

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/136599>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Scene-based Imperceptible-visible Watermarking for HDR Video Content

K. Perez-Daniel and F. Garcia-Ugalde

Engineering Faculty

National Autonomous University of Mexico (UNAM), Mexico City, Mexico

krperezd@hotmail.com and fgarciau@unam.mx

V. Sanchez

Dept. of Computer Science

University of Warwick, Coventry, UK

V.F.Sanchez-Silva@warwick.ac.uk

Abstract—This paper presents the High Dynamic Range - Imperceptible Visible Watermarking for HDR video content (HDR-IVW-V) based on scene detection for robust copyright protection of HDR videos using a visually imperceptible watermarking methodology. HDR-IVW-V employs scene detection to reduce both computational complexity and undesired visual attention to watermarked regions. Visual imperceptibility is achieved by finding the region of a frame with the highest hiding capacities on which the Human Visual System (HVS) cannot recognize the embedded watermark. The embedded watermark remains visually imperceptible as long as the normal color calibration parameters are held. HDR-IVW-V is evaluated on PQ-encoded HDR video content successfully attaining visual imperceptibility, robustness to tone mapping operations and image quality preservation.

Index Terms—HDR, watermark, imperceptibility, scene detection

I. INTRODUCTION

High Dynamic Range (HDR) content extends the contrast between the brightest whites and darkest blacks and incorporates a wider color gamut. New transfer functions (TFs) are used to represent this extended dynamic range with a limited number of digital codes. Specifically, TFs perform the mapping between the linear light components and a non-linear imaging signal. In this way, TFs can emulate the Human Visual System (HVS) using nonlinear operations to quantize the sample values with minimal subjective distortions. Nowadays, the most common TFs to represent HDR content are called OETFs (Opto-Electronic Transfer Functions) and are standardized under the specifications SMPTE ST.2084, also known as Perceptual Quantizer (PQ) [1], and the recommendation REC.2100, also known as the Hybrid Log Gamma (HLG) TF [2].

Several broadcasters, providers of over-the-top (OTT) services and video-on-demand (VOD) streaming companies are already distributing HDR content, both in compressed [3]–[5] and uncompressed formats. As HDR content becomes mainstream, its vulnerability to piracy, unauthorized distribution, and malicious modifications also increases [6].

Watermarking has been widely accepted as a method for copyright protection, copy control, and video authentication

due to its proven efficiency [7]. The discordance between HDR mastering parameters such as white peak values, mid-gray tone values, color gamut, TFs, and metadata usage make the designing of an effective watermarking method a challenging task to protect the ownership of HDR content. Moreover, as stated in [8], an important challenge in watermarking for HDR content is to ensure imperceptibility while guaranteeing that the copyright seal survive changes in resolution, frame rate, and dynamic range, e.g., through tone mapping operations (TMO).

In this paper, we propose HDR-IVW-V, an imperceptible-visible watermarking method for the copyright protection of HDR video content. HDR-IVW-V protects HDR video sequences by exploiting the inaccuracies between the way real-life luminance is acquired and how the brightness of a scene is perceived by the HVS when observed on a display. Moreover, HDR-IVW-V profits from the key advantages of visible watermarking, i.e., providing an easy way to visually recognize the embedded copyright seal while minimizing the distortion introduced to the HDR video. Instead of finding an embedding region for each frame of the video, HDR-IVW-V finds an embedding region for only a few selected frames according to the number of segments (or scenes) comprising the sequence. This is based on the observation that a video sequence is a collection of several frames connected in time and usually redundant in terms of visual content within a scene. Therefore, finding the embedding region for each frame is computationally expensive and unnecessary. HDR-IVW-V watermarks an HDR video in two main stages: the first stage detects the segments of the video, while the second stage estimates the spatial location of the embedding region to be used to watermark all frames within a segment.

The rest of the paper is organized as follows. Section II presents an overview of watermarking methods for HDR content. Section III provides a description of the watermarking methodology. Section IV details the scene detection methodology and the scene-based watermarking embedding strategy. Section V presents the performance evaluation results. Finally, Section IV concludes this work.

II. RELATED WORK

A very limited number of works address the problem of copyright protection of HDR video content. In [9], the

This work was supported by the National Autonomous University of Mexico (UNAM) under the DGAPA Postdoctoral Scholarship Program, the PAPIIT-IT101119 Research Project, and the EU Horizon 2020 project, entitled Computer Vision Enabled Multimedia Forensics and People Identification (acronym: IDENTITY, Project ID: 690907).

authors present a watermarking method based on feature map extraction in the RGB space. They employ the Tucker decomposition to extract the feature map of each color channel. The first feature map is used for watermark embedding and is also transformed using the Schur operator to make a stable size relation between coefficients so the robustness to TMO is guaranteed. Nevertheless, this method requires additional processing for watermark revealing and extraction.

Watermarking of HDR images has been explored in more detail in [10]–[13]. However, these works require a secret key and additional processing for watermark revealing and/or extraction, which may contribute to the non-discriminative dissemination of copyrighted material since the final user may not be aware of the media’s ownership.

For the case of Standard Dynamic Range (SDR) content, scene-based watermarking has been explored before. In [14], authors present a scene-based watermarking method that uses support vector machines, where one watermark is extracted per scene. The embedding region is located in the same position in all the frames within a scene. The extracted watermark is the average of the watermarks obtained from each frame. In [15], the Slant transform on the Y-channel is used for watermark embedding. The embedding is done in the middle frequency bands of slant coefficients to enhance imperceptibility. A portion of the watermark is embedded on each frame of the scene using a single watermark per scene. Recently, in [16] a combination of the discrete wavelet transform (DWT) and a scene detector is used for watermarking video content, where the watermark is embedded on a single frame per scene. The watermark seal has a fixed size of 32×32 and is embedded on the frames where a scene change is detected by using the LL sub-band of the DWT.

Visually imperceptible watermarking methods have been previously proposed for SDR content, e.g., [17], [18], and just recently for HDR images [19]. Differently from the watermarking method in [19], HDR-IVW-V avoids watermarking every single frame by segmenting the video into scenes composed of perceptually similar frames. A single embedding region is estimated for each scene, which is refined to attain imperceptibility on each constituent frame.

III. WATERMARKING METHODOLOGY

We embed an imperceptible visible watermark by measuring the inaccuracies between the way luminance is acquired as HDR content and the way the brightness of an HDR scene is perceived by the HVS when displayed on screen [19]. Such measured inaccuracies allows determining a hiding threshold, ξ_{HVS} , which establishes the variation in luma values that a pixel can suffer without being visually perceived by the human eye, according to the OETF used to encode the HDR content [19]. An EOTF, which is the inverse of an OETF, is used to decode the HDR content so it can be displayed on a monitor. For a given EOTF, there is a \mathcal{L}_{target} value, which specifies the range of luminance values (or luma codes) that are best suited to embed an imperceptible visible watermark on the Y-

channel. For a 10-bit system, for example, $\mathcal{L}_{target} \in [64, 198]$ for PQ-encoded HDR content [19].

To embed a watermark on a single frame, three stages are followed, as illustrated in Fig. 1.

A. Embedding Region Selection

To preserve the visual imperceptibility of the embedded watermark, the luminance of the embedding region (ER) should be large enough with uniform texture and luma codes $\in \mathcal{L}_{target}$. After superpixel segmentation [20] on the Y-channel, the superpixel (SP) with the largest area and largest average luma value ($\in \mathcal{L}_{target}$) is selected as the target SP, SP_{target} . The ER is then defined as the largest inscribed region within SP_{target} [21]. Fig. 2 illustrates the target SP and the corresponding ER for a sample Y-channel.

B. Watermark Embedding

We embed a binary watermark (i.e., binary seal) on the Y-channel as follows:

$$ER_{wtm}(i, j) = \begin{cases} ER(i, j) + \Xi_{ER} & \text{if } BN_{seal}(i, j) = 0 \\ ER(i, j), & \text{otherwise,} \end{cases} \quad (1)$$

where $ER_{wtm}(i, j)$ is the watermarked ER at the pixel position (i, j) , BN_{seal} is the pixel value of the binary watermark at the position (i, j) , and Ξ_{ER} is the overall hiding factor of the cover Y-channel, which is computed as a linear combination of the hiding threshold, ξ_{HVS} , of the whole channel, the ER and the regions surrounding the ER [19].

C. Revealing

The embedded watermark can be revealed following one of two approaches:

- 1) Manual color calibration of the HDR screen, which contributes to exaggerating the contrast between the watermarked luma codes and the surrounding pixels.
- 2) Applying a gamma TF to the TMO version of the watermarked HDR frames, which results in a brighter version of the tone-mapped frames, thus revealing the watermark, as exemplified in Fig. 3 and 4.

IV. PROPOSED HDR-IVW-V METHOD

HDR-IVW-V comprises of two main stages: *scene detection* and *scene-based watermarking*. This last stage uses the methodology explained in Section III.

A. Scene detection

Scene detection of HDR video content is challenging because of the absence of shot production rules and the lack of video processing strategies that support and/or exploit HDR luminance and contrast capabilities. According to experimental results [22], TMO can be efficiently used to reduce the dynamic range without significantly modifying the mid and dark tones of an HDR image. For this reason, the Reinhard-TMO (R-TMO) version of the HDR video is used for scene detection. The proposed scene detection scheme is shown in

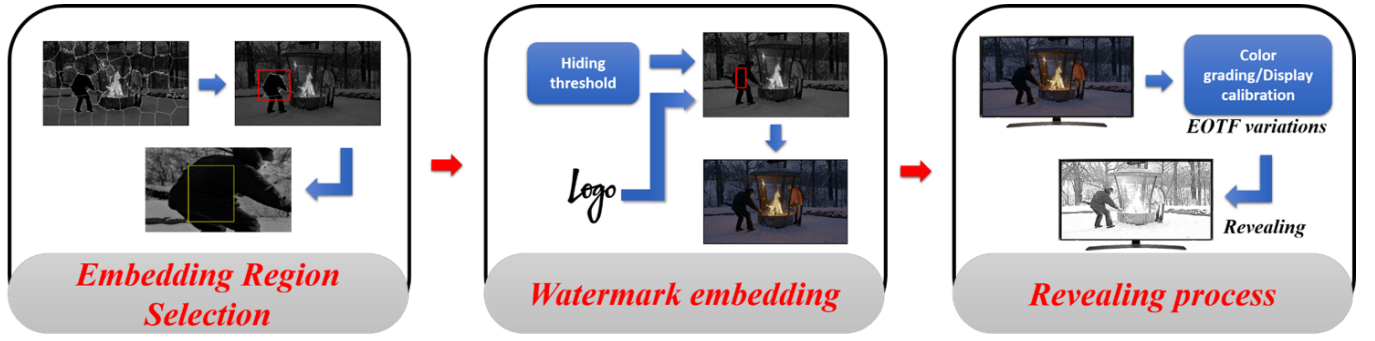


Fig. 1: An overview of the stages to embed a watermark on a single HDR frame.

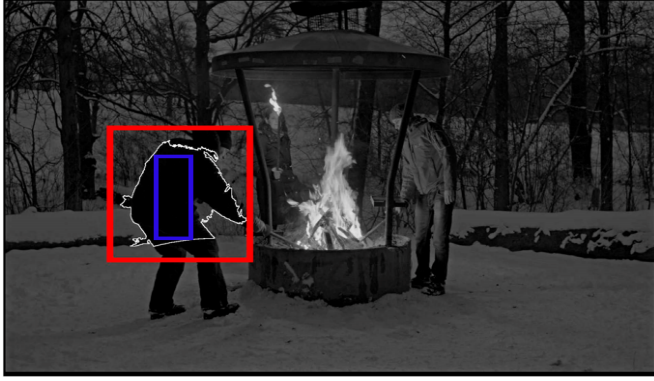


Fig. 2: Example target superpixel (inside the red box) and corresponding ER (blue box).



Fig. 3: Revealed watermark by manual color calibration of the HDR screen.

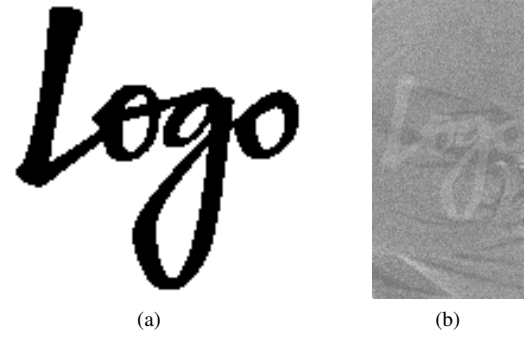


Fig. 4: (a) Binary watermark used in this work. (b) Zoomed-in version of the revealed watermark in Fig. 3

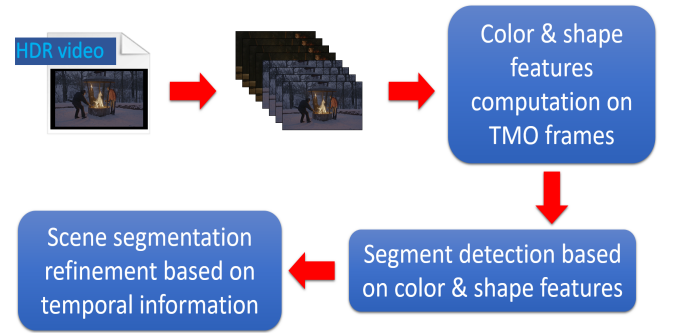


Fig. 5: Proposed scene detection scheme.

Fig. 5. Here, the HDR video is decomposed into k frames and R-TMO is then applied to all frames.

Scene detection aims to find semantic scene changes that are homogeneous in terms of color, shapes and objects. This cross-media space is composed by low level visual features as presented next.

Color features: To describe visual information for scene detection, global frame descriptors provide highly-representative information about the semantic connection among frames. Such global frame descriptors include the Color Structure

Descriptor (CSD) [23] and the Scalable Color Descriptor (SCD) [24]. CSD has been used effectively for shot boundary detection [25], since it can capture distribution of colors and their spatial location as a 64-dimensional vector in the HMMD color space. On the other hand, SCD can capture color and tone variations by computing a histogram in the HSV color space and encoding it using the Haar Transform.

Shape features: To describe the shape of objects on video frames, the Pyramid of Histogram of Oriented Gradients (PHOG) [26] is employed. PHOG has been widely used for recognition of the global shape of objects and human actions.

The scene detection we propose consists in grouping video frames according to temporal connection and visual content similarities. Then, a representative keyframe per segment is selected to estimate the ER of the current video scene. Color and shape features are fused using the early fusion paradigm [27]. Then, scene detection is done by high-level clustering in the complete space (early fused features), as described below.

1) *Scene detection*: A scene is a collection of frames that are close in the feature space and contiguous in time. K -means clustering [28] is used first to find frames close in the feature space, while scene refinement is realized by the partition of frames within a scene that are temporally far from the main segment.

The number of k segments required for K -means clustering is given according to the number of segments found by manual annotation of video segments. K -means partitions the data according to feature similarity and often similar frames can be found in different segments along the video, which may be indeed chronologically far. To solve this problem, we use density based clustering as a post-processing stage.

Let $\{S_1, S_2, \dots, S_k\}$ be a set of video segments obtained by K -means in the complete feature space with $S_i = \{f_1, f_2, \dots, f_m\}$, where i denotes a single video segment in a video with a total of k segments and j is a frame within a particular segment with a time-stamp defined by t_j , such that $\tilde{S}_i = \{t_1, t_2, \dots, t_m\}$. Then, member $t_{(j+1)} \notin \tilde{S}_i$ iff $t_{(j+1)} > (t_j + th)$, where th is a temporal distance threshold set according to the frame rate of the video. Using this temporal refinement for scene detection, new subsets may emerge, such that the final set of chronological ordered scenes is given by $\hat{S} = \{\hat{S}_1, \hat{S}_2, \dots, \hat{S}_q\}$, after a temporal sorting of scenes. The centroid of each cluster/scene is selected and labeled as the *keyframe*.

B. Scene-based watermarking

HDR-IVW-V embeds the watermark on scene basis, as shown in Fig. 6. As stated in section IV-A, scene segmentation is done on the R-TMO version of the HDR frames. However, the watermarking process is done on the Y-channel of the actual HDR video content.

The watermarking of all frames within each scene consists of the following steps:

- 1) The watermarking methodology in Section III is applied to the HDR version of the keyframe of every chronologically ordered scene, \hat{S} .
- 2) The ER location estimated for the keyframe of the scene is taken as the candidate ER (ER_C) for all the other frames within the scene.
- 3) Considering that a scene is given by $S_i = \{f_1, f_2, \dots, f_m\}$. The ER_C of f_j is evaluated to determine if it fulfills the imperceptibility requirements to be the actual ER, i.e., the average luma value of ER_C must be within the range $\mathcal{L}_{target} \in [64, 198]$ for PQ-encoded HDR content. If the average luma value of ER_C does not fulfill this requirement, an ER_C subregion that meets the low-luminance requirement is computed. The

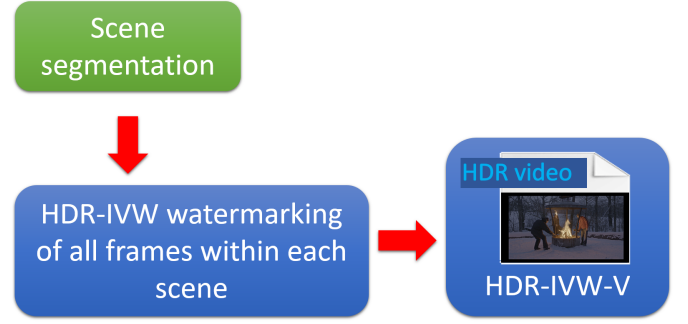


Fig. 6: An overview of the proposed scene-based watermarking embedding process.

ER_C region is binarized such that pixel values larger than 198 are set to 0 and pixel values within the \mathcal{L}_{target} range are set to 1. Then, the largest inscribed region [29] within this binarized region is selected as the ER location.

- 4) Once the ER location is found, the watermark embedding process described in Section III-B is performed.
- 5) Finally, all watermarked frames are concatenated back to obtain final watermarked video sequence.

V. EXPERIMENTAL RESULTS

Two sets of experiments evaluate the imperceptibility and robustness to TMO of the HDR content watermarked using HDR-IVW-V. For these experiments, we use 4 HDR videos encoded with Rec.2020 + PQ-EOTF. Videos are of size 1920×1080 and 25 fps. Fig. 7 shows a sample frame of each video used for evaluation.

The test video sequences are obtained from [30] and [31]. From [30], we use the “BeerFestTeaserClip4000_25_12_P3ct2020_444i” (BeerFest), the “FireplaceTeaserClip4000_24_12_P3ct2020_444i” (FirePlace) and the “ShowGirl2TeaserClip4000_25_12_P3ct2020_444i” (ShowGirl) sequences. “EBU ZurichAthletics2014HD100p_HDREXR” (EBU_100) is obtained from [31].

A. Imperceptibility

Imperceptibility is measured in a subjective and objective way. The Visual Difference Predictor for HDR images, HDR-VDP-2 [32], and the modified Peak Signal to Noise Ratio, mPSNR [33] are used as objective metrics, while MOS (Mean Opinion Score) is used as the subjective metric. HDR-VDP-2 measures the *visibility*, which determines the probability that an observer distinguish differences between two images and the *quality*, which measures the degradation that the original image suffers after watermarking. In this evaluation, the HDR-VDP-2 is given in dB using the Minkowsky distance. On the other hand, the mPSNR measures the quality of the watermarked HDR image frame by frame against the original HDR content, higher mPSNR values are thus desirable. Table



Fig. 7: Sample HDR frames encoded using Rec.2020 + PQ EOTF.

I shows the average HDR-VDP-2 and mPSNR scores obtained per video sequence.

TABLE I: Average HDR-VDP-2 & mPSNR values of the watermarked sequences using HDR-IVW-V.

Video Sequence	HDR-VDP-2 (dB)	mPSNR (dB)
BeerFest	27.2336	54.8274
FirePlace	25.8981	53.4142
ShowGirl	28.5475	53.8429
EBU_100	34.7767	55.5380

According to Table I, HDR-VDP-2 values are within an acceptable range for most of the cases ($\text{HDR-VDP-2}_{dB} > 25.8$ [34]). mPSNR values are high enough to conclude that HDR-IVW-V minimally affects the visual quality of the watermarked videos.

Imperceptibility is evaluated subjectively by means of the MOS metric. Eight observers with various experience levels in HDR imaging have visually inspected each watermarked sequence on a laptop built-in HDR screen with Windows 10 HDR advanced color settings enabled. The observers were asked to assign a score (from 1 to 4) to each sequence according to the detected level of perceptibility without any pause on the playback of the videos, where 1 correspond to full visual perceptibility and 4 to full visual imperceptibility. Results are summarized in Table II

TABLE II: Percentage of MOS scores assigned to each watermarked HDR video sequence.

Video Sequence	Score 1	Score 2	Score 3	Score 4
BeerFest	0	0	0	100
FirePlace	0	0	12.5	87.5
ShowGirl	0	0	0	100
EBU_100	0	0	0	100

The results in Table II confirms that HDR-IVW-V embeds watermarks that are visually imperceptible in most of the cases.

B. Robustness to TMO

TMO is one of the most common transformations that HDR content may suffer. For this reason, six different TMOs are applied to the test watermarked HDR video sequences. [35]. Namely, Clip (C-TM), Gamma (G-TM), Hable (H-TM), Linear (L-TM), Mobius (M-TM) and Reinhard (R-TM). Table

III shows the average Bit Error Rate, in percentage, between the original binary watermark, CRS and the recovered binary watermark, CRS_{rec} , across all frames.

BER values presented in Table III confirm the robustness of HDR-IVW-V to TMO, where the largest BER is only 0.3513%. Let us recall that HDR-IVW-V embeds the watermark in low luminance regions and TMOs do not tend to be aggressive on low luminance tones.

VI. CONCLUSION

In this paper, we proposed HDR-IVW-V for the visually imperceptible copyright protection of HDR video content. To reduce the computational processing of estimating the most suitable region for watermark embedding, HDR-IVW-V computes the embedding region on a single keyframe per scene instead of on every single frame. The MOS, mPSNR and HDR-VDP-2 metrics were used to evaluate imperceptibility and visual quality preservation of the watermarked content, confirming high imperceptibility. Additionally, overall BER values obtained confirm the robustness of HDR-IVW-V to TMO.

REFERENCES

- [1] SMPTE, "SMPTE ST 2084: High Dynamic Range Electro-Optical transfer function of mastering reference displays," 2014.
- [2] ITU-R Rec BT.2100, "Image Parameter values for High Dynamic Range Television for Use in Production and International Programme Exchange, 2016," <https://www.itu.int/rec/R-REC-BT.2100>.
- [3] Y. Zhang, M. Naccari, G. Agraftiotis, M. Mrak, and D. Bull, "High dynamic range video compression exploiting luminance masking," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 26, no. 5, pp. 950–964, 2015.
- [4] K. Perez-Daniel and V. Sanchez, "Luma-aware multi-model rate-control for hdr content in hevc," in *IEEE International Conference on Image Processing*, 2017, pp. 1022–1026.
- [5] Y. Zhang, E. Reinhard, and D. Bull, "Perception-based high dynamic range video compression with optimal bit-depth transformation," in *IEEE International Conference on Image Processing*, 2011, pp. 1321–1324.
- [6] Interactive Advertising Bureau, "What is an untrustworthy supply chain costing the U.S. Digital Advertising Industry?," https://www.iab.com/wp-content/uploads/2015/11/IAB_EY_Report.pdf, 2015.
- [7] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, 2018.
- [8] M. U. Celik and J. Haitsma, "The challenges of forensic watermarking uhd/hdr content," in *SMPTE Annual Technical Conference and Exhibition*, 2016, pp. 1–7.
- [9] T. Luo and et al, "Robust High Dynamic Range color image watermarking method based on feature map extraction," *Journal of Signal Processing*, vol. 155, pp. 83–95, 2019.

TABLE III: Average BER (%) between the recovered binary watermark after TMO and the original watermark.

Video Sequence	C-TM	G-TM	H-TM	L-TM	M-TM	R-TM
BeerFest	0.0820	0.0842	0.0625	0.2174	0.0820	0.0726
FirePlace	0.2836	0.2953	0.2477	0.3513	0.2836	0.3051
ShowGirl	0.2284	0.2329	0.1984	0.3003	0.2284	0.2376
EBU_100	0.1788	0.1790	0.1783	0.1746	0.1788	0.1790

- [10] F. Y. Bakhsh and M. E. Moghaddam, "A robust HDR images watermarking method using artificial bee colony algorithm," *JISA*, vol. 41, pp. 12 – 27, 2018.
- [11] E. Maiorana and P. Campisi, "Multi-bit watermarking of High Dynamic Range images based on perceptual models," *Security and Communication Networks*, vol. 9, no. 8, pp. 705–720, 2016.
- [12] V. Solachidis, E. Maiorana, P. Campisi, and F. Banterle, "HDR image watermarking based on bracketing decomposition," in *18th International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1–6.
- [13] E. Maiorana and et al, "Robust multi-bit watermarking for HDR images in the Radon-DCT domain," in *ISPA*, 2013, pp. 284–289.
- [14] S. Yen, Hsiao-Wei Chang, Chia-Jen Wang, P. S. Wang, and Mei-Chueh Chang, "A scene-based video watermarking technique using svms," in *19th International Conference on Pattern Recognition*, 2008, pp. 1–4.
- [15] K. Meenakshi, C. S. Rao, and K. S. Prasad, "A scene based video watermarking using slant transform," *IETE Journal of Research*, vol. 60, no. 4, pp. 276 – 287, 2014.
- [16] D. Shukla and M. Sharma, "Robust scene-based digital video watermarking scheme using level-3 dwt: Approach, evaluation, and experimentation," *Radioelektronika*, vol. 61, no. 1, pp. 5 – 21, 2018.
- [17] S. C. Chuang, C. H. Huang, and J. L. Wu, "Unseen visible watermarking," in *ICIP*, Sept 2007, vol. 3, pp. III – 261–III – 264.
- [18] L. Zhaotian and et al, "Robust unseen visible watermarking for depth map protection in 3d video," in *AIHMSP*, 2018, pp. 322–332, Springer.
- [19] K. Perez-Daniel, F. Garcia-Ugalde, and V. Sanchez, "Imperceptible Visible Watermarking of HDR Content based on HVS capabilities," *Submitted to IEEE Transactions on Image Processing*, 2019.
- [20] W. Murong, L. Xiabi, G. Yixuan, M. Xiao, and Q. S. Nouman, "Superpixel segmentation: A benchmark," *Signal Processing: Image Communication*, vol. 56, pp. 28–39, 2017.
- [21] C. Knauer, L. Schlupf, J. M. Schmidt, and H. R. Tiwary, "Largest inscribed rectangles in convex polygons," *Journal of Discrete Algorithms, Elsevier*, vol. 13, pp. 78 – 85, 2012.
- [22] E. Reinhard, M. Stark, P. Shirley, and Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267 – 276, 2002.
- [23] D. S. Messing and et al, "The MPEG-7 colour structure descriptor: image description using colour and local spatial information," in *ICIP*, 2001, vol. 1, pp. 670–673.
- [24] W. Yong-ge and P. Sheng-ze, "Research on image retrieval based on scalable color descriptor of mpeg-7," in *Advances in Control and Communication*, Dehuai Zeng, Ed. 2012, pp. 91–98, Springer.
- [25] A. B. Abdelali and et al, "A study of the color structure descriptor for shot boundary detection," *IJSTACCE*, vol. 3, no. 1, pp. 956 – 971, 2009.
- [26] J. Wang, P. Liu, M. F. H. She, A. Kouzani, and S. Nahavandi, "Human action recognition based on pyramid histogram of oriented gradients," in *IEEE ICSMC*, 2011, pp. 2449–2454.
- [27] C. Snoek and et al, "Early versus late fusion in semantic video analysis," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, 2005, MULTIMEDIA '05, pp. 399–402.
- [28] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651 – 666, 2010.
- [29] C. Knauer, L. Schlupf, J. M. Schmidt, and H. R. Tiwary, "Largest inscribed rectangles in convex polygons," *Journal of Discrete Algorithms, Elsevier*, vol. 13, pp. 78 – 85, 2012.
- [30] Stuttgart University, "High Dynamic Range database Graded MPEG Teaser clips," <https://hdr-2014.hdm-stuttgart.de/>.
- [31] EBU-SVT, "Public HDR test sequences," https://tech.ebu.ch/EBU_SVT_Public_Test_Sequences.
- [32] R. Mantiuk, K. J. Kim, and et al, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, 2011.
- [33] J. Munkberg, P. Clarberg, J. Hasselgren, and T. Akenine-Möller, "High dynamic range texture compression for graphics hardware," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 698–706, July 2006.
- [34] K. J. Kim and et al, "A comparison of three image fidelity metrics of different computational principles for JPEG2000 Compressed abdomen CT images," *IEEE Transactions on Medical Imaging*, vol. 29, no. 8, pp. 1496–1503, 2010.
- [35] Multimedia Framework., "FFMPEG v4.0.2," <https://www.ffmpeg.org/>.