Development of a Robust Watermarking Scheme Against Recompression Attacks for AV1 Videos

Ezequiel Magnaye
College of Engineering and Agro-industrial Technology
University of the Philippines Los Baños
Los Baños, Philippines
eamagnaye@up.edu.ph

Melvin Ilang-Ilang
College of Engineering and Agro-industrial Technology
University of the Philippines Los Baños
Los Baños, Philippines
mcilangilang3@up.edu.ph

***Abstract*—This paper proposes an image-on-video watermarking scheme for videos compressed with the AOMedia Video 1 (AV1) format that is made robust against recompression attacks using Discrete Cosine Transform (DCT) techniques. A Two-Dimensional DCT-transformed watermark was first embedded in the high frequency region of each similarly transformed frame of the test sequences. The base watermarked sequences and their attacked copies were then subjected to extractions to determine the robustness of the embedded payload. On average, the embedding algorithm created watermarked videos with fair levels of distortion in terms of Peak Signal-to-Noise Ratio (PSNR) and Root Mean Squared Error (RMSE), and excellent perceptual quality in terms of Video Multimethod Assessment Fusion (VMAF) scores. Upon extraction, the retrieved images had high fidelities with the designed watermark, as computed in terms of Structural Similarity Index Measure, even with those from the attacked videos. The Embedding Capacity Ratio per Frame (ECRF) of the scheme was calculated to be within the typical range of values of watermarking for security and documentation purposes.**

Keywords—AV1, DCT, H264, H265, robustness, recompression attack, video watermarking, VMAF

I. Introduction

Data communication features such as instant messaging, and media sharing between end users are a main focus in the creation of modern computing devices with faster and more powerful processors. To improve these features, researchers have been developing advanced software applications as well involving algorithms designed for a more efficient encoding and decoding of data streams. These reduce file sizes which provide more allocation within a transmission bandwidth [1]. Video transmission in particular benefit more with compression coding due to their greater file sizes compared to image and audio files, and the innate redundancies of information carried within and across video frames. With these considerations, a coding format called AOMedia Video 1 (AV1) was created by the eponymous non-profit organization, Alliance for Open Media, to be the new standard in video transmission. AV1 was released to the public in 2018 and has since followed the open-source and royalty-free system of its predecessor, Google's VP9 codec. It is in direct competition with Advanced Video Coding (AVC a.k.a. H264) and High Efficiency Video Coding (HEVC a.k.a. H265), both created under MPEG LA, to prevent licensing issues that are usually attached to software developed by private companies. AV1 was found to have an increase of 17% in average bitrate savings in video compression compared to VP9 [2], and 53% and 21% when compared to H264 and H265 [3], respectively. By coupling its proven performances and dedication to license-free use with the development of advanced processors and modern content delivery networks, AV1 is projected to become a forerunner in mainstream implementations of encoding video on demand and live streaming contents [4]. However, new opportunities arise for digital piracy to exploit this codec, and so digital watermarking is often used by creators to assert their copyright over their content [5]. In this study, a robust, blind image-on-video watermarking scheme is proposed to certify ownership of AV1 files.

So far, there have been no publicly announced method for AV1 watermarking, but [6] has proposed a text-on-video codec-level steganography method. When the text payload is embedded and the cover video file size increases by at least 13 kilobytes, the watermark becomes irretrievable and the video results with noticeable hazy artifacts. Embedding in the frequency domain offers the possibility of a more effective watermarking scheme. This is especially true when using the Discrete Cosine Transform (DCT), one of the most common transformation techniques used in video compression, on video files having codecs with intrinsic DCT algorithms like AV1. DCT was first applied to H264 videos in the early 2000s with watermarks already being embedded by modifying AC coefficients of I-frame luma blocks [7]. Such watermarks could be preprocessed to have a smaller bit size which would minimize bitrate increase in the cover file. This could be as simple as a binarization of a grayscale watermark [8] or an N×N integer DCT could be applied to it for further compression [9]. The use of spread spectrum for embedding in the frequency domain was also suggested to mask a watermark's location against deliberate attacks to the video [10]. The DCT have been also paired with other transforms either to get a boost in robustness, or to be complemented on the same file where one works with a robust watermark while the other is fragile [11]. The 2010s then saw the release of H265, and with it a Discrete Sine Transform technique was paired with DCT to transform the remaining error of 4×4 intra-predicted blocks [12]. A Discrete Wavelet Transform inside a DCT was also discovered to result with the shortest processing time for watermark embedding [13]. Even while considering these advantages of pairing other transformation techniques, this paper focuses on a DCT-based technique, not only to provide a base from which other AV1 watermarking schemes could be compared to but also to make use of the luma components of the videos. DCT is ideal since it could handle modifications in luminance alone with greater efficiency and accuracy [14].

979-8-3315-6946-4/25/$31.00 ©2025 IEEE

The rest of the paper is structured as follows: Section II gives a brief overview of the AV1 codec, Section III elaborates on the proposed scheme, Section IV shows the results of the watermarking and their relevant calculations, and Section V concludes the findings of the study.

II. AV1 Compression Codec

The AV1 codec follows the general flow of video compression which involves a hybrid block-based encoding-decoding process. When compared with other codecs, it uses downscaling and denoising before prediction to cut down redundant data on the input bitstream, and adds upscaling before passing through the filter to maintain the quality for prediction [15]. Each video frame is partitioned to blocks with a maximum possible size of 128×128 from out of 10 variations, each of which is then scanned for the values it contains [16]. The redundancies that will be observed are either spatial or temporal in nature. In intra-frame prediction of values, neighboring and distant pixels could be called at once with respect to a reference pixel, like in the case of image processing. For inter-frame prediction, the existence of such pixels is also checked in preceding and succeeding frames. Error signals are then obtained by subtracting the predictions from the input bitstream per block. These undergo transformation and quantization before being used in a multi-symbol arithmetic entropy coding to become a part of the output bitstream. AV1 uses 16 transform types which are combinations of DCT, Asymmetric Discrete Sine Transform (ADST), FlipADST, and Identity Transform, and six quantization parameters of (Y', Cb, Cr)×(DC, AC). Since the compression process works block by block, the transformed and quantized signals also undergo inverse quantization and then inverse transformation. They then pass through a filter to be scanned once again with the input bitstream as new predictions for the next block.

III. The Proposed Watermarking Scheme

The watermarking scheme consisted of two scripts, one for embedding and another for extraction, both of which worked with a predefined monochromatic stylized text logo and a set of 11 1920×1080 videos from the Xiph.org test media repository compressed with SVT-AV1. Fig. 1 shows the flow of the embedding process of the proposed scheme. A grayscale image of 64×64-pixel size was first created and was then preprocessed by binarization to obtain the monochrome image, leading to a reduction of payload size by about 24.70%. The cover videos were then split into their I-frames, with each being converted into the YCbCr color space from which their luma components, Y, were isolated. A 2D-DCT technique was then applied to both the preprocessed watermark and the masked, isolated luma components so that the DCT values of the former could be inserted into the high frequency regions of the latter. Equation (1) shows the formula for the 2D-DCT technique [17].

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}, \begin{matrix} 0 \le p \le M-1 \\ 0 \le q \le N-1 \end{matrix} \quad (1)$$

In this equation, A is the input image represented in spatial domain, B is the output image represented in the frequency domain, M and N are the respective row and column sizes of A, and P and Q are the respective row and column sizes of B.

After insertion of the watermark's DCT values to the lower right region of the masked, transformed matrices, inverse 2D-DCT was applied to the result to produce the corresponding luma components. Equation (2) shows the formula for the inverse 2D-DCT technique with the same parameters used for (1) [17].

$$A_{mn} = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \alpha_p \alpha_q B_{pq} \cos\frac{\pi(2m+1)p}{2M} \cos\frac{\pi(2n+1)q}{2N}, \begin{matrix} 0 \le p \le M-1 \\ 0 \le q \le N-1 \end{matrix} \quad (2)$$

The component Y is combined with components Cb and Cr to reform the image which could then be converted back into RGB. The process is recursive until all frames have been embedded and are ready for compiling into one sequence again. On the other hand, watermark extraction works in reverse, although it is not as exhaustive as embedding since any one frame from the AV1 video file carries a watermark. Fig. 2 shows the sequence for this process.

For this study, all available frames per test sequence were considered for extraction to determine how the watermark fared throughout the entirety of the videos. Each watermarked video file was split into I-frames again, and each frame was color-converted into the YCbCr space. Upon isolation of the luma component, it was applied with the 2D-DCT technique resulting with a DCT matrix. DCT values from the embedding region of this transformed component were copied into a new, separate matrix which was then applied with the inverse 2D-DCT technique. The resulting image became the luma component of the retrieved watermark, and its color conversion produced the RGB watermark for that frame.

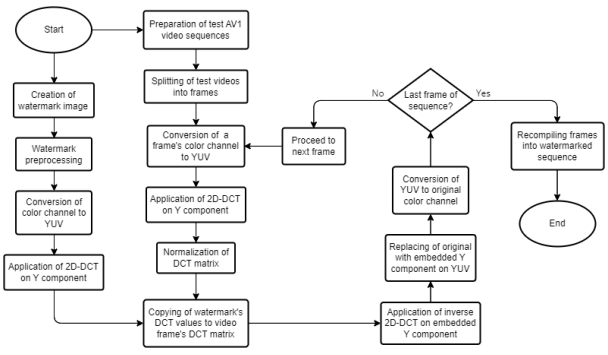


Fig. 1. Watermark embedding flowchart.

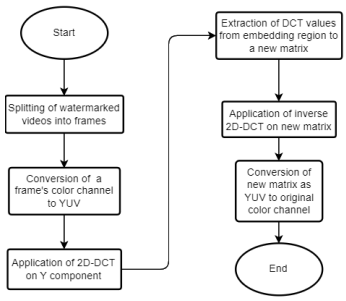


Fig. 2. Watermark extraction flowchart.

The experiment was conducted using a laptop computer running Windows 11 Home Version 24H2 OS with an Intel Core i5-8250U processor operating at the maximum overclocked speed of 3.4 GHz.

IV. Experimental Results

The proposed watermarking scheme was evaluated through the three trade-off properties of watermark efficiency: imperceptibility, robustness, and embedding capacity [18]. Imperceptibility is the watermark's property to be invisible to inspection using human senses, and to be near undetectable to basic modern tools. Robustness is the measure of integrity of the watermark after the medium in which it was placed was subjected to modifications. Lastly, the embedding capacity is also considered such that the addition of the watermark to the file would not contribute a significant increase to the output file size compared to that of the original.

A. Imperceptibility Tests

The scheme's imperceptibility was first analyzed using the Peak Signal-to-Noise Ratio (PSNR) to determine the distortions in the watermarked videos with respect to their covers. This measures the power of the cover frame over the power of the watermarked frame evaluated in terms of the decibel scale [19], [20]. Equations (3) and (4) show how to calculate the PSNR for each video frame.

$$PSNR = 10 * \log_{10}\left(\frac{MAX_A^2}{MSE}\right) \tag{3}$$

$$MSE = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}[A(i,j)-B(i,j)]^2}{m \times n} \tag{4}$$

For the PSNR, MAX_A is the maximum pixel value of the cover frame, MSE is the mean squared error of the cover and watermarked frames, m and n are the dimensions of the frames, i and j are the respective pixel row and column indices, A(i,j) is the pixel value at image A's ith row and jth column, and B(i,j) is the pixel value at image B's ith row and jth column. Root Mean Squared Error (RMSE) is also used as an estimator for the degree the watermarked frame's pixel values deviate from that of the cover frame. It is obtained by simply getting the square root of (4) which results to a unitless pixel value number. The ranges of the watermarked videos' PSNR and RMSE values are presented in Figs. 3 and 4, respectively.

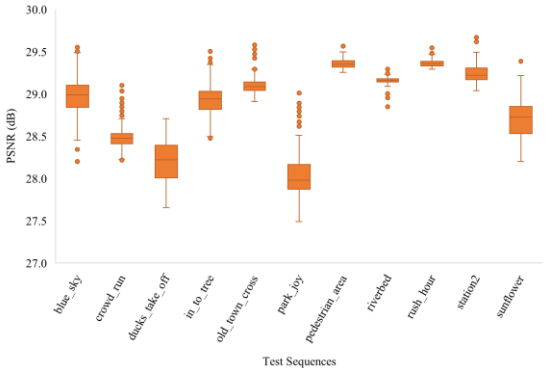


Fig. 3. PSNR levels of the watermarked test sequences.

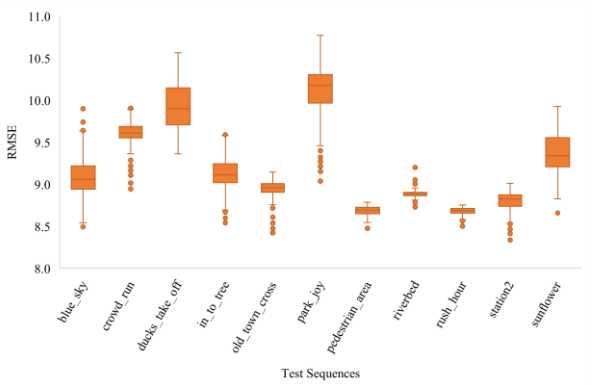


Fig. 4. RMSE values of the watermarked test sequences.

A higher PSNR value means a watermarked frame's visual fidelity is closer to the original frame, while a lower one indicates a greater introduction of noise. Comparing the videos' RMSE values to their PSNR values, the trend follows the relation of a low RMSE corresponding to a high PSNR and vice versa. The test sequences exhibited PSNR levels ranging from 27 dB to 30 dB which, basing on [21], correspond to a perceptual quality of "fair". While 30 dB to 50 dB is the usual range of PSNR values for lossy image compression, 20 dB to 30 dB is still acceptable [22]. However, it is also noted that videos having values below 30 dB show distortions that are noticeable to the Human Visual System (HVS) [23]. On the other hand, since a channel in the RGB color space has 256 possible values, it means that on average the watermarked videos have their pixels deviating from their original color value by 8 to 11. The deviations generally do not make drastic changes in hue but only make minimal effects in a pixel's brightness and saturation. These objective metrics, particularly PSNR, are not reliable as perceptual quality metrics as they do not provide exact correspondence with HVS in video quality [24]. PSNR is solely calculated on a per frame basis while HVS also considers the temporal properties of the cover video [25].

The Video Multimethod Assessment Fusion (VMAF) metric, being based on the HVS, supplements PSNR and RMSE. It was developed to properly correlate objective measurements of video quality with human perception by training a support vector machine model with a dataset evaluated for multiple metrics. It works on three core features of video quality: information fidelity loss, video detail loss, and mean absolute pixel difference between adjacent frames [26], [3]. Each feature is scored by how similar they are with the respective features of the cover videos. The scores from the three features are weighted and then combined to produce the total Differential Mean Opinion Score which ranges from zero to 100, where the latter is equal to a perceptual quality equivalent to that of the reference video. Fig. 5 shows the watermarked videos' ranges of perceptual quality in terms of VMAF. The lowest mean VMAF score attained among the test videos was 88.053. This is still considered as being "excellent" since the score falls between 80 and 100 according to [27]. The only instances where the quality went down were from sequences "ducks_take_off", "riverbed", and "sunflower" when they reached VMAF scores in the 70s which correspond to only "good" ratings. On the other hand, sequences "blue_sky", "crowd_run", "park_joy",

"pedestrian_area", "riverbed", and "rush_hour" were all able to reach a perfect VMAF score for certain frames.

B. *Robustness Tests*

The scheme's robustness was then analyzed using Structural Similarity Index Measure (SSIM) which is another metric for measuring the similarity of two images, but this differs from PSNR in that the frame to be tested was subjected to degradation attacks. SSIM measurements are unitless, and they have their best scores as positive unity, indicating complete similarity between the extracted and the original watermarks. Equation (5) is used to calculate the SSIM of the pairs of watermarks to measure the fidelity of the extracted ones to the reference in terms of luminance, contrast, and structural integrity [28].

$$SSIM(x,y) = [l(x,y)]^{\alpha} * [c(x,y)]^{\beta} * [s(x,y)]^{\gamma} \qquad (5)$$

Calculating the SSIM requires l, c, and s which are the luminance, contrast, and structure comparisons, respectively, of the two images, x and y. The parameters α, β, and γ are parameters of positive values used for adjusting comparisons. Since every single frame of the test videos was processed for extraction, the SSIM values were also computed on a frame-by-frame basis. Fig. 6 shows the ranges of SSIM values of the unmodified watermarked videos, and Fig. 7 presents the pairs of watermarks with the greatest and least similarities with the original for each sequence.

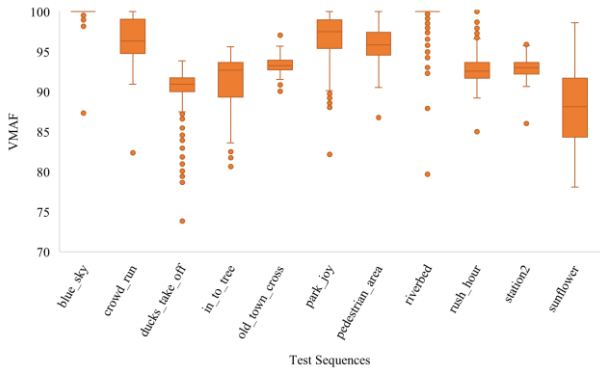


Fig. 5. Total VMAF scores of the watermarked test sequences.

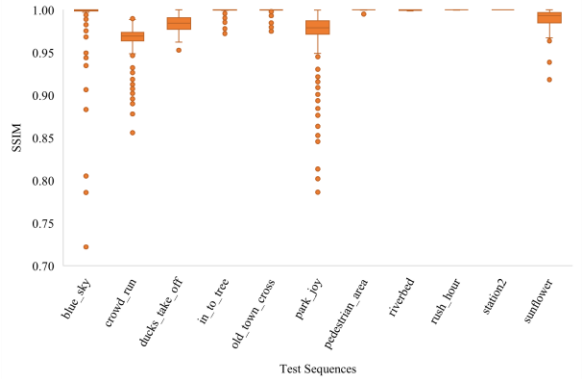
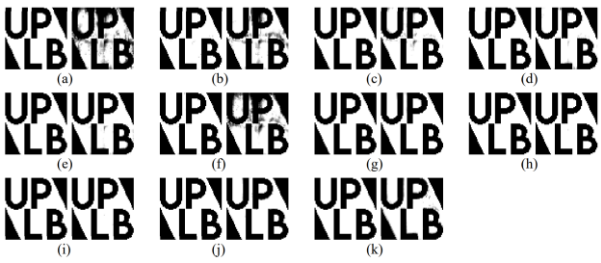


Fig. 6. SSIM values of the extracted watermarks from the test videos.



(a) blue_sky, (b) crowd_run, (c) ducks_take_off, (d) in_to_tree, (e) old_town_cross, (f) park_joy, (g) pedestrian_area, (h) riverbed, (i) rush_hour, (j) station2, and (k) sunflower.

Fig. 7. Extracted watermarks with the highest and lowest SSIM values from the unmodified test sequences.

The extracted watermarks exhibited near perfect similarities with the original, having average values of at least 0.9657. Even though the sequences "crowd_run", "ducks_take_off", and "park_joy" were not able to achieve perfect similarity with any of their watermarks, they were able to come close to positive unity having maximum values from 0.9929 to 0.9999. The least similar watermark extracted from a single frame belonged to "blue_sky" with a value of 0.7220. This watermark has the greatest distortion among all extractions from the unmodified test sequences although, as seen from Fig. 7, the introduction of dark areas is not enough to obscure the important details of the watermark. It must also be noted that "blue_sky" has the greatest deviation of SSIM value from the average with 0.2705, while "pedestrian_area", "riverbed", "rush_hour", and "station2" have deviations that are less than 0.01.

Greater distortions were expected for the extracted watermarks from the test videos that were first converted into another compression codec using FFmpeg and then back into AV1. The recompression tests used in this study are through H264 and H265 since these codecs are still widely used today. Fig. 8 shows the ranges of SSIM values of the extracted watermarks from the recompressed videos. The watermarks having the greatest and least similarities were selected from each type of recompression attack and displayed in Fig. 9. The values from Fig. 8 shows a trend where the extracted watermarks from the test videos subjected to AV1-H265-AV1 recompression attack have greater similarities with the original watermark than their counterparts through the H264 recompression. This suggests that the proposed watermarking scheme has greater compatibility with the H265 codec as compared to H264 when the watermarked AV1 video file is converted through them. The sequences were still able to reach maximum values of at least 0.8359 but only "station2"

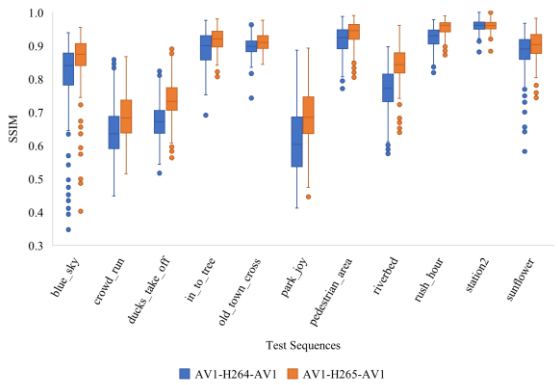


Fig. 8. SSIM values of the extracted watermarks from videos subjected to recompression attacks.

(a) station2 and blue_sky through H264, and (b) station2 and blue_sky through H265, respectively.

Fig. 9. Extracted watermarks with the highest and lowest SSIM values among all test sequences subjected to each recompression attack.

through H265 recompression was able to achieve perfect similarity. The average SSIM values from recompressed files range from 0.6177 to 0.9622. Meanwhile, their lowest values range from 0.3470 to 0.8833, with sequence "blue_sky" having the watermarks with the least similarities for both recompression attacks due to the resampling of color values of the dark canopy from frames near the end of the said watermarked video.

C. Correlation between Perceptual Quality and Robustness

To determine the correlation between the per-frame perceptual quality scores of the test videos and the robustness of their respective sets of extracted watermarks, Pearson correlation analyses were used and the resulting coefficients are presented in Table I. A positive correlation coefficient means that the SSIM values of an attacked test video's extracted watermarks follow the trend of the VMAF scores of the unmodified videos they were originally from. On the other hand, a negative coefficient implies an inverse relationship wherein the proposed scheme was not able to balance the imperceptibility and robustness properties of the watermark. It could mean that the scheme was able to either secure the watermark in an attacked video at the expense of its visual quality or ensure its fidelity with the original by sacrificing the retrievability of the embedded data.

After basing all test videos on the general guidelines for correlation interpretation from [29], an absolute coefficient of value greater than 0.5 was obtained only from sequence "in_to_tree", implying strong degrees of correlation for SSIM values produced by recompression attacks. Next, "rush_hour" exhibited moderate degrees of correlation having an absolute correlation coefficient between 0.3 and 0.5. All other sequence-attack pairs showed weak correlation, especially with sequence "blue_sky" having a coefficient that is infinitesimal which could be approximated to no correlation. This near-zero degree implies that the recompression attacks are strong enough to make the embedded watermarks irretrievable.

TABLE I. PEARSON CORRELATION COEFFICIENTS OF VMAF SCORES OF THE TEST SEQUENCES AND SSIM VALUES OF THE EXTRACTED WATERMARKS FROM THE RECOMPRESSED VIDEOS

Test Sequence	ρ
blue_sky	0.00334
crowd_run	0.23624
ducks_take_off	0.21682
in_to_tree	0.66544
old_town_cross	0.08644
park_joy	-0.16929
pedestrian_area	-0.26242
riverbed	-0.19883
rush_hour	-0.33920
station2	-0.04410
sunflower	-0.25763

D. Embedding Capacity Calculation

Finally, the Embedding Capacity Ratio per Frame (ECRF) was calculated using properties of the original preprocessed watermark and the test sequences in general to determine the embedding capacity of the watermarking scheme. The ECRF is calculated using (6) where $D_{embedded}$ is the bit size of embedded watermark data and F_{size} is the number of cover frame pixels.

$$ECRF = \frac{D_{embedded}}{F_{size}} \times 100\% \quad (6)$$

Since the proposed watermarking scheme used the same watermark for embedding and it solely focused on AV1 videos with 1920×1080 resolution, calculating for the ECRF was straightforward and resulted with 0.0482 bits per pixel (bpp) for all test sequences. The reported average ECRF values for image watermarking are within the range of 0.001 bpp, for highly secure watermarking, and 0.1 bpp, for watermarking as annotation [30], [31]. The proposed scheme's ECRF lies between these two points, signifying a fair mix of imperceptibility and embedding capacity but leans more towards the former since the value is closer to 0.001 bpp than the other average. Alternatively, the ECRF could also be interpreted as the scheme using only 4.8187% of the frame area to embed the watermark.

V. CONCLUSION

The rise in mainstream use of the AV1 compression format is brought about by the need for a royalty-free, open-source video codec that is also more efficient than other formats in creating and delivering high quality videos with low file sizes. However, this entails increased risks of digital piracy related to video content. To counter this type of copyright infringement, a blind image-on-video watermarking scheme in the frequency domain for AV1 videos was developed using DCT techniques. The proposed scheme involves replacing the high frequency coefficients of a video's transformed intra-coded luma frame components with a preprocessed watermark image's DCT equivalent.

The watermark was embedded first in 11 SVT-AV1-encoded videos with 1920×1080 resolution and the resulting test sequences have average RMSE values from 8.6810 to 10.0998, and average PSNR values from 28.0485 dB to 29.3595 dB which are considered as indicators of having fair or acceptable levels of distortion. On the other hand, with average VMAF scores of 88.053 to 99.899, the watermarked videos are considered as having excellent levels of perceptual quality when compared with the original videos. The test sequences were then subjected to separate recompressions through H264 and H265 to test the scheme's robustness against such attacks. The extracted watermarks and their corresponding SSIM values showed that the proposed scheme still produced clear watermarks even after recompression, with attacks made through H265 resulting with extractions of greater similarities with the original watermark than those from H264. The ECRF value of the watermarking scheme is 0.0482 bpp which lies between the average ranges for secure and annotative watermarking use cases.

REFERENCES

[1] N. Barman, M. Martini, and Y. Reznik, "Bjontegaard Delta (BD): A tutorial overview of the metric, evolution, challenges, and recommendations," unpublished, 2024, DOI:10.13140/RG.2.2.18622.05444.

[2] D. Grois, T. Nguyen, and D. Marpe, "Performance comparison of AV1, JEM, VP9, and HEVC encoders," Applications of Digital Image Processing XL. SPIE, p. 120, Feb. 08, 2018. DOI: 10.1117/12.2283428.

[3] M. Uhrina, L. Sevcik, J. Bienik, and L. Smatanova, "Performance comparison of VVC, AV1, HEVC, and AVC for high resolutions," Electronics, vol. 13, no. 5, p. 953, Mar. 2024, DOI: 10.3390/electronics13050953.

[4] I. Trow, "AV1: Implementation, performance, and application," SMPTE Mot. Imag. J, vol. 129, no. 1, pp. 51–56, Jan. 2020, DOI: 10.5594/jmi.2019.2951827.

[5] Q. Tang, Y. Li, Q. Wang, W. He, and X. Peng, "A dual blind watermarking method for 3D models based on normal features," Entropy, vol. 25, no. 10, p. 1369, Sep. 2023, DOI: 10.3390/e25101369.

[6] L. Catania, D. Allegra, O. Giudice, F. Stanco, and S. Battiato, "Introducing AV1 codec-level video steganography," Lecture Notes in Computer Science. Springer International Publishing, pp. 284–294, 2022. DOI: 10.1007/978-3-031-06427-2_24.

[7] G. Qiu, P. Marziliano, A. Ho, D. He, and Q. Sun, "A hybrid watermarking scheme for H.264/AVC video," Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, pp. 865-868 Vol.4, 2004. DOI: 10.1109/icpr.2004.1333909.

[8] J. Zhang and A. Ho, "Robust digital image-in-video watermarking for the emerging H.264/AVC standard," IEEE Workshop on Signal Processing Systems Design and Implementation, 2005. IEEE, pp. 657–662. DOI: 10.1109/sips.2005.1579947.

[9] J. Zhang, A. Ho, G. Qiu, and P. Marziliano, "Robust video watermarking of H.264/AVC," IEEE Trans. Circuits Syst. II, vol. 54, no. 2, pp. 205–209, Feb. 2007, DOI: 10.1109/tcsii.2006.886247.

[10] S. Bouchama, H. Aliane, and L. Hamami, "Watermarking techniques applied to H264/AVC video standard," 2010 International Conference on Information Science and Applications. IEEE, pp. 1–7, 2010. DOI: 10.1109/icisa.2010.5480288.

[11] B. Lei, K. Lo, and H. Lei, "A blind and robust watermarking scheme for H.264 video," 2010 International Conference on Communications, Circuits and Systems (ICCCAS). IEEE, pp. 421–424, Jul. 2010. DOI: 10.1109/icccas.2010.5581966.

[12] S. Gaj, A. Sur, and P. Bora, "A robust watermarking scheme against re-compression attack for H.265/HEVC," 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG). IEEE, pp. 1–4, Dec. 2015. DOI: 10.1109/ncvpripg.2015.7490065.

[13] W. El-Shafai, E. El-Rabaie, M. El-Halawany, and F. El-Samie, "Efficient multi-level security for robust 3D color-plus-depth HEVC," Multimed Tools Appl, vol. 77, no. 23, pp. 30911–30937, Jun. 2018, DOI: 10.1007/s11042-018-6036-z.

[14] W. Mustafa et al., "Image enhancement based on discrete cosine transforms (DCT) and discrete wavelet transform (DWT): A review," IOP Conf. Ser.: Mater. Sci. Eng., vol. 557, no. 1, p. 012027, Jun. 2019, DOI: 10.1088/1757-899x/557/1/012027.

[15] C. Feldmann, Video coding basics – How is this so efficient? (May 20, 2022). Accessed: Jul. 24, 2024. [Online video]. Available: https://www.youtube.com/watch?v=LDeL7-49qm4

[16] T. Terriberry, "The AV1 video codec," Linux.conf.au 2019 Convention (Jan. 21-25, 2019). Accessed: Jul. 24, 2024. [Online video]. Available: https://archive.org/details/lca2019-The_AV1_Video_Codec

[17] MathWorks, "Image transforms," MathWorks. [Online]. Available: https://www.mathworks.com/help/images/image-transforms.html

[18] M. Dalal and M. Juneja, "A secure video steganography scheme using DWT based on object tracking," Information Security Journal: A Global Perspective, vol. 31, no. 2, pp. 196–213, Mar. 2021, DOI: 10.1080/19393555.2021.1896055.

[19] A. Elrowayati, M. Alrshah, M. Abdullah, and R. Latip, "HEVC watermarking techniques for authentication and copyright applications: Challenges and opportunities," IEEE Access, vol. 8, pp. 114172–114189, 2020, DOI: 10.1109/access.2020.3004049.

[20] S. Gupta, G. Kalia, and P. Sondhi, "Video steganography using discrete wavelet transform and artificial intelligence," International Journal of Trend in Scientific Research and Development, 3, 4, pp. 1210-1215, 2019.

[21] J. Gross, J. Klaue, H. Karl, and A. Wolisz, "Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming," Computer Communications, vol. 27, no. 11, pp. 1044–1055, Jul. 2004, DOI: 10.1016/j.comcom.2004.01.010.

[22] D. Bull, "Digital picture formats and representations," Communicating Pictures. Elsevier, pp. 99–132, 2014. DOI: 10.1016/b978-0-12-405906-1.00004-0.

[23] R. Mstafa, "Reversible video steganography using quick response codes and modified ElGamal cryptosystem," Computers, Materials & Continua, vol. 72, no. 2, pp. 3349–3368, 2022, DOI: 10.32604/cmc.2022.025791.

[24] H. Mareen, J. De Praeter, G. Van Wallendael, and P. Lambert, "A novel video watermarking approach based on implicit distortions," IEEE Trans. Consumer Electron., vol. 64, no. 3, pp. 250–258, Aug. 2018, DOI: 10.1109/tce.2018.2852258.

[25] Y. Zhou, C. Wang, and X. Zhou, "An intra-drift-free robust watermarking algorithm in high efficiency video coding compressed domain," IEEE Access, vol. 7, pp. 132991–133007, 2019, DOI: 10.1109/access.2019.2940366.

[26] Netflix, "Features.md," VMAF, Netflix. (Aug. 4, 2022). [Online]. Available: https://github.com/Netflix/vmaf/blob/master/resource/doc/features.md

[27] Z. Li et al., "VMAF: The journey continues," Medium. (Oct. 26, 2018). [Online]. Available: https://netflixtechblog.com/vmaf-the-journey-continues-44b51ee9ed12

[28] W. Zhou, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. on Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004, DOI: 10.1109/tip.2003.819861.

[29] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Elsevier Science, 2013.

[30] Z. Wang, C. Chang, T. Lin, and C. Lin, "A novel distortion-free data hiding scheme for high dynamic range images," 2012 Fourth International Conference on Digital Home. IEEE, pp. 33–38, Nov. 2012. DOI: 10.1109/icdh.2012.49.

[31] C. Yu, K. Wu, and C. Wang, "A distortion-free data hiding scheme for high dynamic range images," Displays, vol. 32, no. 5, pp. 225–236, Dec. 2011, DOI: 10.1016/j.displa.2011.02.004.