A Robust 3D Video Watermarking Scheme Based on Multi-modal Visual Redundancy

Congxin Cheng[1,2], Wei Ma[1,2(✉)], Yuchen Yang[1,2], Shiyang Zhang[1,2], and Mana Zheng[1,2]

[1] Faculty of Information Technology, Beijing University of Technology, Beijing, China
mawei@bjut.edu.cn
[2] Beijing Key Laboratory of Trusted Computing, Beijing, China

Abstract. Image watermarking is a popular research topic in signal processing. The paper presents a blind watermarking scheme for 3D videos. Given a 3D video, each frame of both views is divided into blocks. Watermark information is embedded by modulating selected DCT coefficients of each block. The modulation strength is controlled by multi-modal visual redundancies existing in the 3D video. Specifically, we compute an intra-frame Just-noticeable Distortion (JND) value and an inter-frame reference value for the block to determine the strength. The former reflects the visual redundancies in the image plane. The latter represents the visual redundancies of the block from aspects of motion between sequential frames and disparity between the left and right views. We validate the robustness of the proposed watermarking scheme under various attacks through experiments. More importantly, the visual quality of the 3D videos watermarked by our scheme is proved to be as good as that of the original videos, by a proposed LDQ (Loss of Disparity Quality) criterion specially designed for 3D videos, as well as PSNR of single views.

Keywords: Stereo video watermarking · Blind watermarking Robustness · Perceptual redundancy · Disparity

1 Introduction

With the rapid development of 3D technologies, high-definition stereo media is widely used in many areas, such as 3D movies, virtual reality, etc. Such media is invaluable. How to protect them from piracy has become a significant issue [2,6,9–11,14,15]. Digital watermarking is an effective technique to achieve this task via embedding the owners' information into the 3D media [2,10,17].

There are three main considerations for digital watermarking schemes, transparency, robustness and capacity, respectively [5]. First, the quality of the original images/videos should not be affected after being watermarked. Otherwise, it will decrease the commercial value of the media. Second, watermarks could

Y. Zhao et al. (Eds.): ICIG 2017, Part III, LNCS 10668, pp. 517–526, 2017.
https://doi.org/10.1007/978-3-319-71598-8_46

not be removed intentionally or unintentionally. It means that the watermarking schemes should be robust to various attacks of changing the media, e.g. cropping and scaling. Third, the schemes should be able to embed watermarks carrying enough information, i.e. having high capacity.

3D videos generally have two forms, lightweight depth-image-based rendered videos [4] and high-quality stereoscopic videos with left-right views [16]. The former is generally used in Internet or TV systems. The latter is popular in off-line movie market. We are targeting at watermarking high-quality stereoscopic videos, rather than lightweight ones [1], to claim their copyrights in the off-line market. Watermarking in these videos is more challenging due to the sharp conflict between high visual quality requirements and capacity.

In recent years, there appear many watermarking methods for stereo videos. An intuitive strategy is to apply single view image watermarking schemes [8] to every frames of both views of stereo videos, individually as done in [9]. These methods tend to cause visual quality distortion in the watermarked videos, since consistencies between frames and the two views are corrupted. Wu et al. [15] embedded watermarks by altering the DCT coefficients of the left and right views in opposite directions. Although their method is robust to many attacks, the correspondences between the left and right views are damaged, which leads to visual discomfort. Rana et al. [13] determined the embedding positions in the hosts by considered both depth and motion information. This method keeps the consistencies between frames and those between views. However, it sacrifices capacity for the consistencies. Besides, the embedded positions are fragile to attacks. Mohghegh et al. [7] controlled the embedding strength in order to keep stereo correspondences. This method has small capacity and low robustness against salt & pepper noise attacks.

This paper presents a blind watermarking scheme for 3D videos. It explores multi-modal visual redundancies in the 3D videos, thereby having high capacity and robustness while keeping good visual quality in watermarked videos. Given a 3D video, at first, all frames of both left and right views are partitioned into non-overlapped blocks. For each block, we compute its DCT coefficients. In the meanwhile, we compute a depth map and a set of motion vectors for each frame. The depth values and the motion vectors in a block are used to calculate a reference value for the block, which represents the perceptual redundancies existing between sequential frames and the left and right views. The reference value, together with Just-noticeable Distortion (JND) [14] which represents those perceptual redundancies in single frames, are used to control the embedding strength of the watermark in the DCT coefficients. The control can effectively preserve visual quality of watermarked videos by avoiding over embedding. In the detection phase, we simply perform the above blocking and DCT steps, and extract watermarks by comparing the DCT coefficients. Experimental results show that our scheme provides high capacity, obtaining good visual quality of watermarked videos and high robustness against various attacks.

On the other hand, state-of-the-art watermarking methods [2,9] generally use PSNR as criterions to evaluate the visual quality of watermarked videos.

PSNR is originally designed for single view images. The consistency of the left and right views, which is significant for high-quality 3D videos [5], are not considered. In this paper, we propose a new measurement, called Loss of Disparity Quality (LDQ). It defines the visual quality of 3D videos from the view of depth perception, which is essential to the visual experiences of 3D videos. LDQ is used together with PSNR to evaluate the visual quality of 3D watermarked videos in this paper.

There are two main contributions in this paper. Firstly, a watermarking scheme for 3D videos is proposed and validated. It is demonstrated to have better performances than state-of-the-art methods. Secondly, to the best of our knowledge, we are the first to give an evaluation method specifically designed for 3D video watermarking.

2 Proposed Method

In this section, we explain the embedding and extracting processes of the proposed method, respectively. Since the extracting is simply an inverse process of the embedding, we put more effort in describing the details of the embedding stage.

2.1 Embedding

Given a stereo video, the same operations are carried out on every pair of left and right frames. First, we divide a pair of left and right frames into non-overlapping blocks of size of 8×8 pixels. Then, DCT coefficients of each block are computed. Next, watermark embedding is performed for each pair of blocks at the same positions of the left and right frames. Each bit of a watermark, which takes the form of a binary image in this paper, is repeatedly embedded in five selected pairs of middle-frequency DCT coefficients in the two blocks in order to descend the probability of losing information during malicious attacks. The embedding strength is sophisticatedly controlled by various human visual perception factors in viewing stereo videos.

We choose five middle-frequency DCT coefficients (indicated in blue in the grid of Fig. 1) in both of the two blocks to embed a single bit of watermark for robustness. Given a selected position (i, j) in the left/right view block, we record its average value $g_{(i,j)}$,

$$g_{(i,j)} = (C_{(i,j)} + C_{(i+1,j)} + C_{(i,j+1)} + C_{(i+1,j+1)})/4 \tag{1}$$

where $C_{(i,j)}$ is the DCT coefficient at (i, j).

We define

$$\begin{aligned} G^l_{(i,j)} &= \omega_1 g^l_{(i,j)} + \omega_2 g^r_{(i,j)} \\ G^r_{(i,j)} &= \omega_1 g^r_{(i,j)} + \omega_2 g^l_{(i,j)} \end{aligned} \tag{2}$$

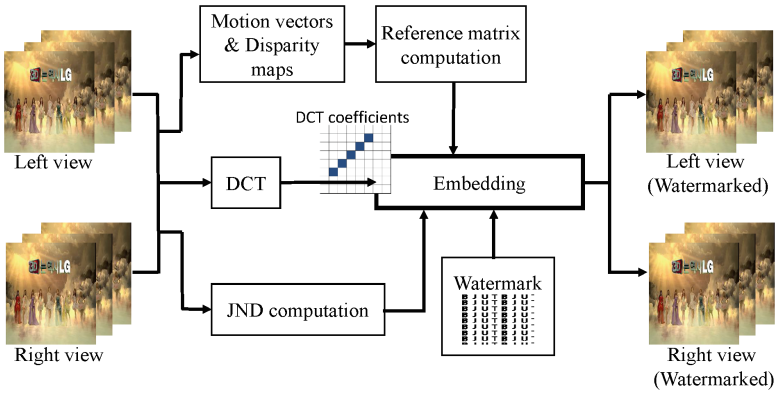

Fig. 1. Pipeline of the embedding process. (Color figure online)

where $g^r_{(i,j)}$ and $g^l_{(i,j)}$ are the average values of the left and right view blocks at (i, j), respectively. $\omega_1 + \omega_2 = 1$. ω_1 is chosen to be six times larger than ω_2 in implementation, so that $G^l_{(i,j)}$ and $G^r_{(i,j)}$ average information from both views while taking a larger portion from their origins, i.e. $g^l_{(i,j)}$ and $g^r_{(i,j)}$, respectively. Embedding a bit of the watermark, which is denoted as $\omega_{(p,q)}$, in the selected DCT coefficients is to alter the coefficients by the state of $\omega_{(p,q)}$ and the values of $G^l_{(i,j)}$ and $G^r_{(i,j)}$. If the bit is 0, $C^l_{(i,j)}$ is modified to be slightly bigger than $G^l_{(i,j)}$. $C^r_{(i,j)}$ is set to be slightly smaller than $G^r_{(i,j)}$. On the contrary, if the bit is 1, $C^l_{(i,j)}$ and $C^r_{(i,j)}$ are set to be slightly smaller than $G^l_{(i,j)}$ and bigger than $G^r_{(i,j)}$, respectively. If $C^l_{(i,j)}$ or $C^r_{(i,j)}$ have already satisfied the above polarity relationship, its modulation will be passed.

The above embedding process can be expressed by the following equations.

$$
\begin{aligned}
C^l_{(i,j)} &= \begin{cases} G^l_{(i,j)} + \beta_{(p,q)}(\alpha JND^l_{(i,j)} + \varepsilon^l_{(i,j)}), \omega_{(p,q)} = 0 \\ G^l_{(i,j)} - \beta_{(p,q)}(\alpha JND^l_{(i,j)} + \varepsilon^l_{(i,j)}), \omega_{(p,q)} = 1 \end{cases} \\
C^r_{(i,j)} &= \begin{cases} G^r_{(i,j)} - \beta_{(p,q)}(\alpha JND^r_{(i,j)} + \varepsilon^r_{(i,j)}), \omega_{(p,q)} = 0 \\ G^r_{(i,j)} + \beta_{(p,q)}(\alpha JND^r_{(i,j)} + \varepsilon^r_{(i,j)}), \omega_{(p,q)} = 1 \end{cases}
\end{aligned}
\tag{3}
$$

where $C^l_{(i,j)}$ and $C^r_{(i,j)}$ denote the final DCT coefficients after modulation, at (i, j) of the left and right views, respectively. $JND^l_{(i,j)}$ and $JND^r_{(i,j)}$ denote the JND values [14] of the left and right views, respectively. α is a parameter controlling the influences of JND. α is empirically selected to be 0.05, which could provide robustness strong enough while ensuring visual quality of the videos after embedding. The modulation range of the DCT coefficients is restrained by the JND values and a minimal value ε. Here, $\varepsilon_{(i,j)} = 0.1 g_{(i,j)}$. Moreover, in order to avoid large changes of small coefficients, the embedding process won't be conducted if the modulation range is more than twice the value of the original DCT coefficient.

Compared to α, fixed for all of the blocks and frames, $\beta_{(p,q)}$ in (3) is flexibly determined by a reference value $R_{(p,q)}$ at block (p,q),

$$\beta_{(p,q)} = \begin{cases} 0.5, & R_{(p,q)} = 0 \\ 1, & R_{(p,q)} = 1 \end{cases} \tag{4}$$

$R_{(p,q)}$ is calculated by referring to the motion and depth properties of block (p, q). At first, we compute the motion factors of pixel (i, j) in block (p, q), along X axis and Y axis, by [3], and recorded as $M^x_{(i,j)}$ and $M^y_{(i,j)}$, respectively. The motion of pixel (i, j) is given by

$$M_{(i,j)} = \sqrt{{M^x_{(i,j)}}^2 + {M^y_{(i,j)}}^2} \tag{5}$$

In parallel, the depth of pixel (i, j), denoted as $D_{(i,j)}$, is computed by the SGBM algorithm given in [6]. By combining the motion and depth information, we obtain a weighted average influence value of each block, denoted as $MD_{(p,q)}$ which is given by

$$MD_{(p,q)} = \frac{\sum_i \sum_j (\mu_1 D_{(i,j)} + \mu_2 M_{(i,j)})}{64} \tag{6}$$

where μ_1 and μ_2 indicate the percentage of depth and motion. μ_1 and μ_2 are empirically chosen to be 5/6 and 1/6, respectively, since in our experiments, the depth information computed by [6] is evaluated to be more reliable than the motion part obtained by [3]. The number 64 in the denominator is the number of pixels in the block.

In implementation, we record the reference values of all the blocks in each frame in a reference matrix. The reference value $R_{(p,q)}$ of the block (p, q) is given by

$$R_{(p,q)} = \begin{cases} 1, & MD_{(p,q)} \geq MD_{avg} \\ 0, & MD_{(p,q)} < MD_{avg} \end{cases} \tag{7}$$

Here, MD_{avg} is the average value in the reference matrix. If the influence factor of block (p, q) is beyond the average value of the reference matrix, $R_{(p,q)}$ is set to 1, which suggests a strong embedding strength in this block, and vice versa. This is consistent with that fact that human vision system is less sensitive to objects with large motion or depth.

2.2 Extracting

In the extracting phase, given a watermarked video, we divide each frame into blocks and compute the DCT coefficients of each block, as done in the embedding step. The total differences between coefficients of a pair of corresponding blocks in the left and right views is computed by

$$z_{(p,q)} = \sum_{(i,j)} (C'^{l}_{(i,j)} - C'^{r}_{(i,j)}) \tag{8}$$

where p and q are block indices. i and j are the indices of the DCT coefficients in the block. As described in the embedding part, the same bit of the watermark is embedded into five DCT coefficients in one pair of blocks for robustness. To extract a bit of the watermark, we compute the differences between corresponding DCT coefficients in the left and right view blocks. Then we sum the differences at the selected five positions to mitigate influences from attacks. The bit of the watermark $w_{(p,q)}$ hidden in blocks at (p,q) is extracted by $z_{(p,q)}$,

$$w_{(p,q)} = \begin{cases} 1, & z_{(p,q)} \geq 0 \\ 0, & z_{(p,q)} < 0 \end{cases} \tag{9}$$

Note that the modification of the selected coefficients in Eq. 3 is moderately controlled for high visual quality of the videos with watermarks. Therefore, the coefficients in the left block is not necessarily changed to be larger or smaller than those in the right block, as supposed in the extraction. It means that the extracted watermark might lose little information as shown in Fig. 2. Nevertheless, the proposed method has good robustness under various attacks as we can see in the experiments.

3 Experiments

In this section, we analyse the performances of the proposed watermarking scheme in aspects of keeping visual quality and the scheme's capacity. Beside, we also test its robustness against various attacks. Keeping visual quality is the most important factor for watermarking high quality stereo videos in off-line movie market. It is also the fundamental consideration during the design of our watermarking scheme. Three 3D videos, shown in Fig. 2, are used in the experiments. The videos, whose specifications are listed in Table 1, all have high resolution. Note that we treat the videos as a series of frames with no video compression. A binary image (shown in Fig. 2(j)) is used as a watermark pattern, which could be repetitively embedded.

3.1 Visual Quality Evaluation and Capacity Analysis

We compare the visual quality of watermarked videos generated by our method with those obtained by four state-of-the-art methods, including a visual-module-based method proposed by Niu et al. [9], a differential watermarking scheme

Table 1. Specifications of 3D video sequences

	Resolution	Frames
Video 1 (Fig. 2(a))	1920 × 1080	24
Video 2 (Fig. 2(b))	1920 × 1080	47
Video 3 (Fig. 2(c))	1920 × 1080	82

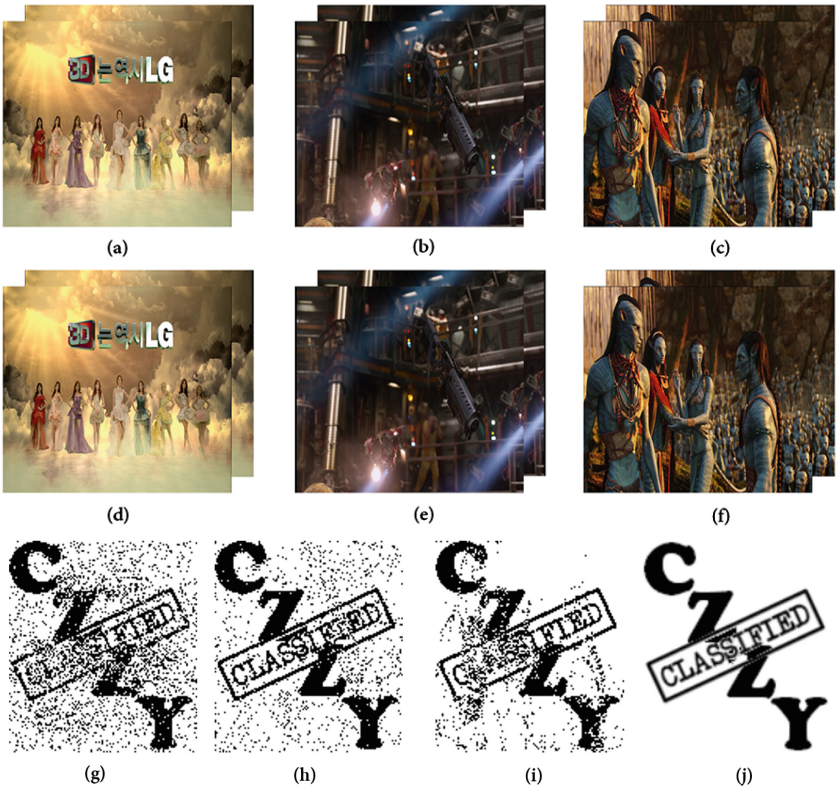

Fig. 2. (a), (b) and (c) are original stereo video frames (video 1, 2 and 3 in Table 1); (d), (e) and (f) are watermarked frames; (g), (h) and (i) are watermarks extracted from (d), (e) and (f), respectively; (j) is the original watermark pattern.

proposed by Wu et al. [15], an adaptive stereo watermarking scheme using non-corresponding blocks proposed by Mohaghegh et al. [7] and a 3D video watermarking scheme based on 3D-HEVC encoder proposed by Rana et al. [13]. The visual quality is evaluated by traditional Peak Signal to Noise Ratio (PSNR) for single-view fidelity and a proposed LDQ for inter-view consistency.

The watermarked videos produced by the proposed method (shown in Fig. 2(d), (e) and (f)) looks totally the same with the original videos. We use PSNR to quantitatively evaluate the similarity between the watermarked videos and the original ones. The higher the PSNR is, the better the visual quality of the watermarked videos is. We compare the PSNR values of our scheme with those of the four state-of-the-art methods (listed in Table 2). It is pretty clear that the proposed method generates watermarked videos with the best visual quality.

PSNR simply quantizes the visual fidelity in single views. For stereo videos, the consistency between left and right views are more important. If the consistency is

Table 2. PSNR, LDQ and capacity of the five schemes.

	Proposed scheme	Niu's scheme [9]	Wu's scheme [15]	Rana's scheme [13]	Mohaghegh's scheme [7]
PSNR	51.11	50.33	48.93	47.80	46.90
LDQ	0.8171	0.7697	0.7639	0.9631	0.8027
Capacity	1/64	1/2048	1/64	1/4098	1/1024

corrupted, the video with watermarks will cause viewers uncomfortable and even dizzy. Inspired by the stereo consistency constraint used for stereo image segmentation in [12], we introduce a new evaluation on the visual quality of watermarked 3D videos, called Loss of Disparity Quality (LDQ), given by

$$LDQ = \sum_{n} \frac{min(B_{(p_i,p_j)}, B'_{(p_i,p_j)})}{max(B_{(p_i,p_j)}, B'_{(p_i,p_j)})} / n \tag{10}$$

Here, p_i and p_j denote a pair of corresponding pixels in left and right views, respectively. n is the number of the pairs of corresponding pixels. $B_{(p_i,p_j)}$ and $B'_{(p_i,p_j)}$ stand for color differences between corresponding pixels p_i and p_j, before and after embedding, respectively. $B_{(p_i,p_j)}$ is given by

$$B_{(p_i,p_j)} = e^{-(\frac{0.5\|c_i - c_j\|^2}{256})^{0.5}} \tag{11}$$

c_i and c_j are the colors of p_i and p_j, respectively.

LDQ represents the variance of the color differences between stereo corresponding pixels in the original and watermarked videos. Ideally, the variation should be close to zero. The closer to 1 the LDQ value is, the better the disparity quality of the watermarked videos is. Table 2 lists the LDQ values of all the five methods. From the table, we can see that our method obtains high LDQ as well.

It's known that low watermark capacity is one of the factors resulting in good visual quality. Since our method performs best in PSNR and LDQ, its capacity might be low. In order to eliminate this concern, we present the capacities, i.e. the number of bits hidden in one pixel in the host videos, of the five methods in Table 2. From the last row of Table 2, we can see that our method and Wu's scheme [15] have the highest capacity among the five methods.

3.2 Robustness Analysis

As explained in the method part, since the scheme is designed to be blind for convenience, the extracted watermarks under no attack might have lost parts of information, compared with the original ones (refer to Fig. 2). This will not result in low robustness of the method against attacks. In order to evaluate the robustness of the proposed method against various attacks, the average NC

values of the watermarks extracted by the five methods under various attacks are calculated and given in Table 3. The attacks, including compression (here we use JPEG compression since we treat the video as separate frames), relative mild salt & pepper noises, gaussian filtering and scaling, are common in reusing the high-quality videos. From the table, we can see that the proposed method obtains the highest NC value under the attacks of Salt & Pepper noise with noise power of 0.01, and the second or third highest values under the other attacks.

Table 3. NC values of the watermarks extracted by the five schemes under various attacks.

Attacks	Attack's parameters	Proposed scheme	Niu's scheme [9]	Wu's scheme [15]	Rana's scheme [13]	Mohaghegh's scheme [7]
JPEG compression	Quality 95	0.7649	0.8447	0.7672	0.6249	0.7547
Salt & pepper noise	Noise power 0.01	0.7275	0.6950	0.6906	0.6358	0.6387
	Noise power 0.03	0.6435	0.6075	0.6324	0.6096	0.6759
Gaussian filtering	Filter size (3, 3)	0.7802	0.7838	0.7737	0.5986	0.7871
Scaling	Resolution 1280×720	0.8409	0.4864	0.8477	0.5782	0.7246
	Resolution 640×480	0.6492	0.5625	0.6449	0.5520	0.6783

4 Conclusion

In this paper, a watermarking scheme for 3D videos was introduced. The highlight of the scheme is to control the embedding strength of watermarks by exploiting multi-modal visual redundancies in 3D videos. The computation of the visual redundancies integrates cues from intra-frame saliency, motion between sequential frames and disparity between the left and right views. Through experiments, we demonstrate the performances of the scheme in visual quality, capacity and robustness. Experimental results show that the proposed method generates watermarked videos with good visual quality, has large capacity and performs well in resisting various attacks. We also presented an evaluation method on the loss of disparity information of watermarked 3D videos, which fills a gap in evaluating the visual quality of 3D watermarking.

There are still limitations in our work. In the future, rather than considering the robustness only in stereoscopic image planes, we will attempted to improve the scheme for robustness against video compression which is commonly used for low storage and fast transmission.

Acknowledgement. This research is supported by Scientific Research Project of Beijing Educational Committee (KM201510005015), Beijing Municipal Natural Science Foundation (4152006), National Natural Science Foundation of China (61672068, 61370113), and Seed Funding for International Cooperation of Beijing University of Technology.

References

1. Asikuzzaman, M., Alam, M.J., Lambert, A.J., Pickering, M.R.: A blind watermarking scheme for depth-image-based rendered 3D video using the dual-tree complex wavelet transform. In: IEEE International Conference on Image Processing, pp. 5497–5501 (2014)
2. Chammem, A.: Robust watermarking techniques for stereoscopic video protection. Evry Institut National Des Tlcommunications (2013)
3. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 363–370. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-45103-X_50
4. Fehn, C.: Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In: Electronic Imaging, pp. 93–104 (2004)
5. Gupta, V., Barve, A.: A review on image watermarking and its techniques. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **4**(1), 92–97 (2014)
6. Hu, F., Zhao, Y.: Comparative research of matching algorithms for stereo vision. J. Comput. Inf. Syst. **9**(13), 5457–5465 (2013)
7. Mohaghegh, H., Karimi, N., Soroushmehr, S.M., Samavi, S.M.R., Najarian, K.: Adaptive stereo medical image watermarking using non-corresponding blocks. In: International Conference of the Engineering in Medicine and Biology Society, pp. 4214–4217 (2015)
8. Mousavi, S.M., Naghsh, A., Manaf, A.A., Abu-Bakar, S.A.R.: A robust medical image watermarking against salt and pepper noise for brain MRI images. Multimedia Tools Appl. **76**(7), 10313–10342 (2017)
9. Niu, Y., Souidene, W., Beghdadi, A.: A visual sensitivity model based stereo image watermarking scheme. In: 3rd European Workshop on Visual Information Processing, pp. 211–215 (2011)
10. Onural, L., Ozaktas, H.M.: Three-dimensional Television: From Science-fiction to Reality. Springer, Heidelberg (2008)
11. Ou, Z., Chen, L.: A robust watermarking method for stereo-pair images based on unmatched block bitmap. Multimedia Tools Appl. **75**(6), 3259–3280 (2016)
12. Price, B.L., Cohen, S.: StereoCut: Consistent interactive object selection in stereo image pairs. In: IEEE International Conference on Computer Vision, pp. 1148–1155 (2011)
13. Rana, S., Sur, A.: Blind 3D video watermarking based on 3D-HEVC encoder using depth. In: Proceedings of Indian Conference on Computer Vision Graphics and Image Processing, pp. 1–8 (2014)
14. Wei, Z., Ngan, K.N.: Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain. IEEE Trans. Circ. Syst. Video Technol. **19**(3), 337–346 (2009)
15. Wu, C., Yuan, K., Cheng, M., Ding, H.: Differential watermarking scheme of stereo video. In: IEEE 14th International Conference on Communication Technology, pp. 744–748 (2012)
16. Yan, T., Lau, R.W.H., Xu, Y., Huang, L.: Depth mapping for stereoscopic videos. Int. J. Comput. Vision **102**(1), 293–307 (2013)
17. Yang, W., Chen, L.: Reversible DCT-based data hiding in stereo images. Multimedia Tools Appl. **74**(17), 7181–7193 (2015)