical wavelet watermarking scheme, including the watermark embedding and detecting flowchart, is described in Section 3. Section 4 provides the experimental results and finally Section 5 concludes the paper.

2 SPHERICAL WAVELET TRANSFORM

For image processing tasks, traditional Euclidean wavelets are constructed on planar images. However, due to the particular spherical nature of 360° VR video, the wavelet on the sphere is needed to exploit precisely the frequency information of the 360° VR video. By extending

Euclidean wavelet analysis to spherical space, the planar Euclidean wavelets were converted to spherical ones through a stereographic projection in [29]. In this paper, we utilize the directional scale-discretized wavelet [12] on the sphere to transform the 360° VR video to the spectrum space.

2.1 Harmonic analysis on the sphere

Any point ω on the sphere can be defined as $\omega = (\theta, \varphi)$, with latitude $\theta \in [0, \pi]$ and longitude $\varphi \in [0, 2\pi)$. For the square integrable signals in $\mathrm{L}^2(\mathbb{S}^2)$ on the two-dimensional sphere $\mathbb{S}^2$, the spherical harmonics $Y_{l,m}(\omega)$ form an orthonormal basis of $\mathrm{L}^2(\mathbb{S}^2)$, with $l \in \mathbb{N}$, $m \in \mathbb{Z}$ and $|m| < l$. In terms of the Legendre polynomials $P_l^m(\cos\theta)$ and the complex exponentials $e^{im\varphi}$, $Y_{l,m}(\omega)$ is given as

$$Y_{l,m}(\theta, \varphi) = \left[\frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m(\cos(\theta))\, e^{im\varphi}. \tag{1}$$

The spherical harmonic decomposition of a square integrable signal $f \in \mathrm{L}^2(\mathbb{S}^2)$ is given as a linear combination of spherical harmonics

$$f(\omega) = \sum_{l \in \mathbb{N}} \sum_{|m|<l} \hat{f}_{l,m} Y_{l,m}(\omega), \tag{2}$$

where the harmonic coefficients are given as

$$\hat{f}_{l,m} = \int_{\mathbb{S}^2} Y_{l,m}^*(\omega) f(\omega) d\Omega(\omega), \tag{3}$$

with the surface element $d\Omega(\omega) = \sin\theta d\theta d\varphi$, where * denotes complex conjugation.

2.2 Spherical wavelet analysis and synthesis

The directional scale-discretized wavelet transform supports the analysis of oriented and spatially localized, scale-dependent features in signals on the sphere. The scale-discretized wavelet transform of a function $f \in \mathrm{L}^2(\mathbb{S}^2)$ on the sphere is given by the directional convolution of f with wavelet $\Psi^j \in \mathrm{L}^2(\mathbb{S}^2)$. The jth scale wavelet coefficient $W^{\Psi^j} \in \mathrm{L}^2(\mathrm{SO}(3))$ is

$$W^{\Psi^j}(\zeta) \equiv (f \otimes \Psi^j)(\zeta) = \int_{\mathbb{S}^2} d\Omega(\omega) f(\omega) (\Re_\zeta \Psi^j)^*(\omega), \tag{4}$$

where $\Re_\zeta \Psi^j \equiv \Psi^j(\mathrm{R}_\zeta^{-1} \cdot \omega)$ denotes the rotated wavelet with three-dimensional rotation matrix R_ζ. Rotation is specified by an Euler angle $\zeta = (\alpha, \beta, \gamma)$ in the three-dimensional rotation group SO(3) [30] with $\alpha \in [0, 2\pi)$, $\beta \in [0, \pi]$, and $\gamma \in [0, 2\pi)$. Based on the zyz Euler convention, the rotation of a physical body in a fixed coordinate system about the z, y and z axes is parameterized by γ, β and α. Eq. (4) probes the directional structure in the signal f, where γ corresponds to the orientation about each point on the sphere $(\theta, \varphi) = (\beta, \alpha)$.

In the harmonic space, the spherical harmonic decomposition of W^{Ψ^j} is given by a weighted product as

$$(W^{\Psi^j})_{m,n}^l = \frac{8\pi^2}{2l+1} \hat{f}_{l,m} \Psi_{l,n}^{j\,*}, \tag{5}$$

where $(W^{\Psi^j})_{m,n}^l =< W^{\Psi^j}, D_{m,n}^{l\,*} >$, $\hat{f}_{l,m} =< f, Y_{l,m} >$, $\Psi_{l,n}^j =< \Psi^j, Y_{l,n} >$, and $< \cdot, \cdot >$ denotes the inner product operation. The Wigner D-functions $D_{m,n}^l \in \mathrm{L}^2(\mathrm{SO}(3))$ with

natural $l \in \mathbb{N}$ and integer $m, n \in \mathbb{Z}$, $|m|, |n| \leq l$ are the matrix elements of the irreducible unitary representation of the rotation group SO(3).

Usually, the wavelet coefficients represent the high-frequency, detail information contained in the signal, and scaling coefficients depict the low-frequency, approximation information in the signal. The scaling coefficients $W^{\Phi}(\omega) \in \mathrm{L}^2(\mathbb{S}^2)$ are obtained from a convolution of f with the scaling function $\Phi \in \mathrm{L}^2(\mathbb{S}^2)$,

$$W^{\Phi}(\omega) \equiv (f \otimes \Phi)(\omega) =< f, \Re_{\omega}\Phi >, \tag{6}$$

where $\Re_{\omega} = \Re_{(\varphi,\theta,0)}$. The harmonic decomposition of the scaling coefficients is

$$W^{\Phi}_{l,m} = \sqrt{\frac{4\pi}{2l+1}} \hat{f}_{l,m} \Phi^*_{l,0} \tag{7}$$

With wavelets and scaling function satisfying an admissibility, the signal f can be synthesized exactly from its wavelet and scaling coefficients by

$$f(\omega) = \int_{\mathbb{S}^2} d\Omega(\omega') W^{\Phi}(\omega')(\Re_{\omega'}\Phi)(\omega) + \sum_{j=J_0}^{J} \int_{\mathrm{SO}(3)} d\vartheta(\rho) W^{\Psi^j}(\rho)(\Re_{\omega'}\Psi^j)(\omega), \tag{8}$$

where $d\vartheta(\rho) = \sin\beta \cdot d\alpha \cdot d\beta \cdot d\gamma$ is the usual invariant measure on SO(3), J_0 and J denote the lowest and highest scale of the wavelet decomposition, respectively. In harmonic space, the wavelet reconstruction is

$$\hat{f}_{l,m} = \sqrt{\frac{4\pi}{2l+1}} W^{\Phi}_{l,m} + \Phi_{l,0} \sum_{j=J_0}^{J} \sum_{n=-l}^{l} (W^{\Psi^j})^l_{m,n} \Psi^j_{l,n} \tag{9}$$

and the admissibility condition [29] that a band-limited signal f can be decomposed and reconstructed exactly is

$$\frac{4\pi}{2l+1} |\Phi_{l,0}|^2 + \frac{8\pi^2}{2l+1} \sum_{j=J_0}^{J} \sum_{m=-l}^{l} \left|\Psi^j_{l,m}\right|^2 = 1, \quad \forall l. \tag{10}$$

Besides the good reconstruction quality, the spherical wavelet transform also shows the excellent localization property [31] in spatial domain. Given any $\zeta \in \mathbb{R}^+_*$, there exist strictly positive constants $C_1, C_2 \in \mathbb{R}^+_*$, make that the directional scale-discretized wavelet $\Psi \in \mathrm{L}^2(\mathbb{S}^2)$, centered on the North pole, satisfies the localization bound

$$|\Psi(\theta, \varphi)| \leq \frac{C_1}{(1 + C_2\theta)^{\zeta}} \tag{11}$$

where $(\theta, \varphi) \in \mathbb{S}^2$ denotes the spherical coordinates with latitude $\theta \in [0, \pi]$ and longitude $\varphi \in [0, 2\pi]$.

3 SPHERICAL WAVELET WATERMARKING SCHEME

The spherical wavelet transform probes spatially localized features in signals on the sphere. This property is desirable for hiding a watermark on the 360° VR video that is spherical shape by nature. By manipulating the VR data on the sphere in SO(3) instead of the Euclidean plane, the spherical domain watermarking can fully exploit the invisibility nature of the spherical geometry space. Furthermore, the spherical wavelet shows an excellent localization property in spatial domain. This property can be naturally exploited to design the watermarking scheme that can resist to the cropping attack that is often used in viewport rendering. Moreover, the wavelet transform on the sphere provides good reconstruction quality by generating the scale-discretized spectrum that is suitable for hosting the watermark.

Taking into account these three advantages, we believe that the spherical wavelet transform is suitable for VR video watermarking.

The proposed spherical wavelet watermarking scheme is based on the idea that the watermark is hidden in the spherical wavelet spectrum domain of the VR video. Three novel operations that are adapted to 360° VR video are proposed. First, based on the particular characteristic of the spherical VR image that its rotation in the sphere corresponds to the translational motion in the ERP image, we propose to utilize the equal angle rotation for several continuous frames to defend against the inter-frame collusion attacks. Second, by accommodating the viewport viewing feature of 360° VR video, we propose to use viewport-oriented JND to mask the embedded watermark to guarantee its imperceptibility on the rendered viewport. Third, unlike the traditional watermarking approach that computes the normalized cross-correlation in the planar domain, we compute the normalized cross-correlation in the spherical domain that considers the 3D rotations to accurately detect the watermark with a spherical shape.

In this section, we first present the viewport-oriented JND model that is used to control the watermarking strength and then introduce the spherical wavelet watermark embedding and extracting schemes.

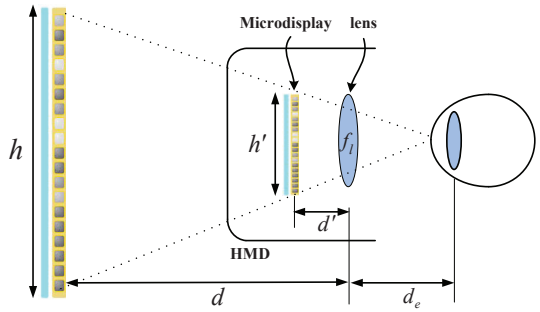


Fig. 2. The optical design for a HMD

3.1 Viewport-oriented JND Model

Usually, 360° VR video is not directly viewed on the screen and it needs to be displayed by rendering a viewport on HMD. Thus, the traditional just noticeable differences (JND) model that is originally dedicated to the planar video cannot work efficiently for 360° VR video. For a VR image, since only a part of VR data are displayed at a moment, the JND needs to be detected in the viewport domain but not in the planar ERP image domain. However, the user's viewport positions are very difficult to be predicted in advance. In contrast, the spherical video is the intermediate format that links all the projections as well as the viewport rendering. In such situation, it is desirable to detect JND on the spherical video. Moreover, the proposed JND model targets watermarking applications and consequently it only focuses on the contrast sensitivity that is introduced by the embedded watermark.

3.1.1 Spatial JND. It is well known that the HVS is sensitive to luminance contrast rather than the absolute luminance value. Thus, the frequency sensitivity of the HVS can also be adopted to capture the perceptual redundancy in the HVS. The discrete wavelet transform (DWT) decomposition of an image provides a representation that mimics the multiple channel models of the HVS [32] and this property offers the potential to predict JND. Similarly,

the spherical wavelet decomposition provides the potential to predict the viewport-oriented JND. To find the visible luminance threshold $\Gamma_y(\upsilon)$ of the uniform noise on an image in a decomposition orientation τ and spatial frequency υ, Watson at al. [33] summarized a spatial frequency threshold model as

$$\log(\Gamma_y(\upsilon)) = \log(a) + \hbar \cdot (\log(\upsilon) - \log(g_\tau \cdot \upsilon_0))^2 \tag{12}$$

where a is a constant (0.495), $\hbar$ the width (0.466), $g_\tau \cdot \upsilon_0$ the minimum of the parabola ($\upsilon_0 = 0.401$ and g_τ is 1.501, 1, and 0.534 for the LL, LH/HL, and HH subbands). Eq. (12) was measured from the psychovisual detection of noise added to the wavelet coefficients. The four filtering orientations were supported in Eq. (12). However, only the scaling subband decompositions are performed for the current SWT. To find the appropriate $g_\tau \cdot \upsilon_0$, we have to establish a correspondence between the scale κ in SWT and the filtering orientation subbands in the DWT. Assume that N_s scaled subbands are obtained by the SWT, and then the LL, LH/HL, and HH subbands in the DWT correspond to $\kappa \leq \lceil N_s/3 \rceil$, $\lceil N_s/3 \rceil < \kappa \leq \lceil 2N_s/3 \rceil$, and $\kappa > \lceil 2N_s/3 \rceil$ in SWT, respectively. In Eq. (12), υ is inferred from the display visual resolution and the corresponding wavelet transform scale as

$$\upsilon = r2^{-\kappa} \text{cycles/degree}, \tag{13}$$

where r is the display visual resolution that is computed as

$$r = d_v \cdot d_r \cdot \tan(\frac{\pi}{180}) \approx d_v \cdot \frac{d_r \cdot \pi}{180} \approx d_v \cdot \frac{d_r}{57.3}. \tag{14}$$

In the above, d_v is the viewing distance in cm and d_r is the display resolution in pixels/cm.

Viewers typically watch 360° VR video using an HMD. The optical design of the HMD introduces a micro-display that is located behind a magnifying lens, as shown in Fig. 2. The distance d' from lens to physical display is slightly smaller than the focal length f_l of lens, such that a magnified virtual image with size of h from a micro-displayed image with size of h' is optically created at a larger distance d. With the Gaussian thin lens formula, the magnification M can be derived as

$$M = \frac{f_l}{f_l - d'} \tag{15}$$

Furthermore, the display visual resolution for an HMD is computed as

$$r = d_v \cdot \frac{d_r}{57.3} = (d + d_e) \cdot \frac{d_r}{57.3} \cdot \frac{1}{M} \tag{16}$$

The embedded watermark information can be taken as a form of quantization error on the image. The visibility of errors is mainly affected by the frequency sensitivity perceived by HVS. The JND value from Eq.(12) is a spatial frequency threshold [33]. Consequently, the JND value that measures the visibility of watermark-induced errors, indicates the strength of embedded watermark signals. The JND model in Eq.(12) is originally designed for measuring the visibility of the uniform quantization errors. Even though the watermarking information in the image is usually not uniformly distributed, its subband can be approximately regarded as a uniform distribution [21] and further the watermark is approximately taken as a complex visual masking result. Thus, the watermarking bits on the image can be assumed as several sets of uniform errors over the discrete subbands. Since the variance within a subband characterizes the spread of errors, it can be utilized to scale the JND in each subband for approximating the uniform distribution of watermark-induced errors.

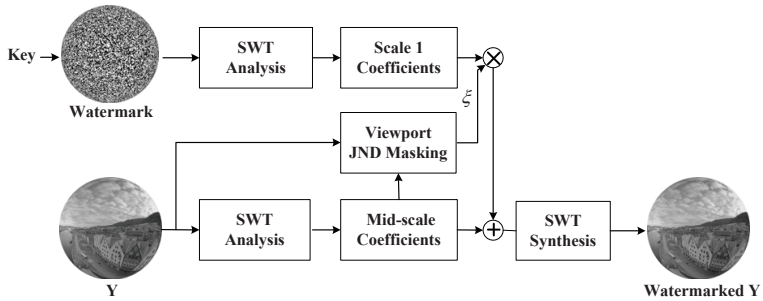


Fig. 3. Watermark embedding in the SWT domain

Based on the measurement of variance within a subband, a simple adjustment is used to scale the JND in that subband,

$$\Gamma'(\upsilon) \approx \Gamma(\upsilon) \cdot (1 + \frac{\sigma^2(\upsilon)}{(\Gamma(\upsilon))^2})^{\frac{1}{2}} \tag{17}$$

where $\sigma^2(\upsilon)$ is the variance within the watermarking region in luminance channel of VR image.

3.1.2 SWT domain JND. The watermark is hidden into the wavelet coefficients. To control the watermarking strength in SWT, the spatial domain JND needs to be converted to the SWT domain JND. Thus, we need to estimate the size of wavelet coefficient error that produced the spatial JND. The spatial JND is expressed as the peak amplitude of the signal detected by an observer and it relates to the peak of the spatial (impulse) response when the wavelet pyramid is reconstructed [34]. In the wavelet transform, the wavelet decomposition combines the low-pass and high-pass synthesis filtering to generate different scales of wavelet coefficients. For the worst case in the wavelet synthesis filtering, the spatial JND corresponds to a combination of maximum coefficient amplitudes for low-pass and high-pass filters. Considering that different filtering coefficient amplitudes are used in different subband of wavelet decomposition, the "worst case" conversion formula [34] for estimating the SWT domain JND is derived as

$$\Pi(\upsilon) = \frac{\Gamma'(\upsilon)}{i_\tau \cdot p_{l_w}^{(l_w-1)}} \tag{18}$$

where l_w is either 1, 2, 3 that specifies the subband and i_τ is either $p_{l_w}^2$, $p_{l_w} \times p_{h_w}$ and $p_{h_w}^2$ for the subbands corresponding to $\kappa \leq \lceil N_s/3 \rceil$, $\lceil N_s/3 \rceil < \kappa \leq \lceil 2N_s/3 \rceil$, and $\kappa > \lceil 2N_s/3 \rceil$, respectively. $p_{l_w} = 0.788845$ and $p_{h_w} = 0.852699$ [34] are the maximum coefficient amplitudes for low-pass and high-pass synthesis filters in the wavelet decomposition, respectively.

3.2 Spherical Watermark Embedding

The flowchart of of proposed watermark embedding scheme is shown in Fig. 3. The watermark I_w is a 2D array that consists of values in $\{+1, -1\}$ generated by a pseudo-random sequence generator where the seed represents the secret key K. By regulating K, I_w can be repeated for t_k continuous frames. To avoid the loss of watermarking strength that brought by the direct embedding of bipolar watermark into the transform domain of the host image, the watermark I_w is first transformed into the SWT coefficients that are then embedded into the SWT spectrum of the host image.

Note that the 360° videos are mostly available in large spatial resolutions so that the viewport-dependent random access to the image data is necessary. To deal with the viewport-dependent partial data transmission, the watermarks need to cover the whole image spatially. We propose to divide the SWT coefficients of the host image into many longitude-latitude grids that they have the same size. We call these grids tiles. The tiling process of SWT coefficients is same with that in [6]. Each tile embeds one watermark and the watermarks are the same in all the watermarked tiles of each frame.

To enhance the watermarking robustness against temporal filtering, we apply the rotation operation in the sphere to regulate the watermark embedding positions. One temporal watermarking group consists of $t_g = 6$ continuous frames and each frame rotates 60° around the south-north axis compared to the previous frame. The watermark is embedded in the SWT domain of the luminance component of the rotated 360° image. After watermark embedding, the cover image that contains the watermark will be inversely rotated to recover the frame. Thus, coupling the temporally various watermarking positions for continuous t_g frames with the duplicate watermark content for continuous t_k frames can keep a strong resistance to the temporal synchronization attack and the watermark estimation attack.

To invisibly embed the watermark that can survive the lossy compression, the middle-high frequency is selected for hosting the watermark. If N_s scales are obtained for the host image after SWT, the middle-high subband $ms = \lceil (N_s + 1)/2 \rceil$ is used to host the watermark. For watermarking, the planar luminance data for the host image is first projected to the spherical luminance data. The spherical data is then analyzed with a directional N_s-scales SWT. Assume the size of the VR image is $\mathrm{W} \times \mathrm{H}$ and the tile size in planar luminance data is $\frac{\mathrm{W}}{m_w} \times \frac{\mathrm{H}}{n_w}$, where m_w and n_w are the tile numbers in the horizontal and vertical directions, respectively. For watermarking, the watermark is embedded in a block that is centered on each tile. The watermark block size can be less than or equal to the tile size. In this paper, we take the watermark size to be same as the tile size.

For the luminance component Y, the middle-high scale subband coefficient matrix $\mathbf{F}^{ms,\,o}_{wav\_y}$ with four orientations $o = 1, 2, 3, 4$ that represent 0°, 90°, 180° and 270° is used to host the watermark signal. To let the watermark cover the whole image, the transform coefficient matrix $\mathbf{W}_{wav}$ of the watermark I_w after a one-scale directional SWT is duplicated and then embedded into each tile in $\mathbf{F}^{ms,o}_{wav\_y}$. To control the imperceptibility of the embedded watermark, we propose to use visual masking to process the transformed watermark data. The visual mask matrix $\mathbf{M}^{ms,o}_{wav\_y}$ is generated as,

$$\mathbf{M}^{ms,o}_{wav\_y} = \left|\mathbf{F}^{ms,o}_{wav\_y}\right| / \eta \tag{19}$$

where η denotes a factor that regulates the magnitude of the masking coefficients.

By masking the watermark signal, the watermarked transform coefficient matrix $\tilde{\mathbf{F}}^{ms,o}_{wav\_y}$ is written as

$$\tilde{\mathbf{F}}^{ms,o}_{wav\_y} = \mathbf{F}^{ms,o}_{wav\_y} + \xi \times (\mathbf{M}^{ms,o}_{wav\_y} \bullet \mathbf{W}) \tag{20}$$

where the symbol $\bullet$ denotes an element-wise matrix multiplication, $\mathbf{W}$ is a matrix that consists of the element $\mathbf{W}_{wav}$, and parameter ξ is used to tune the watermarking strength under the help of viewport-oriented JND model. Assume the element in the ith row and jth column in $\mathbf{M}^{ms,o}_{wav\_y} \bullet \mathbf{W}$ is $\Omega^{ms,o}_{wav\_y}\big|_{i,j}$. Based on the proposed JND model, $\Omega^{ms,o}_{wav\_y}\big|_{i,j}$ is obtained as

$$\Omega^{ms,o}_{wav\_y}\big|_{i,j} = \begin{cases} \Pi_y(\upsilon)/\xi, & \text{if } \xi \cdot \Omega^{ms,o}_{wav\_y}\big|_{i,j} > \Pi_y(\upsilon) \\ \Omega^{ms,o}_{wav\_y}\big|_{i,j}, & \text{otherwise} \end{cases} \tag{21}$$

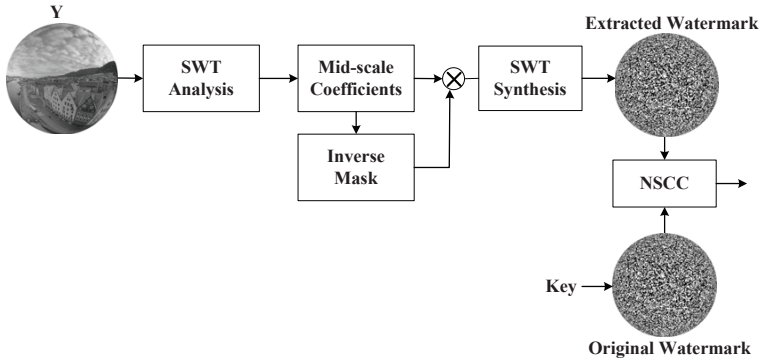


Fig. 4. The watermark detection in spherical domain

where $\Pi_y(\upsilon)$ is the JND threshold in viewport domain. It is computed from the viewport-oriented JND model in Eq.(18). Finally, $\tilde{\mathbf{F}}^{ms,o}_{wav\_y}$ is inversely transformed to reconstruct the watermarked luminance signal. The synthesized luminance component after inverse rotation is next combined with the chrominance components to construct the complete 360° VR video.

3.3 Watermark Detection

3.3.1 Watermark detection on the sphere. The flowchart of spherical domain watermark detection is shown in Fig. 4. For watermark detection, only the key for generating the original watermark is available. After the SWT transform of the luminance component, the middle-high scale coefficient matrixes $\breve{\mathbf{F}}^{ms,o}_{wav\_y}$ $(o = 1, 2, 3, 4)$ is constructed. Based on the visual masking operation in Eq. (19), the inverse masking matrix for luminance component is constructed as

$$\breve{\mathbf{M}}^{ms,o}_{wav\_y} = \begin{bmatrix} \frac{1}{\tilde{\mathbf{M}}^{ms,o}_{wav\_y}(0,0)} & \cdots & \frac{1}{\tilde{\mathbf{M}}^{ms,o}_{wav\_y}(0,\mathrm{W})} \\ \vdots & \ddots & \vdots \\ \frac{1}{\tilde{\mathbf{M}}^{ms,o}_{wav\_y}(\mathrm{H},0)} & \cdots & \frac{1}{\tilde{\mathbf{M}}^{ms,o}_{wav\_y}(\mathrm{H},\mathrm{W})} \end{bmatrix} \tag{22}$$

where $\tilde{\mathbf{M}}^{ms,o}_{wav\_y}$ is the visual mask matrix that operated on the $\breve{\mathbf{F}}^{ms,o}_{wav\_y}$ $(o = 1, 2, 3, 4)$. Correspondingly, the extracted transform coefficients form the watermark matrix $\breve{\breve{\mathbf{W}}}_y$ with size of $m_w \times n_w$,

$$\breve{\breve{\mathbf{W}}}_y = \breve{\mathbf{M}}^{ms,o}_{wav\_y} \bullet \breve{\mathbf{F}}^{ms,o}_{wav\_y} \tag{23}$$

From $\breve{\breve{\mathbf{W}}}_y$, the embedded watermarks $\breve{\breve{\mathbf{W}}}_{wav\_y}$ in the tiles are extracted. By performing the spherical wavelet synthesis transform for $\breve{\breve{\mathbf{W}}}_{wav\_y}$, the extracted watermark $\breve{I}_{w\_y}$ is constructed. Because $\breve{I}_{w\_y}$ is still the spherical signal, we compute the normalized spherical cross-correlation (NSCC) between the extracted watermark $\breve{I}_{w\_y}$ and the original watermark I_w for each frame to judge whether the watermark is present. The computation of cross-correlation in the spherical domain, when compared to that in the ERP plane, can mitigate the effect of distortion that is introduced by the projection from the sphere to ERP plane. Based on the 3D rotation operation $\Lambda(\mathrm{R})$ in SO(3) with $\mathrm{R}(\alpha, \beta, \gamma) \in \mathrm{SO}(3)$, the normalized

spherical cross-correlation between reference image signal $\breve{I}_{w\_y}(\omega)$ and template signal $I_w(\omega)$ is defined as [35]

$$\mathrm{NSCC(R)} = \frac{\int_{\mathbb{S}^2} (\breve{I}_{w\_y}(\omega) - \bar{I}_{w\_y}(\omega))\Lambda(\mathrm{R})(I_w(\omega) - \bar{I}_w(\omega))d\omega}{\sqrt{\int_{\mathbb{W}} \left[\breve{I}_{w\_y}(\omega) - \bar{I}_{w\_y}(\omega)\right]^2 d\omega \int_{\mathbb{W}} \left[I_w(\omega) - \bar{I}_w(\omega)\right]^2 d\omega}} \tag{24}$$

where $\mathbb{W}$ is the image window defined by the support of the template, and $\bar{I}_{w\_y}(\omega)$ and $\bar{I}_w(\omega)$ denote the mean values of the signals $\breve{I}_{w\_y}(\omega)$ and $I_w(\omega)$ in $\mathbb{W}$, respectively.

Since $\int_{\mathbb{S}^2} \Lambda(\mathrm{R})(I_w(\omega) - \bar{I}_w(\omega))d\omega = 0$ and $\bar{I}_{w\_y}(\omega)$ is also a constant, the numerator in Eq. (24) is written as

$$\int_{\mathbb{S}^2} \breve{I}_{w\_y}(\omega)\Lambda(\mathrm{R})(I_w(\omega) - \bar{I}_w(\omega))d\omega \tag{25}$$

The integral in the denominator of Eq. (24) is computed over the support window $\mathbb{W}$. By using a support mask $p_{\mathrm{R}}(\omega) = \Lambda(\mathrm{R})p(\omega)$ that rotates with the template signal, the integral over $\mathbb{W}$ can be transferred to the sphere $\mathbb{S}^2$ as

$$\begin{aligned}
&\int_{\mathbb{W}} \left[\breve{I}_{w_y}(\omega) - \bar{I}_{w_y}(\omega)\right]^2 d\omega \\
&= \int_{\mathbb{S}^2} p_{\mathrm{R}}(\omega)\left[\breve{I}_{w_y}(\omega) - \bar{I}_{w_y}(\omega)\right]^2 d\omega \\
&= \int_{\mathbb{S}^2} p_{\mathrm{R}}(\omega)(\breve{I}_{w_y}(\omega))^2 d\omega - 2\bar{I}_{w_y}(\omega) \int_{\mathbb{S}^2} p_{\mathrm{R}}(\omega)\breve{I}_{w_y}(\omega)d\omega + (\bar{I}_{w_y}(\omega))^2 \int_{\mathbb{S}^2} p_{\mathrm{R}}(\omega)d\omega
\end{aligned} \tag{26}$$

With the constant integral of the support mask, the mean value of the reference image $\bar{I}_{w\_y}(\omega)$ can be computed as

$$\bar{I}_{w\_y}(\omega) = \frac{\int_{\mathbb{S}^2} p_{\mathrm{R}}(\omega)\breve{I}_{w\_y}(\omega)d\omega}{\int_{\mathbb{S}^2} p(\omega)d\omega} \tag{27}$$

Thus, NSCC(R) in Eq. (24) can be computed by the integrals over $\mathbb{S}^2$.

After finishing the spherical correlation computation, we proceed to project the NSCC to the plane by using the stereographic projection. Thus, the NSCC can be converted to a planar form as the common normalized cross-correlation (NCC). Theoretically, the position of the peak value of the computed planar NSCC approximately locates at central position $(\bar{x}, \bar{y})$ in the image tile for embedding watermark. Assume that the peak value of the planar NSCC for the ith frame is T_p, its actual position is (x_i, y_i), and the Euclidean distance between (x_i, y_i) and $(\bar{x}, \bar{y})$ is $d_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}$. Normally, d_i is equal to zero. When the position of the planar NSCC peak is not accurate for a certain frame, d_i will be more than zero. We set a threshold value T_h to identify the watermark. T_h was set to 0.1 to guarantee a probability of false detection lower than 10^{-6} based on the approach in [36]. If d_i is less than an empirical value of 10 and $T_p > T_h$, the frame that is currently detected is identified with the watermark. Here the threshold value d_i is selected as 10 after the careful studies in several watermarking experiments. Once fifty percent of the frames in one video sequence are identified with the watermark, the video sequence will be considered to contain the watermark.

3.3.2 Watermark detection from viewport. During 360° VR video playback, the viewport is rendered in real-time. Usually, the viewport data is not allowed to be saved for redistribution. However, in some cases, the viewport data may be used without authorization. To avoid

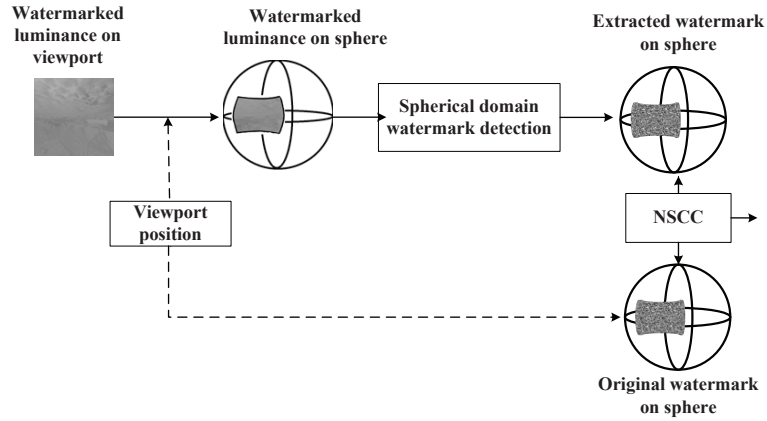


Fig. 5. Watermark detection from the viewport

the illegal distribution of the viewport data, the proposed watermarking scheme is capable of detecting the watermark from the viewport.

The flowchart of the proposed watermark detection scheme from the viewport is shown in Fig. 5. During detection, besides a key for generating the original watermark, the original 360° video is also available for locating the position of the watermarked viewport in the sphere. In the general case, the viewport needs to be inversely projected to the sphere to detect the watermark since the watermark is embedded in the spherical domain. For the inverse projection, the exact viewport position on the sphere needs to be known in advance and this can be finished by registering the viewport data on the sphere. For viewport registration, we use the "Planar vs. Omni" matching based on SIFT on the sphere [37] to find the exact viewport position in the watermarked image. The spherical SIFT can efficiently capture the local features in spherical coordinates and further find the viewport position accurately by using the feature point matching. Since the feature point matching is performed on the sphere, it generally requires several minutes to find the accurate viewport position. After the viewport data is projected on the sphere, the spherical domain watermarking detection approach is used to find the embedded watermark. Finally, the NSCC between the original watermark and the extracted one is computed. If the peak position of the planar NSCC deviates from the correct position by no more than an empirical Euclidean distance of 10, the watermark for this viewport frame is considered to be present. If fifty percent of the frames in one viewport sequence are predicted to contain the watermark, the entire viewport video is considered as watermarked video.

4 EXPERIMENTAL RESULTS

The proposed watermarking scheme was tested on a group of 360° videos under different types of attacks. The 360° VR video dataset included fifteen test sequences containing KiteFlite, Harbor, Trolley, Gaslamp, and AerialCity in MPEG [38]. These videos that contained different scenes were split into sixty short clips with a duration of 4 seconds (120 frames) for each in watermarking experiments. The spatial resolution of the sequences is 4096×2048 while the viewport covered 110° horizontal and 90° vertical field of view.

In our experiments, parameters d_r, d and d_e that are needed for deriving viewport-oriented JND were set to 240pixels/cm, 156cm and 1.8cm, respectively. For watermarking, the overall number of scales N_s in the SWT decomposition was set to 3 and the scaling factor η in Eq.

(a) Watermarked image and viewport by SWT-w-JND scheme

(b) Watermarked image and viewport by SWT-w/o-JND scheme

Fig. 6. Sample images for different watermarking schemes.(The left is the ERP image, the middle is the viewport image, and the right is enlarged picture for the region covered by red rectangle)

(19) was set to 2 and 100 for watermark embedding and detection, respectively. m_w and n_w were both set to 8. ξ was set to 0.1.

Until now, the existing VR video watermarking schemes are all performed in the planar domain and our proposed watermarking scheme is the first work that performs this task in the spherical domain. To evaluate the performance of the proposed spherical domain watermarking, we compared it with the state-of-the-art DCT-based [28] and DWT-based [44] planar ERP-domain watermarking scheme. They are currently the only two available 360° image watermarking schemes. Similar to the proposed 360° video watermarking scheme, the texture-based 3D watermarking scheme [43] also demanded a watermark detection from the 2D view of 3D object, and hence we also selected it as a baseline of VR video watermarking to evaluate the performance of our proposed scheme.

4.1 Watermark Transparency

Watermarking adds redundant bits into the 360° VR video and inevitably introduces pixel variations on the image. However, the HVS cannot always sense every pixel variation due to its near-threshold properties. In this subsection, we evaluate the imperceptibility of the added watermark via subjective viewport quality assessment. During the evaluation, the watermarked and original 360° VR videos are viewed by a HTC vive focus HMD.

The Double Stimulus Impairment Scale (DSIS) method for 360° VR video in [39] was used for evaluation. In DSIS, the raw reference video was presented first, and after a three-second mid-grey display, the watermarked video followed. The specific test procedures followed the specification in ITU-RBT.500 [40]. In these tests, the viewports were evaluated by a total of 30 subjects (23 male, 7 female, ages 17 to 42). Subjects were asked to rate the watermarking-induced degradations on a five-point scale (5: Imperceptible, 4: Perceptible but not annoying, 3: Slightly annoying, 2: Annoying, 1: Very annoying). During the tests,

Table 1. Subjective viewport quality rating

Sequence	Watermarked viewport for SWT-w-JND			Watermarked viewport for SWT-w/o-JND		
	MOS	CI	PSNR(dB)	MOS	CI	PSNR(dB)
KiteFlite	4.53	0.09	43.35	4.11	0.11	42.38
Harbor	4.31	0.12	42.84	3.89	0.10	41.47
Trolley	4.45	0.11	43.29	4.03	0.08	41.45
Gaslamp	4.42	0.13	43.17	3.92	0.14	40.64
AerialCity	4.34	0.13	42.93	3.97	0.09	40.66
Average	4.41	0.12	43.12	3.99	0.10	41.32

subjects were instructed to freely explore the 360° VR video scene. The mean opinion scores (MOS) for the subjective quality evaluation tests with 95% confidence interval (CI) and the objective PSNR values are summarized in Table 1.

The average MOS value in Table 1 is higher than 4 for the proposed SWT-based watermarking scheme with viewport-oriented JND model (let's abbreviate it to SWT-w-JND) and it provides a larger value than the scheme without viewport-oriented JND model (let's abbreviate it to SWT-w/o-JND). The PSNR values for the ERP images in Table 1 also verify this observation. It shows that the proposed SWT-w-JND watermarking scheme can efficiently hide some information without damaging the perceptual quality of the viewport. There are two reasons for this. On the one hand, the modification effects of coefficients on the frequency domain by watermarking are spread over a large image area in the spatial domain. On the other hand, the JND model controls the watermark embedding strength efficiently based on the perceptual viewport quality metric.

For a visually intuitive insight into the result, we offer the viewport images and ERP images for the KiteFlite video in Fig. 6 for different watermarking schemes. In Fig. 6 (b), a few noising artifacts (for example, the artifacts near the palisade) can be perceived in the watermarked viewport for SWT-w/o-JND scheme as shown in the enlarged picture. The visual comparison among the viewport images illustrates the effectiveness of the proposed JND model in viewport watermark transparency.

The watermark visibility is also related to the embedding strength. We measured the perceptual qualities for different values of watermark strength parameter ξ and the results are shown in Fig. 7(a). The MOS value is the average result over all the test clips. It can be seen from Fig. 7(a) that the SWT-w-JND scheme has always a higher perceptual quality than the SWT-w/o-JND scheme. When the value of ξ is less than 1, the MOS values for the SWT-w-JND scheme are above 3.0. When the value of ξ is larger than 1, the MOS value for SWT-w/o-JND scheme deteriorates rapidly (less than 3.0,) while the MOS value for SWT-w-JND scheme is still above 3.0. This indicates that the watermark for SWT-w/o-JND scheme will be visible when the value of ξ is larger than 1. Fig. 7(b) shows the NSCC values for the increasing ξ. In comparison with Fig. 7(a), when the value of ξ is less than 0.1, the two schemes have relatively small NSCC values despite the large MOS values in Fig. 7(b). Thus, the trade-off between the watermarking invisibility and detection robustness is achieved in the vicinity of 0.1 for ξ.

4.2 Viewport-dependent watermark detection

In the experiments, we tested all the sequences for watermark detection from the viewport. The viewports were randomly selected. In the watermark detection, the viewport position estimated by the spherical SIFT sometimes deviates a bit from the correct one. Because the original viewport watermark pattern also uses the same position as the estimated one, this small deviation of the viewport position will be tolerated when seeking the peak value

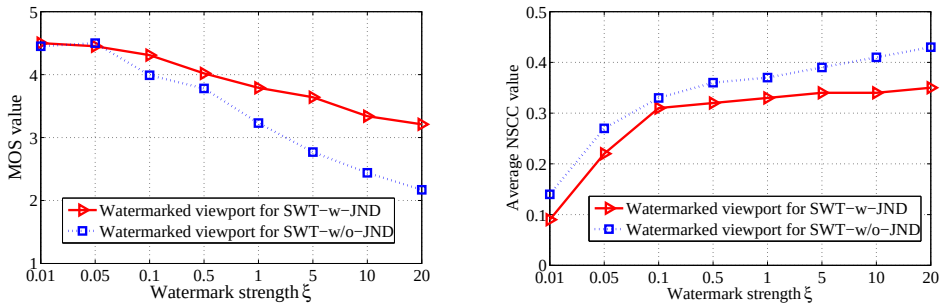


Fig. 7. The perceptual qualities and NSCC values for the increasing ξ

for NSCC computation. Thus, the small error of the viewport position has almost no effect on the robustness of the viewport watermark detection.

Fig. 8 shows the average NCC and planar NSCC peak value over all the frames for viewport watermark detection. It can be seen from Fig. 8 that when the watermark is detected from the viewport, spherical domain watermarking has higher watermark strength than the DCT-based and DWT-based ERP-domain watermarking. Due to the additional pre-processing procedure to avoid embedding the watermark into the area with large spherical projection distortion, the DWT-based ERP-domain watermarking obtains higher watermark strength than the DCT-based one. When compared with spherical domain watermarking, viewport watermark detection in the ERP-domain watermarking requires more forward and backward projections from viewport to ERP plane. This introduces larger projection-induced distortion on the video, while it also leads to a negative effect on the performance of viewport watermark detection.

During ERP to sphere projection, the video distortions at the two poles are larger than those at the equator of the sphere. In Fig. 8, we notice that the gap between the planar NSCC of the SWT-based scheme and the NCC of ERP-domain scheme (either the DCT-based or the DWT-based scheme) for the Harbor sequence is larger than that of the other sequence. After analyzing the viewport position data, we found that more viewports at the two poles in the sphere were selected for the Harbor sequence which led to this larger gap.

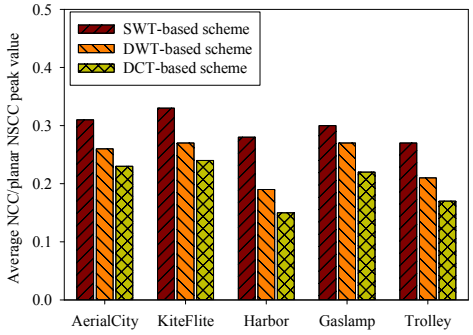
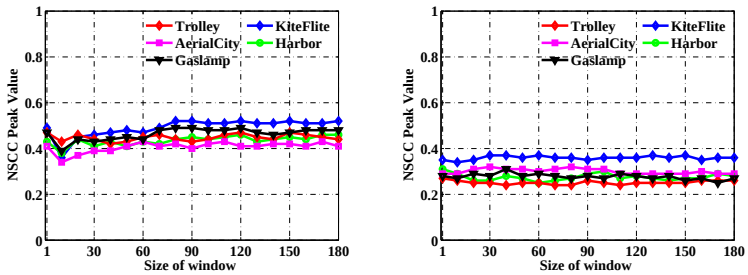


Fig. 8. The average NCC (or planar NSCC) peak value for viewport watermark detection

4.3 Security analysis

Video watermarking is vulnerable to the inter-frame watermark attacks. Watermark estimation remodulation (WER) [46], and temporal frame averaging (TFA) [47] are both typical inter-frame attack approaches for removing video watermarks. In this subsection we evaluate the performance of the proposed watermarking scheme in presence of the above attacks. Since the watermark can be detected in either ERP-domain or viewport-domain of VR video, these domains are also assumed as the targets of attacks. In the WER attack, the difference between a frame and its low-pass filtered version was computed as the rough estimate of the watermark in one frame, and after that the refined estimate of the watermark was obtained by averaging the rough estimations of the watermark extracted from a large number of consecutive frames that constituted a temporal window. Regarding the TFA attack, the low-pass filtering and temporal averaging were used to process the watermarked frames over a temporal window.

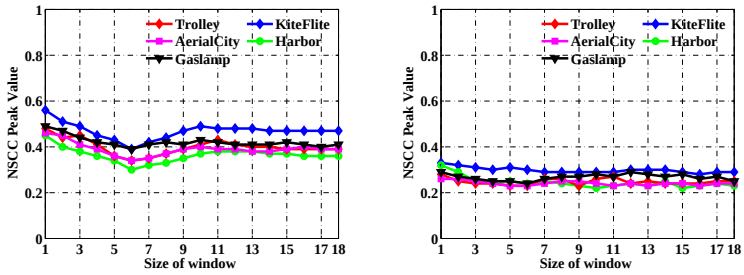
The NSCC peak values for detecting watermarks of several VR video clips after WER attack over different sizes of temporal window are shown in Fig. 9. It can be seen from Fig. 9(a) that initially the NSCC peak values decrease slowly with the enlarged size of window, but from the 12th frame, the NSCC peak values gradually increase and finally keep approximately constant over the increased size of window. In Fig. 9(a), the average NSCC peak values after ERP-domain attack for all test clips are still larger than 0.4. It is because the same watermark is repeated for the consecutive $t_k = 12$ frames, and the embedding position of the watermark changes frame by frame within successive $t_g = 6$ frames. Consequently, the watermarks embedded at the same spatial positions in different frames are almost uncorrelated within a short period but intermittently correlated within a long period. As a result, the WER attack achieves a very low probability to successfully remove the watermark in ERP-domain. Similarly, due to the dynamic change of viewport in temporal dimension, the watermarks among the consecutive viewport frames are nearly uncorrelated and also the image contents in successive viewport frames are different. This leads to almost the same NSCC peak values with those before attack in Fig.9(b). It indicates that the WER attack to the viewport video is not very useful.



(a) ERP-domain attack (b) Viewport-domain attack

Fig. 9. Average NSCC peak values after WER attacks using different size of window.

Fig.10(a) shows the NSCC peak curves for different sequences after ERP-domain TFA attack. It can be seen from Fig. 10(a) that the NSCC peak values for each video clip gradually decrease from the start until the size of window is 6 and then slightly increase until the size of window is 12. The short-term uptrend of the NSCC peak values in the middle of the curve in Fig10(a) illustrates that the proposed scheme has a light-weight



(a) ERP-domain attack (b) Viewport-domain attack

Fig. 10. Average NSCC peak values after TFA attacks using different size of window.

defense against the TFA attack in ERP-domain. However, the final decline trend of the NSCC peak curve shows that the TFA attack in ERP-domain might be able to extract the watermark. But it should be noted that the attack will introduce the significant visual distortions. As for viewport-domain attack, Fig. 10(b) shows the NSCC peak curves for different VR video sequences after TFA attack. It can be seen from Fig. 10(b) that the NSCC peak values are mostly below 0.3, that is a value slightly less than those before attack. It indicates that the TFA attack is possible to remove the watermark. But it is obviously that the attack will result in a very poor visual quality due to averaging the totally different contents in successive viewport frames.

4.4 Robustness to attacks

4.4.1 Compression and scaling attack. Since the proposed approach hides the information into the raw signal, compression and spatial scaling can both be used to attack the watermark. We simulated the scaling attacks that downscale the 360° VR video from the original resolution to 3072×1536, 2048×1024 and 1536×768. The false negative rates (FNRs) under the scaling attacks are summarized in Table 2. The results are averaged over all the tested 360° videos. From Table 2, it can be seen that the proposed watermarking scheme and the ERP-domain (DCT-based and DWT-based) watermarking schemes can all successfully detect watermarks in their embedded domain under the attacks with scaling ratio exceeding 0.5. However, when detecting the watermark from the viewport, the proposed watermarking scheme presents significantly smaller detection errors than the DCT-based and DWT-based schemes for all the resolutions.

Table 2. FNR (%) of watermark detection for downscaling attacks

Scaling ratio	Detection from 360° data			Detection from viewport		
——	Proposed scheme	DCT-based scheme	DWT-based scheme	Proposed scheme	DCT-based scheme	DWT-based scheme
4096×2048 (1×1)	0	0	0	1.12	12.24	8.27
3072×1536 (0.75×0.75)	0	0	0	4.35	18.87	14.33
2048×1024 (0.5×0.5)	0	0	0	9.13	39.29	25.44
1536×768 (0.375×0.375)	1.17	3.17	2.21	14.92	56.74	43.29

To simulate the compression attack, we used HEVC [41] to encode the watermarked 360° VR video. For the simulation, the hierarchical B coding structure with a GOP =16 was used. Quantization parameters (QP) with values 25, 30, 35 were tested. Usually, compression attacks are accompanied by image scaling. We simulated the combined attacks with different encoding QPs (25, 30 and 35) and scaling ratios (0.75×0.75 and 0.5×0.5). The watermark

Table 3. FNR (%) of watermark detection for different compression attacks

QP	Scaling ratio	Detection from 360° data			Detection from viewport		
—	—	Proposed scheme	DCT-based scheme	DWT-based scheme	Proposed scheme	DCT-based scheme	DWT-based scheme
25	0.75×0.75	0	0	0	6.39	22.74	18.51
25	0.5×0.5	1.24	2.89	1.91	10.51	41.18	39.23
30	0.75×0.75	0	0.87	0.49	8.82	26.27	22.12
30	0.5×0.5	2.31	5.59	4.81	14.41	44.79	36.35
35	0.75×0.75	4.96	7.32	6.34	13.17	38.33	34.33
35	0.5×0.5	6.18	11.25	9.38	18.78	53.11	46.95

Table 4. FNR (%) of watermark detection for noise and contrast change attacks

Attack	Detection from 360° data			Detection from viewport		
—	Proposed scheme	DCT-based scheme	DWT-based scheme	Proposed scheme	DCT-based scheme	DWT-based scheme
Gaussian noise (0.04)	0	0	0	0	0.43	0
Gaussian noise (0.1)	0	1.83	1.27	2.45	3.21	2.87
Contrast change (30%)	0	0	0	0	0	0
Contrast change (70%)	0	0.62	0.34	0.52	1.1	0.96

detection results are shown in Table 3. It can be seen from Table 3 that the proposed scheme provides similar FNRs with the DCT-based and DWT-based schemes when the watermark is detected on the originally embedded domain. But for the detection from viewport, the proposed scheme offers significantly lower FNR values when compared to the DCT-based and DWT-based ERP-domain schemes.

4.4.2 Noise and contrast change attack. As the commonly-used image processing operations, adding white Gaussian noise and changing contrast are often used to attack the watermarked video. In the experiments, the pollution to the watermarked video by adding the zero-mean white Gaussian noise (default variance value of 0.04 and 0.1) was tested for watermark detection. In the contrast change attack, the contrast of watermarked frame was adjusted based on the image's histogram, and the 30% and 70% contrast change rates were simulated. Table 4 shows the VR video watermark detection results after noise and contrast change attacks. It can be seen from Table 4 that the proposed scheme works better against the Gaussian noise pollution compared to the DCT-based and DWT-based schemes. For the contrast change attack, it alters slightly the intensity of image pixel in spatial domain, so that it makes little impact on transform-domain watermarks. The low FNR values for all the three schemes in Table 4 verify this observation.

4.4.3 Geometry attack. Fig. 11 shows the results for a rotation attack in the 360° VR video watermarking system. In these results we performed rotations within 10° of the compressed videos (QP=25) with accompanying the appropriate cropping for the ERP image. It can be seen from Fig. 11 that the proposed spherical domain watermarking is robust to a rotation attack and especially for the viewport watermark detection, it shows lower FNRs than the DCT-based and DWT-based ERP-domain schemes. The proposed scheme supports viewport-dependent watermark detection, and it is naturally robust to the cropping attack. For watermark detection from viewport, the raw 360° VR video in the ERP format is usually needed to seek the viewport position and hence it is also used to register the rotated version of the watermarked 360° VR video, and then the watermark is detected in the spherical domain of the registered version. The registration improved the correlation between the extracted watermark and the original one. For watermark detection from the spherical domain and the ERP domain, the registrations were not used since they still keep the blind

Table 5. FNR (%) of watermark detection after Cubemap conversion attack

QP	Scaling ratio	Detection from 360° data			Detection from viewport		
——	——	Proposed scheme	DCT-based scheme	DWT-based scheme	Proposed scheme	DCT-based scheme	DWT-based scheme
25	0.75×0.75	8.29	13.74	12.45	22.67	34.21	30.22
25	0.5×0.5	14.25	21.63	17.39	38.83	47.28	42.73
30	0.75×0.75	9.45	16.22	14.47	27.53	39.67	34.39
30	0.5×0.5	11.64	18.47	16.33	35.49	61.39	52.25
35	0.75×0.75	19.38	21.49	17.32	29.17	43.47	39.48
35	0.5×0.5	26.91	35.21	31.49	41.78	62.94	57.39

watermark detection feature. This is why the FNRs for watermark detection from viewports are sometimes lower than those for spherical or ERP domain in Fig. 11.

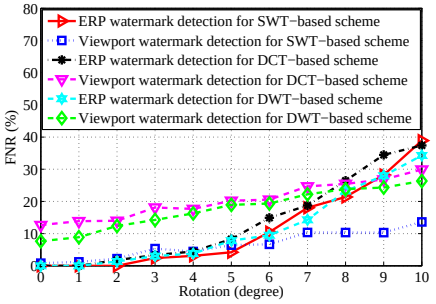


Fig. 11. Watermark detection results for rotation attacks (Averaged over all sequences)

4.4.4 Format conversion attack. The ERP format is often used for 360° VR video. However, after embedding the watermark in the ERP version, the other projection formats transformed from the watermarked ERP version can also be used to attack the watermark in 360° VR video. When the other format is used for 360° VR video applications, the other format will be first projected to the embedding domain and then the watermark is extracted from that domain. To evaluate the effects of format conversion on watermark detection, we tested the 4×3 Cubemap format attack. The attack combines compression and scaling, and the related watermark detection results are shown in Table 5. It can be seen from Table 5 that the proposed scheme shows a clear superiority to the DCT-based and DWT-based ERP-domain schemes in watermark detection rates even the larger FNRs compared with Table 3.

4.4.5 Camcording attack. The VR viewport display is enclosed by the helmet, and it is difficult to capture the 360° VR video with a camcorder. HMD viewing is naturally immune to the camcording attack. However, 360° VR video is sometimes played on a desktop screen. In such case, the 360° VR video can be captured by a camcorder. We used an iPhone 7 camera to simulate the camcording attack to 360° VR video playback (viewport viewing) on a desktop screen. During the simulation, the viewports at various positions were tested. Fig. 12 shows the watermark detection results averaged over all 360° videos captured by a camcorder. The iPhone 7 camera adopts the AVC/H.264 codec to encode the captured video. Counting in the captured distortion, camcording actually simulated a comprehensive attack by combining compression, scaling, rotation and cropping. It can be seen from Fig. 12 that under the desktop screen camcording attack, although the three watermarking schemes offer larger FNRs with increasing scaling ratio, the proposed scheme provides the best robustness among all the three schemes for viewport watermark detection.

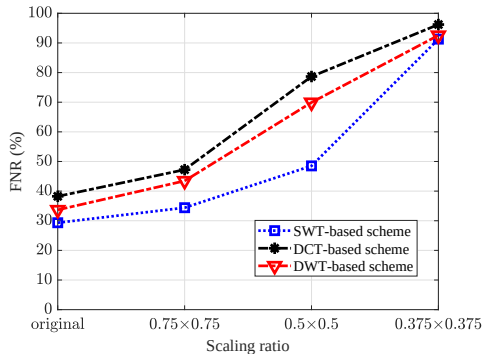


Fig. 12. Viewport watermark detection results under camcording attacks

Table 6. FNR (%) of watermark detection from viewport for different attacks

Attacks	Proposed scheme	3D watermarking scheme [43]
Compression(QP=35)	10.83	24.21
Scaling (0.5×0.5)	9.13	18.32
Gaussian noise (0.1)	2.45	13.64
Contrast change (70%)	0.52	16.32
Rotation (5°)	6.23	51.35
Format conversion	21.87	65.64
Camcording	28.53	79.72

4.5 Comparison with 3D watermarking scheme

The proposed watermarking scheme embeds the watermark in the spherical 360° data in 3D space. It has similar application properties with texture-based 3D watermarking. The texture-based 3D watermarking scheme embedds the watermark into the textures of 3D objects and extracts the watermark from the partially reconstructed texture of the 2D view of the 3D object. In texture watermarking, the Eurëmark algorithm [43] was used. When the texture-based 3D watermarking scheme is used for VR video, the ERP image and viewport image are regarded as the texture image and the 2D view of object, respectively, and the watermark is extracted from the recovered viewport data in the ERP image. Table 6 presents the viewport watermark detection results of the proposed scheme compared with that of the texture-based 3D watermarking scheme after different types of attacks. Table 6 shows that all of the results for the proposed scheme are superior to those of the texture-based 3D watermarking scheme. It illustrates that the proposed scheme that embeds the watermark in the spherical domain is more robust to attacks than the texture-based 3D watermarking scheme that embeds the watermark in the planar texture domain.

5 CONCLUSION

Aiming for 360° VR video copyright protection, this paper proposes a spherical domain watermarking scheme. By considering the advantage of the spherical shape in 360° VR video representation, the spherical wavelet transform for 360° VR video is exploited to hide the watermark in the sphere spectrum. At the same time, the viewport-oriented JND model is proposed to regulate the watermark strength to guarantee watermark transparency by accommodating the HMD viewing conditions. Watermark detection both from the spherical domain and the viewport projection are supported. When compared to the state-of-the-art DCT-based or DWT-based ERP domain watermarking, the proposed system provides similar robustness to the common attacks that are used for planar videos but offers better

robustness to the viewport-dependent transmission attack that is engineered specifically for 360° VR video.

ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China under Grant 61771469 and Ningbo Natural Science Foundation under Grant 2019A610109.

REFERENCES

[1] MPEG Experts. 2016. Summary of survey on virtual reality. ISO/IECJTC 1/SC 29/WG 11, m16542.
[2] Huawei-iLab. 2018. Cloud VR Network Solution White Paper. http://www.huawei.com/.
[3] L. E. Beausoleil. 2017. Copyright Issues and Implications of Emerging Virtual Reality Technologies. Boston College Intellectual Property and Technology Forum, http:// http://bciptf.org/.
[4] N. P. Sheppard, R. Safavi-Naini and P. Ogunbona. 2002. Digital watermarks for copyright protection. Journal of Law and Information Science 12, 1(2002),110-130, 2002.
[5] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, T. Kalker. 2008. Digital Watermarking and Steganography, Morgan Kaufmann Publisher.
[6] K. Liu, Y. Liu, J. Liu, A. Argyriou, X. Yang. 2017. Joint Source Encoding and Networking Optimization for Panoramic Streaming Over LTE-A Downlink. In Proc. IEEE Globecom 2017.
[7] Y. Liu, J. Liu, A. Argyriou, S. Ci. 2019. MEC-assisted Panoramic VR Video Streaming over Millimeter Wave Mobile Networks. IEEE Transactions on Multimedia 21, 5(2019),1302-1316.
[8] Y. He, B. Vishwanath, X. Xiu, Y. Ye. 2016. AHG8: InterDigitals projection format conversion tool. Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET, D0021.
[9] P. Schröder, W. Sweldens. 1995. Spherical Wavelets: Efficiently Representing Functions on the Sphere. In Proc. SIGGRAPH, 161-172.
[10] F. Simons, I. Loris, G. Nolet, I. C. Daubechies, S. Voronin, et al. 2011. Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. Geophysical Journal International 187, 969-988.
[11] Z. Wang, C. Leung, Y. Zhu, and T. Wong. 2004. Data compression with spherical wavelets and wavelets for the image-based relighting. Computer Vision and Image Understanding 96, 3(2004), 327-344.
[12] J. D. McEwen, P. Vandergheynst, Y. Wiaux. 2013. On the computation of directional scale-discretized wavelet transforms on the sphere. In Proc. SPIE Wavelets and Sparsity XV, 8858.
[13] P. W. Wong and N. Memon. 2001. Secret and public key image watermarking schemes for image authentication and ownership verification. IEEE Transactions on Image Processing 10, 10(2001), 1593-1601.
[14] J. S. Tsai, W. B. Huang and Y. H. Kuo. 2011. On the selection of optimal feature region set for robust digital image watermarking. IEEE Transactions on Image Processing 20, 3(2011), 735-743.
[15] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. 1997. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing 6, 12(1997), 1673-1687.
[16] M. Amini, M. Ahmad, and M. Swamy. 2018. A robust multibit multiplicative watermark decoder using vector-based hidden markov model in wavelet domain. IEEE Transactions on Circuits and Systems for Video Technology 28, 2(2018), 402-413.
[17] Y.-S. Lee, Y.-H. Seo, and D.-W. Kim. 2019. Blind image watermarking based on adaptive data spreading in n-level DWT subbands. Security and Communication Networks 2019, 1-11.
[18] D. Xu, R. Wang, and Y. Q. Shi. 2014. Data hiding in encrypted H.264/AVC video streams by codeword substitution. IEEE Transactions on Information Forensics and Security 9, 4(2014), 596-606.
[19] T. Dutta and H. P. Gupta. 2016. A robust watermarking framework for high efficiency video coding (HEVC) encoded video with blind extraction process. Journal of Visual Communication and Image Representation 38, 29-44.
[20] G. Hua, J. Goh, and V. L. L. Thing. 2015. Time-spread echo-based audio watermarking with optimized imperceptibility and robustness. IEEE/ACM Transactions on Audio, Speech, and Language Processing 23, 2(2015), 227-239.
[21] C. H. Chou and K. C. Liu. 2010. A perceptually tuned watermarking scheme for color images. IEEE Transactions on Image Processing 19, 11(2010), 2966-2982.
[22] M. Urvoy, D. Goudia, and F. Autrusseau. 2014. Perceptual DFT watermarking with improved detection and robustness to geometrical distortions. IEEE Transactions on Information Forensics and Security 9, 7(2014), 1108-1119.

[23] E. Halici and A. A. Alatan. 2009. Watermarking for depth-image-based rendering. In Proc. IEEE Int. Conf. Image Process., 4217-4220.
[24] Y. Lin and J. Wu. 2011. A Digital Blind Watermarking for Depth-Image-Based Rendering 3D Images. IEEE Transactions on broadcasting 57, 2(2011), 602-611.
[25] H. D. Kim, J. W. Lee, T. W. Oh, and H. K. Lee. 2012. Robust DT-CWT watermarking for DIBR 3D images. IEEE Trans. Broadcast. 58, 4(2012), 533-543.
[26] A. Koz, C. Cigla, and A. A. Alatan. 2010. Watermarking of free-view video. IEEE Transactions on Image Processing 19, 7(2010), 1785-1797.
[27] Y. Miura, X. Li, S. Kang, and Y. Sakamoto. 2018. Data hiding technique for omnidirectional JPEG images displayed on VR spaces. In Proc. International Workshop on Advanced Image Technology, 1-4.
[28] J. Kang, S. Ji, and H. Lee. 2019. Spherical panorama image watermarking using viewpoint detection. In Proc. Digital Forensics and Watermarking. Cham: Springer International Publishing, 95-109.
[29] J. D. McEwen, M. P. Hobson, D. J. Mortlock, and A. N. Lasenby. 2007. Fast directional continuous spherical wavelet transform algorithms. IEEE Trans. Signal Process. 55, 2(2007), 520-529.
[30] E. Eade. 2017. Lie Groups for 2D and 3D Transformations", In: URL: http://www.ethaneade.com/lie.pdf.
[31] J. D. McEwen, C. Durastanti, and Y. Wiaux. 2018. Localisation of directional scale-discretised wavelets on the sphere. Appl. Comput. Harmon. Anal. 44, 59-88.
[32] S. Mallat. 1996. Wavelets for vision. Proc. IEEE 84, 604-614.
[33] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor. 1997. Visibility of wavelet quantization noise. IEEE Trans. Image Processing 6, 8(1997), 1164-1175.
[34] A. P. Bradley. 1999. A wavelet visible difference predictor. IEEE Trans. Image Process. 5, 8(1999), 717-730.
[35] L. Sorgi and K. Daniilidis. 2004. Normalized cross-correlation for spherical images. In Proc. European Conference on Computer Vision (ECCV).
[36] M. L. Miller and J. A. Bloom. 1999. Computing the probability of false watermark detection. In Proc. 3rd Int. Workshop Inf. Hiding, 146-158.
[37] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J.-P. Thiran. 2012. Scale invariant feature transform on the sphere: Theory and applications. International Journal of Computer Vision 98, 217-241.
[38] E. Alshina, J. Boyce, A. Abbas, Y. Ye. 2017. JVET common test conditions and evaluation procedures for 360°video", JVET, H1030.
[39] A. Singla, W. Robitza, A. Raake. 2018. Comparison of Subjective Quality Evaluation Methods for Omnidirectional Videos with DSIS and Modified ACR. In Proc. Human Vision and Electronic Imaging (HVEI), 2018.
[40] ITU-R Recommendation BT.500-13. 2012. Methodology for the subjective assessment of the quality of television pictures. Geneva, Switzerland, International Telecommunication Union.
[41] High Efficiency Video Coding (HEVC). Jan. 2013. Rec. ITU-T H.265 and ISO/IEC 23008-2.
[42] C. Burini, S. Baudry, and G. Doërr. 2014. Blind detection for disparity-coherent stereo video watermarking. In Proc. SPIE 9028, Media Watermarking, Security, and Forensics 2014, 90280B.
[43] E. Garcia, and J. Dugelay. 2003. Texture-based watermarking of 3D video objects. IEEE Trans. Circuits Syst. Video Techn. 13 (2003), 853-866.
[44] S. Baldoni, M. Brizzi, M. Carli, and A. Neri. 2019. A watermarking model for omni-directional digital images, 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 240-245.
[45] J. Jin, M. Dai, H. Bao, Q. Peng. 2004. Watermarking on 3D mesh based on spherical wavelet transform. Journal of Zhejiang University-SCIENCE A 5, 3(2004), 251-258.
[46] S. Voloshynovskiy, S. Pereira, A. Herrigel, N. Baumgartner, and T. Pun. 2000. Generalized watermarking attack based on watermark estimation and perceptual remodulation. In Proc. SPIE, vol.3971, 358-370.
[47] G. Doërr and J. Dugelay. 2004. Security pitfalls of frame-by-frame approaches to video watermarking. IEEE Trans. Signal Process. 52, 10(2004), 2955-2964.
[48] J.Kang, J. Hou, S. Ji, H. Lee. 2020. Robust Spherical Panorama Image Watermarking Against Viewpoint Desynchronization. IEEE Access 8 (2020), 127477-127490.