Image Watermarking based on Saliency Detection and Multiple Transformations

Ahmed Khan*, KokSheik Wong* and Vishnu Monn Baskaran*
* School of Information Technology, Monash University Malaysia, Malaysia.
E-mail: {ahmed.khan1, wong.koksheik, vishnu.monn}@monash.edu

***Abstract*—An ideal image watermarking (IW) scheme aims to manage the trade-off among quality, capacity, and robustness. However, our literature survey reveals some flaws in the form of poor robustness and quality or low embedding capability. In this paper, multiple frequency domain based image watermarking scheme using salient (eye-catching) object detection is applied. Specifically, the host and the watermark images are partitioned into background and foreground regions by the proposed multi-dimension decomposition, which accumulates image patches and combining them to form the salient map. Next, the watermark image is encrypted by multiple applications of the 3D Arnold and logistic maps, then embedded into both the identified foreground and background regions of the host image by using different embedding strengths. The proposed method can embed 1 color pixel of the watermark image into 1 color pixel in the host image while maintaining high image quality. In the best case scenario, we could embed a 24-bit image as the watermark into a 24-bit image of the same dimension while maintaining an average RGB-SSIM of 0.9999. Experiments are carried out (with 10K MSRA dataset images) to verify the performance of the proposed method and to compare our proposed method against the state-of-the-art (SOTA) watermarking methods.**

I. Introduction

Nowadays, digital content can be easily generated, modified, and broadcasted. Therefore, digital content of commercial or political value is subjected to various threats, and many solutions including watermarking are put forward to protect them [1]. Watermarking is a process of inserting some data (e.g., ownership information) into the host content with the aim to survive various forms of attacks [2]. In the case of image watermarking, the attack usually includes operations designed to remove the inserted data without compromising the quality of the watermarked image.

Over the years, various innovative watermarking schemes are put forward by researchers to solve different problems [3]. In most cases, the image of interest H is transformed into another domain, and the watermark information W is embedded in the transformed domain. In particular, discrete wavelet transform (DWT) has been widely considered in the watermark community due to its multi-resolution and dimension characteristics. As a result, many DWT based watermarking methods are proposed. For example, a 3-level DWT based watermarking scheme is proposed where the watermark is embedded into the low frequency areas of the host image. On the other hand, Arnold map, DWT-SVD (Singular value decomposition) and homomorphic transforms are combined to form robust watermarking scheme [4]. Researchers also combined 4-level DWT and the discrete cosine transformation (DCT) to embed the watermark while improving robustness [5].

Recently, the combination of saliency detection [6] and DWT has received much attention in the community [7]. For example, Liu et al. first detect saliency by analyzing the Laplacian distribution of the wavelet coefficients in DWT and then embed watermark into the visually sharp edges [8]. Zhang et al. proposed to use the sharp salient features in the LL subband of contourlet transform to embed watermark [9]. Similarly, Singh et al. proposed universal structural salient based on enhanced edges to detect salient object [10]. It generates local salient map by using iterative central surround contrast and Laplacian of Guassian. Subsequently, graph model based saliency progression is performed to locate the foreground for the watermarking purpose. Pang et al. put forward a foreground noise reduction method based on the distribution of the coefficients to generate a background map [11]. The background-foreground map is then utilized for the watermarking purposes.

While more combined methods are emerging, wavelet and contourlet transformation, and salient based watermarking methods [5], [8], [9] have some drawbacks [12]. For example, the embedding process is limited to the selected (e.g. salient) areas and hence the capacity is low [8], [13]. In addition, some methods are producing watermarked image of relatively low quality even when embedding a small amount of data [9], [14]. Hence, a watermarking method with the balanced robustness, imperceptibility and high payload capacity needs to be developed.

In this paper, a multi-region watermarking scheme is proposed. Specifically, salient (i.e., foreground) and background regions are identified by using the proposed salient object detection (SOD) method [15]. The watermarking information is first encrypted by applying multiple rounds of chaotic permutations, then embedded into the background and foreground regions of the host image using different embedding strengths. Experiments are then carried out to verify the performance of the proposed method and compare it against the conventional methods.

II. Proposed Method

An image watermarking scheme based on SOD model is proposed applied [15]. Specifically, the background and foreground of the host image H are first identified using SOD method. The watermark W is then divided into two regions depending on the characteristics of H, and each region

is encrypted before it is embedded into H. The following subsections provide further details.

A. Salient Object Detection

SOD aims to detect salient objects in the host image [15]. First, multi-level features including color information and edges are extracted from the host image. These features plays vital roles in merging different set of pixels to form patches and a complete object. Suppose H is a color image and let $c \in \{1, 2, 3\}$ be the RGB channels, respectively. Let $\tau_c > 0$ be a threshold applied to the c-channel of H, denoted by H_c, to extract the edges information E_c:

$$E_c(p_1, p_2) = \begin{cases} \delta(p_1, p_2) & \text{if } \delta(p_1, p_2) > \tau_c; \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $\delta(p_1, p_2) = \sqrt{p_1^2 + p_2^2}$ is the Gaussian distance and $\theta = \tan^{-1}(p_2/p_1)$ is the phase angle between pixels p_1 and p_2. Similarly, for extracting the strong edges E_s, a convolution with Laplacian filter is performed, i.e., $E_s = H * A$ where $A = [0, 1, 0; 1, -4, 1; 0, 1, 0]$. Note that $E_s \subseteq E_c$.

Next, multi-dimensional patch decomposition (MPD) is performed to create and merge patches. Here, MPD performs dilation on the generated edges to ensure the number of patches (independent of size) agree across all channels and eventually the resulting patches are identical in all channels. The patches in each channel of E_c are processed by using a morphological operation. Specifically, $T_j = SE \oplus E_c$ is computed for each channel E_c using the structuring element, $SE = [1, 0, 1; 0, 0, 0; 1, 0, 1]$. Note that the zeros in the filter ignore the pixels which are associated with the outer boundary of the generated edges (but not part of patch) while performing dilation. This SE refines boundary of all patches in each channel, potentially merging ≥ 2 nearby patches to form a larger patch for representing a complete object. Subsequently, patches are combined to form a shadow-like mask for salient object as foreground of H.

Next, brightness allocation is performed on the processed patches to indicate the importance of each pixel in each patch. Here, the brightness allocation process applies the Gaussian distribution, where the highest value appears at the centre while the value decreases as one moves away from the centre. For example, consider a 1D patch of pixel values $[99, 101, 103, 105, 111, 121, 141]$. Next, we form 4 combinations, 141+99, 121+101, 111+103, 105, where the arrangement of pixels obeys that of the bell shape (Gaussian) distribution across the patch. Specifically, larger value (141+99) is assigned to the center of the patch, and other numbers (121+101, 111+103, and 105) are spread out repeatedly until they fill the positions up to the patch boundary. The resulting patch is $[(121 + 101), (121 + 101), (121 + 101), (121 + 101), (141 + 99), (111 + 103), (105), (105), (105)]$. Next, Gaussian smoothing is performed on each resulting patch and they collectively produce the output G.

Finally, a salient map S is produced by performing a multi-level bilateral filtering on G, i.e., $S = [1, 0, 1; 0, 0, 0; 1, 0, 1] \oplus G$. The above bilateral filtering is repeated for $d > 0$ times to obtain a refined salient map S. Subsequently, the binary salient map S^B is obtained by performing histogram equalization for improving the contrast on the patch areas and applying a threshold value τ. The resulting S^B is utilized for binary segmentation and identification of foreground and background. Specifically, the host background is obtained by performing the AND operation on H and $1 - S^B$, i.e., $H_b = H \wedge (1 - S^B)$. Similarly, the host foreground is obtained by performing $H_f = H \wedge S$. Here, the proposed SOD framework is also applied to the watermark image W to produce the foreground W_f and background W_b watermark images, respectively. Intermediate images produced by the proposed method are shown in Fig. 1.

B. Watermark embedding process

Our design aims at diffusing the watermark throughout the host image so that a high fidelity watermarked image is produced and at the same time being robust against attacks. The steps stipulated below are implemented to embed watermark W into the host image H.

1) All RGB channels of $\{W_f, W_b\}$ are encrypted (to obtain $\{W_f^e, W_b^e\}$) using 3D chaotic logistic map [16]:

$$\begin{aligned} x_{n+1} &= c_1(1 - x_n)x_n + c_2 x_n y_n^2 + c_3 z_n^3 \\ y_{n+1} &= c_1(1 - y_n)y_n + c_2 y_n z_n^2 + c_3 x_n^3, \\ z_{n+1} &= c_1(1 - z_n)z_n + c_2 z_n x_n^2 + c_3 y_n^3 \end{aligned} \quad (2)$$

   whered (x, y) is the pixel coordinate and z is the value. c_1, c_2 and c_3 are the chaotic coefficients.

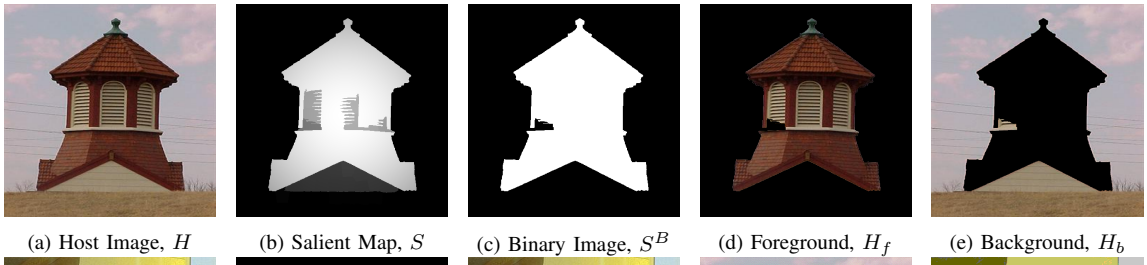
2) The segments of red channel of watermark $\{W_f^e, W_b^e\}^R$ are resized to half the size of red channel of host image $\{H_f, H_b\}^R$, referred as $\{W_f^e, W_b^e\}^a$. Next, one and two degree 3D-Arnold chaotic maps [13] are applied to $\{W_f^e, W_b^e\}^a$. Furthermore, $\{H_f, H_b\}^R$ is transformed by 1-level DWT, i.e., $[LL, HL, LH, HH]_f^R, [LL, HL, LH, HH]_b^R = DWT(\{H_f, H_b\}^R)$. The $\{LL_f, LL_b\}^R$ subbands are modified to embed the watermark regions $\{W_f^e, W_b^e\}^a$ as follows:

$$\begin{aligned} LL_f^R &\leftarrow (\alpha * W_f^{e^a} + LL_f^R) + (\beta * r) \\ LL_b^R &\leftarrow (\alpha * W_b^{e^a} + LL_b^R) + (\beta * r) \end{aligned} \quad (3)$$

   where r is PRNG sequence. The output image's unpredictability and watermark intensity are controlled by α and β, respectively. Finally, all subbands are combined, an inverse DWT is applied to form $\tilde{H}^R = \{\tilde{H}_f, \tilde{H}_b\}^R$.

3) One and two degree 3D-Arnold maps are applied to segments of the green channel $\{W_f^e, W_b^e\}^G$. The regions of green channel of the host image $\{H_f, H_b\}^G$ are transformed using a Fast Fourier Transform (FFT) and the resulting components are utilized as the host for $\{W_f^e, W_b^e\}^G$. Generally, FFT decomposes an image into real R and imaginary I parts. I holds the pixels sensitivity and correlations that maintain the image quality. However, R can be modified as follows:

$$\begin{aligned} R_f^G &\leftarrow (\alpha * W_f^{e^G} + R_f^G) + (\beta * r) \\ R_b^G &\leftarrow (\alpha * W_b^{e^G} + R_b^G) + (\beta * r) \end{aligned} \quad (4)$$



(a) Host Image, H (b) Salient Map, S (c) Binary Image, S^B (d) Foreground, H_f (e) Background, H_b

(f) Watermark, W (g) $W_f = W \wedge S^B$ (h) $W_b = W \wedge (1 - S^B)$ (i) Watermarked, H^W (j) Extracted W'

Fig. 1: Intermediate images produced by the proposed salient object detection method and watermarking scheme.

(a) Coin (b) Jump (c) Crackers

Fig. 2: Additional host images for experiments.

After combining $\{R, I\}$, an inverse FFT is applied to form $\tilde{H}^G = \{\tilde{H}_f, \tilde{H}_b\}^G$.

4) The segments of the blue channel of watermark $\{W_f^e, W_b^e\}^B$ are resized to half size of $\{H_f, H_b\}^B$ and output is referred to as $\{W_f^e, W_b^e\}^a$. Next, the one and two degree 3D-Arnold chaotic maps are applied to $\{W_f^e, W_b^e\}^a$. Furthermore, $\{H_f, H_b\}^B$ is transformed using 1-level Fast lifting wavelet transform (FLWT), i.e., $[LL, HL, LH, HH]_f^B, [LL, HL, LH, HH]_b^B = FLWT(\{H_f, H_b\}^B)$. The $\{LL_f^B, LL_b^B\}$ subbands are subsequently modified to embed $\{W_f^e, W_b^e\}^a$ as follows:

$$\begin{aligned} LL_f^B &\leftarrow (\alpha * W_f^{e^a} + LL_f^B) + (\beta * r); \\ LL_b^B &\leftarrow (\alpha * W_b^{e^a} + LL_b^B) + (\beta * r). \end{aligned} \tag{5}$$

After combining all subbands, an inverse FLWT is applied to form $\tilde{H}^B = \{\tilde{H}_f, \tilde{H}_b\}^B$.

5) The modified channels are combined to obtain the watermarked image $H_W = \tilde{H}^R + \tilde{H}^G + \tilde{H}^B$.

It is noteworthy that the watermark can be extracted by applying the reverse steps of the embedding process, and we omit the details here.

III. Experiments

The proposed method is implemented in Matlab 2020 running on a Windows 10 platform with AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx 2.10 GHz and 16GB of memory. 10K images from the MSRA dataset [10] and standard image processing images (SIPI) are considered for experiment and comparison purposes. Here, the watermark and the host images are resized to $512 \times 512 \times 3$, and the Haar wavelet is utilized. Furthermore, we set $\alpha = 0.04$ and $\beta = 0.02$, which makes the background slightly blurry while the salient object is kept completely imperceptible. Fig. 1(a) and Fig. 2 shows the host images. It is verified that the embedded watermark can be extracted.

TABLE I: Image quality and quality of watermark for the proposed and conventional methods.

Image	Our†	[17]	[14]	[9]
	Tower			
PSNR (dB)	**44.83**	43.30	40.35	40.33
SSIM	0.9990	**0.9995**	0.8904	0.9978
NC, μ	0.9401	0.8121	0.9112	**0.9792**
	Coin			
PSNR (dB)	**44.76**	43.20	40.23	40.03
SSIM	0.9999	**1.000**	0.9596	0.9720
NC, μ	**0.9259**	0.8276	0.9000	0.8251
	Jump			
PSNR (dB)	**45.13**	44.10	40.12	40.32
SSIM	0.9996	**0.9999**	0.9287	0.9141
NC, μ	0.9221	0.8425	0.8915	**0.9733**
	Crackers			
PSNR (dB)	**44.75**	44.11	40.34	40.37
SSIM	**0.9999**	0.9992	0.9757	0.9155
NC, μ	**0.9567**	0.8545	0.7240	0.9523

†The lowest PSNR among RGB channels is recorded.

Table I shows the results for four representative images from [10] as the host and common watermark image. To quantify the image quality, PSNR of each channel and SSIM [14] of marked image H^W is considered. It is observed that the watermarked image quality of the proposed method is consistently high across all channels, i.e., > 44.7dB. The SSIM results also confirms the image quality is well maintained after embedding a watermark by using the proposed method. However, the extracted watermark shown in Fig. 1(j) show some distortions in terms of contrast. It is due to the channel attacks over the encrypted watermarked image, where some grayscale values are modified. Also, this implies that the edges and structures

(a) Mean-filtering (b) Median-filtering (c) Shearing (d) Noise (e) Rotation (f) Cropping (g) JPEG-Compress

Fig. 3: Extracted watermark from the attacked watermarked image

TABLE II: Results μ after applying watermark attack.

Attack	Our	[17]	[14]	[9]
Mean-filtering	0.9375	**0.9750**	0.9567	0.8500
Median-filtering	0.9256	**0.9693**	0.9382	0.9211
Shearing	0.9182	0.9406	0.9380	**0.9427**
Noise	**0.9783**	0.9215	0.9466	0.9587
Rotation	**0.9793**	0.9604	0.9691	**0.9761**
Cropping	**0.9742**	0.9208	0.9188	0.9022
JPEG-Compression	**0.9821**	0.9389	0.9212	0.9189

are extracted well, but pixels intensities got damaged. Next, the normalized correlation (NC) [17], $\mu \in [0, 1]$ is considered to evaluate the quality of extracted watermark.

$$NC = \frac{\sum_i \sum_j (W_{(i,j)} - \mu_W)(W'_{(i,j)} - \mu_{W'})}{\sqrt{\sum_i \sum_j (W_{(i,j)} - \mu_W)^2} \sqrt{\sum_i \sum_j (W'_{(i,j)} - \mu_{W'})^2}} \tag{6}$$

where a value closer to unity implies higher similarity, and vice versa. Here, μ_w and μ_{w_e} refer to the mean intensity of original and extracted watermarks, respectively. To further evaluate the robustness of the proposed watermarking scheme, the watermarked images are attacked by independently applying mean-filter, median-filter, affine shearing with $[100; 0.510; 001]$, additive noise (0.01), rotation of 45^o, cropping with dimension $[75 : 424, 68 : 424]$ and JPEG compression. The NC results are summarized in Table II. The average NC value $\mu > 0.95$ is observed for all MSRA 10k images, which suggests the high robustness performance of the proposed watermarking scheme against the aforementioned attacks. Figure 3 shows the extracted watermark from the attacked watermarked images. It is concluded that the extracted watermarks are still visible, which agree with the numerical values recorded in Table II.

For completeness of discussion, we also compare the proposed method against the SOTA watermarking methods, namely, Jiang et al.'s method [17], Najafi et al.'s method [14] and Zhang et al.'s method [9]. First, using the same host images, [17] and [14] can only embed 64×64 bits and 512×512 bits, respectively, while the proposed method can embed $\geq 24\times$ more, i.e., on average, diffusing 1 color pixel of W into 1 color pixel of H. As reported in Table I, in terms of SSIM, the watermarked image produced by the proposed method is comparable with the existing methods although our method embeds significantly more data. In addition, despite embedding less data, the image quality attained in [14] and [9] are slightly inferior for a direct comparison. Next, the same watermark attacks are applied to the watermarked images and the average NC value μ is compared. While the results are comparable across the considered methods, the proposed method yields best results in four categories (namely, rotation, noise, JPEG compression, and cropping) out of seven. The proposed method also has a better performance in withstanding the cropping attack. A potential reason to this outcome is due to the fact that the watermark is encrypted (hence the information is spreaded) before it is embedded into the host image, while the conventional methods do not spread the watermark information. It can be observed that our performance for mean-filter and median-filter attacks is relatively lower then the SOTA methods, however it should be noted that the proposed method can embed significantly more data. As a solution to retrieve the lost image pixels, we could embed the same information repeatedly and apply majority voting or to error correction codes to improve the results.

With the aforementioned results and justifications, we conclude that the proposed method performs better than the conventional methods in terms of imperceptibility (image quality), robustness and the embedding capacity.

IV. Conclusion

In this work, a multi-region image watermarking scheme is proposed. The background and salient/foreground regions of the host image are identified using the proposed SOD model. The watermark is then encrypted to spread information throughout the host image, where different strengths are applied when embedding the encrypted watermarks into the background and foreground regions. Our method can embed a 24-bit watermark into a host color image of same dimension while maintaining an average SSIM of 0.9999 and achieving high robustness against commonly applied attacks. Experiment results suggest that our proposed method outperforms the SOTA methods, where higher imperceptibility, robustness and embedding capacity are simultaneously achieved.

As future work, our aim is to analyze other blind watermarking, such as those based on diffusion techniques, to further enhance robustness and embedding capacity. In addition, more extensive experiments will be conducted and analyze. Different wavelets are also considered for decomposition purposes as well as computationally efficient encryption algorithm.

Acknowledgment

This work is supported by Advanced Engineering Platform's Funding (account number AEP-2021-Cluster-04), Monash University Malaysia.

REFERENCES

[1] X.-b. Kang, G.-f. Lin, Y.-j. Chen, F. Zhao, E.-h. Zhang, and C.-n. Jing, "Robust and secure zero-watermarking algorithm for color images based on majority voting pattern and hyper-chaotic encryption," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 1169–1202, 2020.

[2] N. K. Jain, N. K. Rathore, and A. Mishra, "An efficient image forgery detection using biorthogonal wavelet transform and improved relevance vector machine," *Wireless Personal Communications*, vol. 101, no. 4, pp. 1983–2008, 2018.

[3] P. Puteaux, S. Ong, K. Wong, and W. Puech, "A survey of reversible data hiding in encrypted images – the first 12 years," *Journal of Visual Communication and Image Representation*, vol. 77, p. 103085, 2021.

[4] P. Khare and V. K. Srivastava, "A reliable and secure image watermarking algorithm using homomorphic transform in dwt domain," *Multidimensional Systems and Signal Processing*, 2020.

[5] J.-Y. Wu, W.-L. Huang, W.-M. Xia-Hou, W.-P. Zou, and L.-H. Gong, "Imperceptible digital watermarking scheme combining 4-level discrete wavelet transform with singular value decomposition," *Multimedia Tools and Applications*, vol. 79, pp. 22 727–22 747, 2020.

[6] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, and R. Yang, "Salient object detection in the deep learning era: An in-depth survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[7] S. Koley, "Visual attention model based dual watermarking for simultaneous image copyright protection and authentication," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 6755–6783, 2021.

[8] H. Liu, J. Liu, and M. Zhao, "Visual saliency model-based image watermarking with laplacian distribution," *Information*, vol. 9, no. 9, p. 239, 2018.

[9] Y. Zhang and Y. Sun, "An image watermarking method based on visual saliency and contourlet transform," *Optik*, vol. 186, pp. 379–389, 2019.

[10] S. K. Singh and R. Srivastava, "A robust salient object detection using edge enhanced global topographical saliency," *Multimedia Tools and Applications*, pp. 1–18, 2020.

[11] Y. Pang, Y. Wu, C. Wu, and M. Zhang, "Salient object detection via effective background prior and novel graph," *Multimedia Tools and Applications*, pp. 1–17, 2020.

[12] H.-C. Ling, R. C.-W. Phan, and S.-H. Heng, "Attacks on svd-based watermarking schemes," in *International Conference on Intelligence and Security Informatics*. Springer, 2008, pp. 83–91.

[13] A. Pourhadi and H. Mahdavi-Nasab, "A robust digital image watermarking scheme based on bat algorithm optimization and surf detector in swt domain," *Multimedia Tools and Applications*, pp. 1–25, 2020.

[14] E. Najafi and K. Loukhaoukha, "Hybrid secure and robust image watermarking scheme based on svd and sharp frequency localized contourlet transform," *Journal of information security and applications*, vol. 44, pp. 144–156, 2019.

[15] A. Khan, K. Wong, and V. Monn Baskaran, "Trade-off independent image watermarking using enhanced structured matrix decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–13, 2022 (Under review).

[16] S. Patel, K. Bharath, and R. Kumar, "Symmetric keys image encryption and decryption using 3d chaotic maps with dna encoding technique," *Multimedia Tools and Applications*, vol. 79, no. 43, pp. 31 739–31 757, 2020.

[17] F. Jiang, T. Gao *et al.*, "A robust zero-watermarking algorithm for color image based on tensor mode expansion," *Multimedia Tools and Applications*, pp. 1–16, 2020.