

Received March 5, 2018, accepted April 12, 2018, date of publication April 30, 2018, date of current version June 5, 2018. *Digital Object Identifier 10.1109/ACCESS.2018.2831599*

Performance Evaluation of a Document Image Watermarking Approach With Enhanced Tamper Localization and Recovery

LAMRI LAOUAMER¹ AND OMAR TAYAN^{19[2](https://orcid.org/0000-0002-1481-8976)}

¹ Department of Management Information Systems, College of Business and Economics, Qassim University, Buraydah 51452, Saudi Arabia ²NOOR Research Center, Department of Computer Engineering, College of Computer Science and Engineering, Taibah University, Medina 41411, Saudi Arabia Corresponding author: Omar Tayan (otayan@taibahu.edu.sa)

This work was supported by the NOOR Research Center, Taibah University, Al-Madinah Al-Munawwarah, Saudi Arabia, under Grant NRC1-126B, 2014-2015.

ABSTRACT Digital transmission of sensitive images and documents over unsecure networks, such as the Internet, has become a general practice. As a result, the digital content has become vulnerable to intentional and unintentional modifications during transmission. Prior to considering the reliability of such digital content, it is important that the authentication and the integrity of the content can be confirmed. Such issues were largely considered in the literature for natural and texture-based images, with only minimal work found to address the challenge of sensitive document images with known constraints. In this paper, we present an evaluation of a non-blind robust-watermarking approach with linear interpolation for tamper-detection, localization and recovery. Performance of the proposed approach and its resistance to random paint-based and Stirmark-based attacks for sensitive documents are investigated. Throughout this paper, a sensitive Arabic scripture was used as a case study of a sensitive document image, in which such operations were performed. The proposed model presents a superior tamper detection and recovery capability in comparison to other models in the related literature. Simulation results had demonstrated that the proposed method was robust to malicious attacks and was capable of localizing and correcting tampered regions with a high degree of accuracy. Significantly, it was noted that an average of 43 dB was obtained for the peak signal-to-noise ratio results, while generally low-BER results were achieved, with a rate of 0% in some cases. Finally, the proposed approach possessed a further advantage in its broad applicability to other sensitive digital image content.

INDEX TERMS Document-images, malicious attacks, performance-evaluation, tamper-localization, tamper-recovery, watermarking-approach.

I. INTRODUCTION

With the rapid advancements in modern communication technologies and signal processing techniques, the exchange and reproduction of digital content has become easier and faster. Such benefits are associated with challenges when ensuring digital content integrity-verification; all of which are critical requirements when transmitting sensitive and specialized content including; formal, legal, financial, and religious document images as well as medical images [1], [2]. During transmission, such sensitive content may be intentionally or accidentally manipulated, leading to undesired and even dangerous consequences [2]. In light of this urgency for protecting sensitive images in critical applications, a number of authentication and tamper detection schemes have emerged in the literature to address such demands [2], [3], [6], [9], [30].

Digital watermarking was found to provide an effective approach for integrity and authentication of multimedia content [2]–[5]. Essentially, invisible-watermarking involves embedding the author's signature into the host image such that the quality of the host image is not reduced [6]. The watermarked image is then transmitted online, which may undergo intentional or unintentional modifications before being decoded at the receiver-end to confirm its authenticity [6]. Watermark capacity and imperceptibility provide two important trade-off design metrics, which are usually balanced according to the target application requirements [6]. As a result, a number of watermark-based techniques have been proposed, which operate either in the spatialdomain or in the transform-domain. Essentially, with spatialwatermarking, the embedding process operates directly on

the image-pixels and has low-complexity, however is vulnerable to various attack-types [5]. On the contrary, transformbased watermarking is more complex, with the benefit of high robustness to various attacks [5]. Today, transform-based approaches have gained popularity in addressing contentauthentication and copyright protection [6].

Recently, a number of critical applications have raised the need to detect any tampering made and to correct such modifications so that an image identically-similar to the original copy can be recovered [3]–[6]. Consequently, several fragile and semi-fragile watermarking schemes with recovery capability have been proposed in the literature [3], [6]. However, it was pointed out in He *et al.* [7], that many of those schemes had suffered from limitations such as undetectablemodifications, localization failures and poor recovery-quality under known circumstances [3].

In this paper, we propose a robust semi-blind watermarking approach based on linear-interpolation that operates in the spatial-domain for the purpose of accurate tamper localization and high quality image recovery. This work contrasts with non-blind watermarking approaches since the proposed scheme requires the original watermark to perform the watermark extraction following an attack. Such an approach is useful in particular applications wish need to compare and minimize the error between the original watermark and the extracted watermark, which is difficult to achieve in nonblind systems. Another key motivation was to enable efficient detection of tampered zones and to facilitate the recovery process. We used the Arabic Quran scripture as a case study that provides highly sensitive textual symbols for use in our study. We evaluate our approach against a wide-range of attack scenarios and compare with other existing schemes to demonstrate the superior performance of the proposed technique. The remainder of this paper is organized as follows: Section 2 provides the related work on fragile and semifragile watermarking schemes, and Section 3 presents the proposed approach. Section 4 discusses the evaluation of the proposed approach, and finally, Section 5 concludes the paper.

II. RELATED WORK

Most of the related studies on image-watermarking for tamper-detection and image-restoration had involved fragile and semi-fragile techniques which were simulated under specific attack scenarios [3], [6]. The work in [3] presented a self-recovery fragile watermarking scheme using blockneighborhood tamper detection characterization and analyzed its performance against a number of attacks, including: general attacks, collage attacks, content-only tampering and constant-average attacks with multi-region and multiattack scenarios. Three optimization strategies were also investigated to improve the quality of localization and recovery, yielding satisfactory results for image-recovery (∼25dB PSNR) when upto 60% tampering was done.

In Lui [8], a self-embedding watermarking scheme was proposed that operates in the spatial-domain for colour image

FIGURE 1. Overview of the Watermark Generation Process.

tamper proofing and recovery. In the proposed scheme, a dual-option parity-check method with morphological operations was used to enhance tamper-proofing detection rates. Furthermore, increased quality image recovery was achieved by embedding the feature information of the host-image as a distribution over the luminance and chrominance components of the colour image in the YC_bC_r colour space. Collage, vector-quantization (VQ) and cut-and-paste attacks were used to evaluate the proposed method.

Patra and Patra [6] presented a Chinese Reminder Theorem (CRT) based fragile self-recovery watermarking scheme for tamper-detection and recovery of digital images/documents. In their proposed method, only modular arithmetic was involved in the computations of the CRT-based scheme, which produced reduced algorithmic complexity. The authors claimed that both the capacity and imperceptibility performance metrics could be improved with their approach (with upto 41dB PSNR in their samples). However, it was notable that only few attack scenarios were considered in their analysis.

The work by Rosales-Roldan *et al.* [10] describes two transform-based watermarking algorithms for tamper detection and recovery of official documents. In the first algorithm, the inverse-wavelet transform (IWT) was used for watermark embedding, while in the second algorithm, the discretecosine transform (DCT) was used for embedding. For each algorithm, an image digest of the original image was selfembedded into the frequency-components of the host image, and a multi-layer perceptron neural network (NLP) was employed in the inverse half-toning process to improve quality of the recovered image. In both those schemes, the watermark imperceptibility and robustness to JPEG compression attacks was evaluated and compared against other studies.

Qiang *et al.* [11] presented an image watermarking scheme for tamper detection and image recovery in which each block of the original image is compressed using set partitioning hierarchical trees (SPHIT), followed by scrambling

FIGURE 2. Overview of the Watermark Embedding Process.

FIGURE 3. Overview of the Extraction Watermark Process.

and embedding of the bit-streams into the LSBs of the corresponding offset block. Experimental results showed some improvement in recovery performance compared to other algorithms.

More recently, the potential of tamper detection and recovery schemes for medical image applications was explored [2], [9]. For instance, Tareef *et al.* [2] presented a tamper detection and recovery approach for medical image authentication, whereby electronic patient records (EPR) and the reshaped region of interest (RoI) were embedded in the transform-domain of the region of non-interest (RoNI). Following transmission of the images, the embedded RoI was extracted to recover the tampered image. Results had demonstrated that good recovery performance was achieved under several attack scenarios.

A number of other relevant studies have recently emerged in the literature with the aim of providing efficient image watermarking techniques for various application domains [13]–[19]. For instance, Chauhan *et al.* [14]

FIGURE 4. Samples of original images used in experimental tests.

proposed a spread-spectrum watermarking technique in the wavelet domain for protecting medical images. Initially, the DWT transform is applied with the Mexican hat transform during the preparation phase. Next, the imperceptibility of the watermark image is adjusted during embedding, in accordance with the document-to-watermark (DWT) ratio. The robustness of the proposed scheme is demonstrated with results for various types of attack and while varying the DWT parameter. In [16], Singh *et al.* present a robust and secure multiple watermarking technique (e.g. combines multiple transforms) for protecting sensitive data sent via unsecure social networks. A key benefit of this approach is that the robustness of the image watermark is enhanced using Back Propagation Neural Network (BPNN), with selective encryption applied on the important sensitive data in the watermarked image. Finally, the effects of BCH and Hamming codes on the robustness of the sensitive data was investigated, and extensive tests were analyzed against known attacks, which had shown that performance enhancements were achieved in terms of robustness, security and capacity requirements compared with the other existing techniques.

Chauhan *et al.* [17] describes an improved DWT-based medical image watermarking approach for achieving compact and secure medical data communications. Notably, this approach had addressed the problem of channel noise distortion prior to embedding. Performance analysis for various attack scenarios had demonstrated the effectiveness of the proposed approach in terms of BER and capacity, and had shown that Turbo codes had improved the performance as compared with BCH error correcting codes. On a similar theme, Zear *et al.* [18] has also used the concept of multiple watermarking based on DWT, DCT and SVD for security and authentication of medical images. Additionally, BPNN was used for achieving improved robustness against noiseeffects, and the Arnold transform was applied for the purpose of enhanced security. Results were analyzed for variable gain factors, text watermarking sizes and cover-image modalities. It was found that the proposed approach had achieved significant enhancements in robustness, imperceptibility, capacity and security as compared with the other reported methods. Finally, the work proposed by Pandey *et al.* [19] describes a new secure multiple image and text watermark method for cover-eye images used in Telophthalmology by combining

FIGURE 5. Samples of attacked watermarked images using the paint-application.

DWT and SVD transforms. Essentially, embedding is performed using four watermark components, whereby the text and image watermarks are embedded into the regions of noninterest (NROI), with SHA-512 being applied on the iris of the eye-image for enhanced security. Normalized correlation (NC) and bit-error rate (BER) results of the proposed scheme were evaluated against various signal-processing attacks and 'Checkmark' attacks, in which the robustness performance was shown for the considered attacks.

Finally, it was noted that other works found in the domain of tamper detection and recovery had suffered from medium/low recovery performance [8], or were restricted in application and dependent on the image type, as in [12], which was only applicable to grey-scale images. Moreover, many of the studies found were further limited since they had only demonstrated performance-results against few or a very specific set of attack scenarios. This paper proposes an efficient tamper-localization and recovery watermarking system with enhanced image-recovery performance compared to other related works, with results demonstrated using a large set of attack scenarios. The next section describes the proposed system model.

III. PROPOSED APPROACH

The proposed approach operates as a non-blind watermarking scheme, in which the original watermark is retrieved during the extraction process. Any modifications performed on the watermarked image following transmission do not affect the extraction of the watermark image. Hence, the proposed approach presents a robust watermarking scheme for tamper localization and self-recovery based in the spatial domain. Details of the proposed scheme are now described in the following subsections.

A. WATERMARK GENERATION PROCESS

The generation of a robust watermark image is done by partitioning the original image ($M \times N$) into 2 \times 2 blocks and computing the average of the spatial values for each block.

 (a)

FIGURE 5. Continued. (a) The extracted wa, recovered ia with tampered zones for image1_attack1.

 (b)

FIGURE 5. Continued. (b) The extracted wa, recovered ia with tampered zones for image1_attack2.

FIGURE 5. Continued. (d) The extracted wa, recovered ia with tampered zones for image2_attack1.

 (e)

FIGURE 5. Continued. (e) The extracted wa, recovered ia with tampered zones for image2_attack2.

 (f)

FIGURE 5. Continued. (f) The extracted wa, recovered ia with tampered zones for image2_attack3.

IEEE Access

 (g)

FIGURE 5. Continued. (g) The extracted wa, recovered ia with tampered zones for image3_attack1.

 (h)

FIGURE 5. Continued. (h) The extracted wa, recovered ia with tampered zones for image3_attack2.

FIGURE 5. Continued. (i) The extracted wa, recovered ia with tampered zones for image3_attack3.

The resultant watermark image consequently will be of size $M/2 \times N/2$. The pseudo code and overview of the watermark generation process are presented in algorithm 1 and in Figure 1, respectively.

The process of informed embedding is formulated as an optimization problem under robustness and distortion constraints. Selecting a 2×2 window size of the host image to generate the watermark increases the compatibility between the host signal and the watermark signal. That in turn helps to eliminate the interferences between the host signal and watermark signal, and causes a less noticeable visual distortion in the image. Notably, the correlations between pixels of adjacent blocks are very high. A window size of 2×2 is

the smallest size that can be selected to generate a watermark visually close to the original image. The pixels in any 2×2 window are more homogenous, and their average value is closer to the original one. Selecting a large window size produces a high difference between the average value and the original one, causing an increasingly noticeable visual distortion in the image.

B. WATERMARK EMBEDDING PROCESS

After obtaining the watermark image of size $M/2 \times N/2$, the embedding process is achieved by applying the linear interpolation technique on four parts of the original image to obtain the watermarked image i_w , sized $M \times N$. The linear interpolation equation is given as follows:

$$
i_w = (1 - t) w + t * i
$$
, where $0 < t < 1$

where t is the linear interpolation factor and i, w, iw represent the host image, the watermark and the watermarked image, respectively. The visibility/invisibility of the watermark can be controlled during the embedding phase by setting the value of 't' as in the following expressions:

$$
\xrightarrow{t} 0 \xrightarrow{\text{visibility}} i_w = (1 - t) w + t \times i \to w
$$

$$
\xrightarrow{t} 1 \xrightarrow{\text{involivity}} i_w = (1 - t) w + t \times i \to i
$$

Image	Type of Attack					
Name	JPEG 90	Median 9	Noise 80	PSNR 90	Rot 45	
imgl	12052 يسي المراكل التحسينات الكنديمورب السلين أَلَهُمَّنَ الرَّحِيبِ الْكَسَائِلِ بِلَامِرِ الْدِيَبِ الْ يَاتَفَعَبُدُوَيَاتَفَتَّعِمَتُ لِمَّاعَدِنَا المن المنفيذة بسرط الين أننت عَلَيْهِمْ عَيْرِ الْمُغْضُوبِ يَلْتَهُمْ midle in.	$\ldots \mathcal{A} \ldots \mathcal{B} \mathcal{F} \ldots \mathcal{A}$ الكاسي الأنفياري المأموات والتحصدان المحا والمعتمر والتعديد المصاغرة كربط أستور والمدرة ألار أتسب كمهياز استسريكه 0.0506		بسيافرانش لقسنان الكندية وبالعليم الأمن الأحب بالكليبوم الذين يَاتَنَعَبُنُوَايَاتَشَنَعِينُ لَ آهَدِنَا المنارط الشنيذ الصارط اليك لننت عَلَيْهِمْعَيْرِ الْمَغْضُوبِ يَكْتِهِمْ ١٧٠ استاليز		
img2	r BELISSIN لســـــــــم الله الزكني الزكيــــــــمّ الَّذَ لَّ ذَلِكَ الْكُنْثُ لَارْسَ فِهِ هُدَى لَلْتُقِينَ ﴾ الَّذِينَ يُؤْمِنُونَ بِالَّذِيبَ وَيُقَيِّمُونَ الصَّاؤة وسَّارزَقْتَهُمْ يَنْفِقُونَـَ> @ وَالَّذِينَ يُؤْمِنُونَ بِمَأْتُولَ إِلَيْكَ وَمَأْتُولُ مِنْ قَبْلِكَ وَبِأَلْأَخِرَةِهُمْ يُوقِنُونَ @ أَوْلَتِكَ عَلَ هُدًى مِن رَّبِّهِمْ وَأَوْلَتِهَا هُمُ الْمُفْلِحُونَ ٢	with 19 status in سألف ابن لمروز الل عل \circ arti a consideration C and realize the		TESTELLO لســـــــم الله الزكمن الزكيـــــــمّ الَّذَ لَّ الْكَانُ لَكُنْتُ لَارْبَ فِهِ هُدَى لِلْمُنْقِينَ ﴾ الَّذِينَ يُؤْمِنُونَ بِالْغِيبَ وَيُقْبِمُونَ الصّاؤة وَمَمَّارِزَقْتَهُمْ سَفِقُوتَ ﴾ وَالَّذِينَ يُؤْمِنُونَ بِمَآأَتَرَلَ إِلَيْكَ وَمَآأَتَرِلَّ مِنَ قَبْلِكَ وَبِأَلْاَخِرَةِهُمْ تُوقِنُونَ ۞ أَوَلَيْكَ عَلَىٰ هُدًى مِن زَّبِّهِمْ وَأَوْلَتِكَهُمُ ٱلْمُفْلِحُونَ ١		

FIGURE 6. Samples of attacked watermarked images using the StirMark benchmark test-suite. (a) The extracted wa, and recovered ia, with tampered zones for image1_JPEG_90 attack.

 (a)

FIGURE 6. Continued. (a) The extracted wa, and recovered ia, with tampered zones for image1_JPEG_90 attack.

FIGURE 6. Continued. (b) The extracted wa, and recovered ia, with tampered zones for image1_Median_9 attack.

FIGURE 6. Continued. (c) The extracted wa, and recovered ia, with tampered zones for image1_Noise_80 attack.

 (d)

FIGURE 6. Continued. (d) The extracted wa, and recovered ia, with tampered zones for image1_PSNR_90 attack.

 (e)

FIGURE 6. Continued. (e) The extracted wa, and recovered ia, with tampered zones for image1_ROT_45 attack.

FIGURE 6. Continued. (f) The extracted wa, and recovered ia, with tampered zones for image2_JPEG_90 attack.

 (g)

FIGURE 6. Continued. (g) The extracted wa, and recovered ia, with tampered zones for image2_Median_9 attack.

IEEE Access®

 (h)

FIGURE 6. Continued. (h) The extracted wa, and recovered ia, with tampered zones for image2_Noise_80 attack.

 (i)

FIGURE 6. Continued. (i) The extracted wa, and recovered ia, with tampered zones for image2_PSNR_90 attack.

FIGURE 6. Continued. (j) The extracted wa, and recovered ia, with tampered zones for image2_Rot_45 attack.

The interpolation factor (t) controls the visibility of watermark in the watermarked image. A less noticeable visual distortion on image is obtained when the value of 't' approaches one. Finding the optimal value of (t) value can be defined using artificial intelligent techniques. For example, genetic algorithms can be used to optimize this value. It requires defining a fitness function that considers imperceptibility and robustness ratios in order to determine the optimum (t) value.

As the interpolation factor, "t", approaches 1, the higher the obtained imperceptibility, which is based on the principle of the linear interpolation. In our tests, the optimal value of t was obtained when $t = 0.98$.

The pseudo code (e.g. a description of a code-fragment that closely resembles the English language) and overview diagram of the embedding process is shown in algorithm 2 and Figure 2, respectively. Algorithm 2 proceeds by partitioning the host image (i) of size $M \times N$ into 4 blocks each of size $M/z \times N/2$. Then, the watermark of size $M/z \times N/2$ is embedded in each block using linear interpolation technique. Finally, the four parts are concatenated to compose the watermarked image (iw) of size $M \times N$.

Practically, one advantage of informed watermarking is that it eliminates interference between the host signal and the watermark signal, and produces a less noticeable visual distortion in the image. This goal can be achieved by constructing a watermark from the original image itself to be

Algorithm 2 Code-Segment for Watermark Embedding

Input: the original image *I sized M*×*N, the watermark image w* sized *M/2*×*N/2,* linear interpolation factor *t* **Initialization:** partitioning *i*by *M/2*×*N/2*blocks result of 4 blocks

Start:

 $i_{w1} = (1-t) \times w + t \times i(1:M/2,1:N/2)$ $i_{w2} = (1-t) \times w + txi(1:M/2,N/2+1:N)$ $i_{w3} = (1-t) \times w + t \times i(M/2+1:M,1:N/2)$ $i_{w4} = (1-t) \times w + t \times i(M/2+1:M, col/2+1:N)$ $part1 = \text{horzcat}(i_{w1}, i_{w2})$ $part2 = horzcat(i_{w3}, i_{w4})$ i_w =vertcat(part1,part2) **END Output:** the watermarked image i_w sized $M \times N$

more homogenous. In the proposed model, the watermark is generated based on computing the average value of original pixels in every 2×2 window and the result is a watermark of size: $M/z \times N/2$. The pixels in any 2 × 2 window are more homogenous, and their average value is closer to the original value. Selecting a large window size produces a large difference between the average value and the original value, causing an increasingly noticeable visual distortion in the image.

FIGURE 7. Comparative Performance of various schemes for the case of multi-region attacks.

C. WATERMARK EXTRACTION PROCESS

Transmission of the watermarked image is normally done through an unreliable public network, and as such, the transmitted watermarked image is susceptible to different kinds of attacks during the transmission. Thus, the receiver may receive an attacked watermarked image (i_{wa}) that is embedded by four similar watermarks, so the extraction process will result with four attacked watermarks: w_{a1} , w_{a2} , w_{a3} , w_{a4} . The developed extraction process is semi-blind, and hence, applies the constructed watermark in the embedding process in order to extract the attacked watermark, w_a , from i_{wa} (e.g. the attacked watermarked image). In the proposed model, the host (cover) image is not required at the receiver side. On the contrary, it only requires the original watermark (w) to be applied with the attacked watermarked image (iwa) as input for extraction. However, prior to such communications, the sender and the receiver must first exchange the watermark secretly (by encryption). Thereafter, the hidden watermark can only be read by the authorized recipient during extraction.

The idea of informed watermarking, as considered in this approach, had originated from an observed connection between digital watermarking and the problem of communications that used side information at the encoder [20]. In particular, several studies had investigated how such communications (that had involved side information) could be adapted to the case of watermarking. For instance, the technique in [20] had applied this concept to eliminate the interference of a host signal on the watermark by adapting the watermark to the host signal when constructing the watermarked image/signal. This connection had enabled digital watermarking to achieve large capacity, high robustness, and good imperceptibility as reported in [21]–[23].

The extraction process involves partitioning the attacked watermarked image (i_{wa}) into four parts, and for each part, the following linear interpolation equation is applied to obtain the attacked watermark w_a , of size $M/2 \times N/2$.

$$
w_a = (1/t) w - (1-t)/t * i_{wa}, where 0 < t < 1
$$

We can select one part of the extracted attacked watermark image and consider it as a recovered attacked watermark image of size $M/2 \times N/2$. The pseudo code and overview diagram of the extraction process is shown in algorithm 3 and in Figure 3, respectively.

Algorithm 3 Code-Segment for Extraction of Attacked Watermark

Input: the attacked watermarked image i*wa*, the watermark image *w* sized *M/2*×*N/2,* linear interpolation factor *t* **Initialization:** partitioning i_{wa} by $M/2 \times N/2$ blocks result of 4-blocks

Start:

 $w_{a1} = (1/t) \times w - (1-t)/t \times iwa(1:M/2,1:N/2);$ $w_{a2} = (1/t) \times w - (1-t)/t \times iwa(1:M/2,N/2+1:N);$ $w_{a3} = (1/t) \times w - (1-t)/t \times iwa(M/2+1:M,1:N/2);$ $w_{a4} = (1/t) \times w$ - (1 t /t \times iwa(M/2+1:M,N/2+1:N); **END**

Output: the attacked watermark images w_{a1} , w_{a2} , w_{a3} , wa4each of which sized *M/2*×*N/2*

D. TAMPER DETECTION PROCESS

In order to detect the tampered zones in the original image due to the impact of different kinds of attack, the detection process is done by presenting the difference between the attacked watermarked image i*wa*, and the original watermark, *w*. This process requires matching the sizes of i*wa* and *w*. The pseudo code for this process is given in algorithm 4. In algorithm 4, the tampered zones are detected by subtracting the attacked watermarked image, iwa, from the watermark w after resizing

iwa using the same size as w. Advantageously, this enables perfect detection of the altered zones.

IV. RESULTS ANALYSIS

In this work, tests were conducted to evaluate the performance of the proposed approach. This paper considers two main kinds of attacks including; attacks generated using the Windows paint application and attacks generate during the StirMark test suite. All tests in this section were conducted on three grayscale images of size 256×256 , and were used with both types of attack scenarios. Samples of the original images are illustrated in Figure 4.The experimental results with paint-based and StirMark-based attacks are discussed in the following subsections.

A. EXPERIMENTAL RESULTS OF PAINT-BASED ATTACKS

In this experiment, the Microsoft Windows paint-software was used to simulate three types of attacks including: swapping between Arabic diacritics, swapping a word and Arabic diacritics, and finally, omitting a word. Section IV-B describes experiments conducted using a more realistic set of attack scenarios.

Figure 5 illustrates samples of attacked watermark images following paint-based attack scenarios. The results obtained for the watermark generation, extraction and recovery processes are now illustrated in Figure 5a to Figure 5i. The attack scenarios are presented and correspond to each image separately. Additionally, the tampered zones for each image are illustrated using a yellow circle.

The performance of the proposed approach in terms of robustness and similarity factors is measured from two aspects, and against different types of attacks. Firstly, the robustness and similarity results between the original image (i) and watermarked image (i_w) are presented in Table 1. Secondly, the robustness and similarity results between the original image (i) and the attacked watermarked image (i_{wa}) are presented between original watermark (w) and the attacked watermark image (w_a) in Table 2. Generally, the peak signal-to-noise ratio (PSNR) and correlation coefficient(CC) are the most common metrics that are used to measure the robustness and similarity between two images. In addition, we use the bit-error rate (BER) metric to measure the stability of the watermarked image and the watermark.

The table below illustrates the tamper-detection rate and the tamper-ratio for our processed images in terms of the

different attacks applied. The equations of tamper-detection rate and the tamper-ratio are now presented below with the definitions of our different test attack scenarios: Attack 1, Attack 2, and Attack 3.

Tamper Detectionrate =
$$
\frac{\text{number of detected pixels}}{\text{number of tampered pixels}} \times 100
$$

Tamper Ratio =
$$
\frac{\text{number of tampered pixels}}{\text{total number of image's pixels}}
$$

Attack 1: Swapping between Arabic diacritics. Attack 2: Swapping a word and Arabic diacritics. Attack 3: Omitting a word.

From Table 1, we note that the measured robustness against the three types of attacks from images 1 and 3 outperforms the robustness in image 2, where the PSNR ratio exceeds 45 dB, whilst in case of image 2 it did not exceed 42.8 dB. Furthermore, the BER ratio in image 1 was equal to 12 $\%$, whereas the BER for image 2 and image 3 reached 16%. The CC ratio in all cases was equal to 99%.

In Table 1, the variation in the similarity ratios between the host images and the watermarked images after some attacks can be explained due to the variation of tamper ratios. The tamper ratios for images 1 and 3 were 3.8% and 3.3% respectively, while the ratio for image 2 was 19%. Consequently, it is expected to observe a more noticeable visual distortion for watermarked image 2 than when compared with watermarked images 1 and 3.

From Table 2, we note that the robustness against attacks for image 1 and image 2 outperforms the robustness of image 3. The PSNR ratio for image 1 and image 2 exceeded 42 dB, and the BER ratio had not exceeded 17%. In contrast, the PSNR result for image 3 was 40 dB, while the BER ratio had reached 21%. The similarity between w and w_a was very interesting, where it was equal to 99% for all the images.

In Table 2, the variation in the robustness and similarity ratios between the original watermarks and the extracted watermarks against various attacks (for images 1-3) can be explained due the differences in the image properties. Generally, some images are more condensed with more text as

Image	Matrices	Kind of attack			
name		Swapping between Arabic diacritics (Attack1)	Swapping a word and Arabic diacritics (Attack2)	Omitting a word (Attack3)	
img1	PSNR	46.8767	46.8767	46.8767	
	BER	0.1247	0.1247	0.1247	
	CC	0.9997	0.9997	0.9997	
img2	PSNR	42.8350	42.8350	42.8350	
	BER	0.1633	0.1633	0.1633	
	CC	0.9996	0.9996	0.9996	
img3	PSNR	45.2801	45.2801	45.2801	
	BER	0.1639	0.1639	0.1639	
	CC	0.9997	0.9997	0.9997	

TABLE 2. The performance of proposed model in terms of PSNR, BER, and CC between the watermark (w) and attacked watermark image (wa).

compared to others. Furthermore, we can find more texture, background uniformity, color degrees, etc, in particular images as compared to others. We note that we have some images with no smooth background, whilst other images have a uniform background color, whilst yet other images are more condensed (e.g. as in image3). Such variations in the image properties explain why there is a variation in robustness and similarity ratios when comparing between images 1-3.

B. EXPERIMENTAL RESULTS OF STIRMARK-BASED **ATTACKS**

This section applies the set of Stirmark benchmarks representing image attack scenarios. Notably, the Stirmarkbenchmark suite has been used extensively in the literature for simulating attacks for image-watermarking schemes and contains a rich set of benchmarks that include the most known attack scenarios. Therefore, we found that the Stirmark benchmark suite would be particularly ideal to apply on our scheme in order to compare with many other related studies that had also considered attack-scenarios based on the Stirmark-benchmark.

In the following experiments, all results relating to the StirMark-based attacks are presented. In particular, the Stirmark-based attacks include: the JPEG_90, Median_9, Noise_80, PSNR_90, and Rot_45 attacks, with the tests being applied on image 1 and image 2. Samples of attacked watermarked images using the StirMark software are illustrated in Figure 6. Simulation results of the watermark generation, extraction and recovery processes are now illustrated in Figures 6a through to 6j. Tampered zones under the various attack conditions are also shown for each case.

The experimental results that follow demonstrate the performance of proposed approach in terms of PSNR, BER, and CC between original image (i) and watermarked image (i_w) , in addition to those results for the watermark (w)

TABLE 3. The performance of proposed model in terms of PSNR, BER, and CC between original image (i) and watermarked image (i_w).

TABLE 4. The performance of proposed model in terms of PSNR, BER, and CC between watermark (w) and attacked watermark image (w_a).

TABLE 5. Results of the Tamper-Ratio and Tamper-Detection rates for the Stirmark attack Scenarios

and attacked watermark image (w_a) . Those results are now presented in Table 3 and Table 4, respectively.

From Table 3, we find that the PSNR and BER ratios between the original image (i) and watermarked image (i_w) for image 1 outperforms the corresponding results for image 2 by a factor of 4 dB for the PSNR, and

4% for the BER ratio. The CC ratio in both images had reached 99%.

From Table 4, it is observed that the results in terms of the PSNR and BER ratios for both images are convergent where the difference is below 2%. However, the CC ratio in both cases was similar at 99%. By observing Table 4,

it can also be concluded that the results with JPEG_90, Median_9 and PSNR_90 increases beyond the results for Noise_80 and Rot_45 in terms of the PSNR and BER metrics. Table 5 presents results of the tamper-detection rate and tamper-ratio for the Stirmark attack scenarios using the same three images from Figure 4. A notable advantage in our approach was observed with the 100% detection-rate being achieved under all attack scenarios.

V. PERFORMANCE EVALUATION

A. PERFORMANCE EVALUATION OF PROPOSED APPROACH FOR NATURAL/HIGH-TEXTURE IMAGES

Table 6 illustrates the experimental results of our proposed approach for three natural images that include: the Peppers, Sailboat and F16 images. Table 6 displays the PSNE, BER, and CC results between the original watermark and attacked watermark. Table 7 presents the tamper-ratios and the corresponding tamper-detection rates for those sample images used.

B. COMPARATIVE EVALUATION AGAINST OTHER TAMPER-DETECTION ALGORITHMS

This section first provides a comparison of our approach with five other algorithms previously used for natural and high-texture images, as found in Patra and Patra [6],

TABLE 7. Results of the Tamper ratio and Tamper-Detection rates using our Approach for Samples of Natural and High-Texture Images.

TABLE 8. Comparative Results for our approach against other schemes for natural images.

Host image	PSNR(dB) of watermarked host with original host					
	Lin et al. $[27]$	Patra et al. [6]	Lee et al. [26]	Ton g et al. [25]	Ansari et al. [24]	Proposed Scheme
Sailboat	43. 85	4 3 4 5	4 0 7 0	4 0 5 8	4 3 5 6	4 8 4
Peppers	43. 83	43.9 4	40.73	40. 32	44.04	49.72
Baboon	44. 21	43.1 5	40.73	40. 71	44.35	53.63

Ansari *et al.* [24], Tong *et al.* [25], Lee *et al.* [26], and Lin *et al.* [27]. The comparative study was based on the PSNR metric between the watermarked image and the original image, where the original image in each case was of size 256×256 pixels. From Table 8, the superior performance of our approach compared to the other approaches can be clearly observed when applied for natural and high**TABLE 9.** Comparative results with Ansari et al. [24] following common tampering attacks.

TABLE 10. Comparative results with Singh et al. [28] following common tampering attacks.

texture images. In particular, our PSNR results in some cases had exceeded 53dB, whilst the next best algorithm (Ansari *et al.* [24]), which had only reached 44dB for the same image.

In Table 9, we compare our approach with Ansari *et al.* [24] in terms of the PSNR between the watermarked image and the recovered image against common tampering attacks. In all the following test cases, image sizes of 256×256 were employed. Notably, the approach proposed by Ansari *et al.* was based on the frequency domain using SVD. In contrast, our proposed scheme operates in the spatial-domain, which

reduces the computational complexity and processing-time required as compared with Ansari *et al.* [24]. Despite the fact that improved extraction results are possible using frequency domain schemes, additional benefits were obtained in our approach since our goal of tamper-localization and recovery was achieved in the spatial domain, while completely removing the need for any complex image transformations.

In Table 10, we compare our approach with Singh *et al.* [28] in terms of the PSNR between the watermarked image and the recovered image against common tampering attacks as illustrated in the Table 10. In all

Multi Regions Tampering			
Methods	Tamper ratio (Rt)	PSNR(dB)	
Lin[27]	35.77	15.9	
Lee[26]	28.09	23.2	
He[7]	44.24	12.42	
He et al. $[3]$	28.09	27.65	
Proposed Scheme	27.12	32.05	

TABLE 11. Comparative results with other schemes against known attack scenarios.

the following test cases, image sizes of 512×512 were employed.

Finally, we compare our approach with He *et al.* [3], He [7], and Lee and Lin [26] in terms of the tamper- ratio and the PSNR between the original watermark and the attacked watermark images for the case of multi region attacks. The attacks include: 1) some letters from ''ITP Southwest Jiaotong University,'' 2) the transverse columns tampering, 3) the longitudinal columns tampering, 4) the diagonal columns tampering, 5) many small square regions, and, 6) a triangle region. Tamper detection performance results and the recovered images for the aforementioned attacks are presented in Table 11. In all the following test cases; image sizes of 512×512 were employed.

VI. DISCUSSION AND CONCLUSIONS

This paper has proposed a robust non-blind watermarking approach that operates in the spatial-domain and was shown to possess remarkable tamper-localization accuracy and image-recovery quality. A notable contribution/novelty in this paper was derived from the use of linear-interpolation with our robust watermarking scheme in order to enhance performance when extracting the watermark image, which enabled our approach to provide full detection and recovery capability. The proposed approach was first applied on sensitive document-images, characteristic of fine-granular textsymbols and minimal use of colours/textures, which imposes constraints for achieving an imperceptible embedding. Performance of the proposed approach and its resistance to random paint-based and Stirmark-based

attacks for sensitive document images was then demonstrated. Experimental results had shown that the proposed method was robust to a large set of malicious attacks, including random paint-based and Stirmark-based attacks, while maintaining high-quality and an exceptionally high degree of

accuracy for localizing and correcting tampered regions (with recovery rates reaching 100%).

The proposed approach was then evaluated when applied with natural and high-texture based images that are commonly used in the related literature. In comparison to other methods from the literature, the results had demonstrated the superiority of our scheme over other tamper-detection and recovery algorithms for those images. Moreover, our proposed approach possessed a further advantage in its broad applicability to many types of sensitive digital image content as well as for natural and high-texture images, with our PSNR results exceeding

53dB in some cases of natural images. Future enhancements to this work involve extending and testing our approach for use with other sensitive images, including; medicalimages and colour video applications.

REFERENCES

- [1] O. Tayan, M. N. Kabir, and Y. M. Alginahi, "A hybrid digital-signature and zero-watermarking approach for authentication and protection of sensitive electronic documents,'' *Sci. World J.*, vol. 2014, Aug. 2014, Art. no. 514652.
- [2] A. Tareef, A. Al-Ani, H. Nguyen, and Y. Y. Chung, ''A novel tamper detection-recovery and watermarking system for medical image authentication and EPR hiding,'' in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Chicago, IL, USA, Aug. 2014, pp. 5554–5557.
- [3] H. He, F. Chen, H.-M. Tai, T. Kalker, and J. Zhang, "Performance analysis of a block-neighborhood-based self-recovery fragile watermarking scheme,'' *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 1, pp. 185–196, Feb. 2012.
- [4] L. Laouamer and O. Tayan, "An enhanced SVD technique for authentication and protection of text-images using a case study on digital Quran content with sensitivity constraints,'' *Life Sci. J.*, vol. 10, no. 2, pp. 2591–2597, 2013.
- [5] L. Laouamer and O. Tayan, "A semi-blind robust DCT watermarking approach for sensitive text images,'' *Arabian J. Sci. Eng.*, vol. 40, pp. 1097–1109, Apr. 2015.
- [6] B. Patra and J. C. Patra, "CRT-based fragile self-recovery watermarking scheme for image authentication and recovery,'' in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Taipei, Taiwan, Nov. 2012, pp. 430–435.
- [7] H.-J. He, J. S. Zhang, and F. Chen, ''Adjacent-block based statistical detection method for self-embedding watermarking techniques,'' *Signal Process.*, vol. 89, pp. 1557–1566, Aug. 2009.
- [8] K.-C. Liu, "Colour image watermarking for tamper proofing and patternbased recovery,'' *IET Image Process.*, vol. 6, no. 5, pp. 445–454, Jul. 2012.
- [9] R. Eswaraiah and E. S. Reddy, ''Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest,'' *IET Image Process.*, vol. 9, no. 8, pp. 615–625, 2015.
- [10] L. Rosales-Roldan, M. Cedillo-Hernández, M. Nakano-Miyatake, and H. Pérez-Meana, ''Watermarking-based tamper detection and recovery algorithms for official documents,'' in *Proc. 8th Int. Conf. Elect. Eng. Comput. Sci. Autom. Control (CCE)*, Merida City, Mexico, Oct. 2011, pp. 1–6.
- [11] S. Qiang, W. Jiawen, and Z. Hongbin, "Tamper detection and self-recovery of image based on self-embedding,'' in *Proc. Asia-Pacific Conf. Inf. Process.*, 2009, pp. 76–79.
- [12] C. K. R and N. Shivananda, "A new fragile watermarking approach for tamper detection and recovery of document images,'' in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, New Delhi, India, 2014, pp. 1494–1498.
- [13] A. K. Singh, B. Kumar, G. Singh, and A. Mohan, *Medical Image Watermarking: Techniques and Applications* (Multimedia Systems and Applications). New York, NY, USA: Springer, 2017.
- [14] D. S. Chauhan, A. K. Singh, A. Adarsh, "Combining Mexican hat wavelet and spread spectrum for adaptive watermarking and its statistical detection using medical images,'' *Multimedia Tools Appl.*, pp. 1–15, Nov. 2017.
- [15] C. Kumar, A. K. Singh, and P. Kumar, "A recent survey on image watermarking techniques and its application in e-governance,'' *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 1–26, 2018.
- [16] A. K. Singh, B. Kumar, S. K. Singh, S. P. Ghrera, and A. Mohan, ''Multiple Watermarking technique for securing Online social network contents using back propagation neural network,'' *Future Generat. Comput. Syst.*, vol. 2016, Nov. 2016. [Online]. Available: https://www. sciencedirect.com/science/article/pii/S0167739X16306628?via%3Dihub
- [17] D. S. Chauhan, A. K. Singh, J. P. Saini, and B. Kumar, ''Quantization based multiple medical information watermarking for secure e-health,'' *Multimedia Tools Appl.*, pp. 1–13, Jun. 2017.
- [18] A. Zear, A. K. Singh, and P. Kumar, "A proposed secure multiple watermarking technique based on DWT, DCT and SVD for application in medicine,'' *Multimedia Tools Appl.*, vol. 77, pp. 1–20 Feb. 2018.
- [19] R. Pandey, A. K. Singh, and B. Kumar, ''Iris based secure NROI multiple eye image watermarking for teleophthalmology,'' *Multimedia Tools Appl.*, vol. 75, no. 22, pp. 14381–14397, Nov. 2016.
- [20] J. Chou, S. S. Pradhan, and K. Ramchandran, "On the duality between distributed source coding and data hiding,'' in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, vol. 2. Oct. 1999, pp. 1503–1507.
- [21] C. Wang. "An enhanced informed watermarking scheme using the posterior hidden Markov model,'' *Sci. World J.*, vol. 2014, Jan. 2014, Art. no. 345892.
- [22] A. Abrardo and M. Barni, "Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding,'' *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 824–833, Feb. 2005.
- [23] C. Wang, J. Ni, and J. Huang, "An informed watermarking scheme using hidden markov model in the wavelet domain,'' *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 853–867, Jun. 2012.
- [24] I. A. Ansari, M. Pant, and C. W. Ahn. "SVD based fragile watermarking scheme for tamper localization and self-recovery,'' *Int. J. Mach. Learn. Cyber*, vol. 7, no. 6, pp. 1225–1239, 2015, doi: 10.1007/s13042-015- 0455-1.
- [25] X. Tong, Y. Liu, M. Zhang, and Y. Chen, ''A novel chaos-based fragile watermarking for image tampering detection and self-recovery,'' *Image Comm.*, vol. 28, no. 3, pp. 301–308, 2013.
- [26] T-Y. Lee and S. D. Lin, ''Dual watermark for image tamper detection and recovery,'' *Pattern Recognit.*, vol. 41, no. 11, pp. 3497–3506, 2008.
- [27] P. L. Lin, C.-K. Hsieh, and P.-W. Huang, "A hierarchical digital watermarking method for image tamper detection and recovery,'' *Pattern Recognit.*, vol. 38, no. 12, pp. 2519–2529, 2005.
- [28] P. Singh and S. Agarwal, "An efficient fragile watermarking scheme with multilevel tamper detection and recovery based on dynamic domain selection,'' *Multimedia Tools Appl.*, vol. 75, no. 14, pp. 8165–8194, 2016, doi: 10.1007/s11042-015-2736-9.
- [29] J.-D. Chang, B.-H. Chen, and C.-S. Tsai, "LBP-based fragile watermarking scheme for image tamper detection and recovery,'' in *Proc. IEEE Int. Symp. Next-Generat. Electron. (ISNE)*, Kaohsiung, Taiwan, Feb. 2013, pp. 173–176.
- [30] M. Iwata, T. Hori, A. Shiozaki, and A. Ogihara, ''Digital watermarking method for tamper detection and recovery of JPEG images,'' in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Taichung, Taiwan, Oct. 2010, pp. 309–314.

LAMRI LAOUAMER received the B.Sc. degree in computer science from the University of Setif, Algeria, in 1999, the M.Sc. degree in the field of computer science and applied mathematics from the University of Quebec, Trois Rivieres, Canada, in 2006, and the Ph.D. degree in computer science in the field of information security from the University of Bretagne Occidentale, Brest, France, in 2012. He is currently an Assistant Professor with the Department of Management Information

Systems, Saudi Arabia. He is also an Associate Researcher with the Laboratoire des Sciences et Techniques de l'Information, de la Communication et dela Connaissance, University of Bretagne Occidentale.

His research interests include multimedia watermarking, cryptology, information security. He is an Associate Editor of the *Journal of Telecommunication Systems* (Springer) and the *Journal of Innovation in Digital Ecosystems* (Elsevier).

OMAR TAYAN received the B.Eng. and Ph.D. degrees in computer networks from the University of Strathclyde, Glasgow, U.K. He was a Consultant with the Strategic and Advanced Research and Technology Innovation Unit, Taibah University, Saudi Arabia. He is currently an Associate Professor with the NOOR Research Center, College of Computer Science and Engineering, Taibah University. He is one of the Founding Members of the NOOR Research Center. He currently has over

50 journals, conference papers, technical reports, invited talks to his credit, and a book publication in computer-networks. He has successfully completed about 10 research and development projects as a Principle Investigator and a co-investigator in projects funded by King AbdulAziz City for Science and Technology, Ministry of Higher Education, and the Deanship of Research at Taibah University. His research interests include information security, e-learning and multimedia technologies, Quranic computing, image processing, modeling and simulation, computer networks and networks-on-chip, wireless sensor networks for intelligent transportation systems including Hajj transportation systems, and crowd management.