FREQUENCY AND SPATIAL DOMAIN-BASED IMAGE AUTHENTICATION AND TAMPERING DETECTION

Nidaa Hasan Abbas[1,2*], Sharifah Mumtazah Syed Ahmad[1,3], Sajida Parveen[4] , Wan Azizun Wan[1], Abd. Rahman Bin Ramli[1]

[1] Department of Computer and Communication System Engineering, Universiti Putra Malaysia Serdang,Selangor,Malaysia
[2] Electrical Department, Almustansirya University, Baghdad, Iraq
[3] Research Center of Excellence for Wireless and Photonic Network, Serdang 43400, Malaysia
[4] Faculty of Electrical, Electronics and Computer Systems Engineering, Quaid-e-Awam University of Engineering Science and Technology, Nawabshah 67450, Pakistan
*nidaahasan71@gmail.com

ABSTRACT: *In this research, an efficient self-embedding algorithm for image authentication and tamper detection and localisation in the frequency and time domain is proposed. For more detection accuracy, two fragile watermarks are generated from the original image and are embedded into the same image. The first watermark is generated in the time domain using block wise method and embedded in the Least Significant Bits (LSB) of the same derived block while the other watermark is generated and embedded into the frequency domain utilising Bi- Empirical Mode Decomposition (BEMD). To ensure that the security requirements of the algorithm are met, a block encryption technique is employed whereby a password chosen by the user is encrypted and used as a key. The algorithm has been subject to several tampering attacks and has been proven very efficient in terms of processing high-quality watermarked images with high security and the capability to detect small tampered areas at pixel* level.

Keywords: Image watermarking, Image tampering, Image authentication, Fragile watermark, BEMD.

1. INTRODUCTION

Image authentication has become a mandatory requirement because of the rapid development of technology and digital editing tools, which give a novice user the ability to manipulate digital content. In some cases, verifying the integrity of images may change the life of a person when the court statement depends on a tampered image [1]. To complete the authentication procedure, localising the tampered area is required, along with verifying the originality of the image in question. Separating the authorised parts of an image from its unauthorised parts is a crucial issue in certain cases, such as medical imagery or military maps [2].

Traditional strategies, such as digital signatures, that have been used for digital content authentication have their limitations [3]. As a result, digital watermarking techniques that provide effective detection mechanisms are critically needed. Digital watermarking algorithms are classified into three categories: robust, semi-fragile and fragile watermarking, and they are employed depending on the application to be used. A robust watermark is used for copyright protection. For this purpose, the embedded watermark must be robust and resistive towards deliberate attacks [4]. Semi-fragile watermarks are designed to allow an acceptable level of alteration, such as slight contrast adjustment or low-level lossy compression in images [5]. Meanwhile, fragile watermarks, which are used for tampering detection, do not require the same robustness level as those used for copyright protection mainly because it needs the capability to detect even the slightest modification to the media [2]. As a result, this type of watermarking is suitable for authentication and tamper localisation applications. The image authentication procedure includes the extraction of the verification code from certain features of the original image and the embedding of such code in the same image. The embedded code, which represents the watermark, is then extracted from the watermarked image during the decoding procedure to be compared with the original code. The image is considered authentic if the two compared codes are alike; otherwise, the image is considered as not genuine.

In this study, a fragile watermark algorithm for image authentication and verification that meets the fundamental requirements for an efficient algorithm is presented. The important features of an efficient algorithm are as follows: perceptual quality, localisation of minor tampering, and high security. The fidelity or quality of the watermarked image is measured using the peak signal-to-noise ratio (*PSNR*). The proposed algorithm achieves a *PSNR* of about 50 dB and thus outperformed other similar algorithms. Tampered area localisation is the second important factor. Localisation tampering can be categorised into three levels: localisation at the pixel level, at the block level, and at the whole image level. The most efficient algorithm can detect minor tampered pixels [6]. In this regard, the proposed algorithm is built at the pixel level detection. Security is another essential property that should be considered. [7]. In the proposed algorithm, the security issue is carefully considered through the encryption of a password chosen by the user and the use of such password as a secret key to prevent an unauthorised user from accessing the main algorithm.

The rest of the paper is organized as follows: Section 2 presents the methods of image authentication. Section 3 describes the theoretical background of Bi Empirical Mode Decomposition (BEMD). The proposed algorithm is detailed in Section 4. Evaluation of the performance of the algorithm is explained in section 5. Section 6 reports the most promising research trends along with the conclusions of this paper.

2. Digital image authentication methods

Fragile watermarking algorithms can operate directly in the spatial or the transform domain [8]. Spatial domain fragile watermarking algorithms can either be designed and implemented at the block level [9, 10] or at the pixel level [7, 11, 12]. However, block-wise algorithms lack the localisation detection functionality [15, 13]. Consequently, the concept of pixel-wise fragile watermarking can be considered as an alternative solution

[12]. In the pixel-wise scheme, the fragile watermark is extracted from the grey value of the pixels and embedded in the same image. In this case, the alteration of pixel value will reflect on the watermark and can be easily observed at the receiver side [14].

While the pixel-wise method is simple, fast, and suited for real-time applications, this technique is unsecured because the pixel can handle only a limited number of discrete values and the system can be forged easily [2]. Accordingly, fragile watermarking based on frequency domain is used to increase payload capacity [15]. Few works adopted frequency domain, one of which is the work proposed and implemented in discrete cosine transform domain or DCT by [16] and Lin *et al*. [17]. Although the authors in [17] increased the capacity of the embedded watermark by dividing the original image into 16×16 pixel blocks, the algorithms based on DCT still have limited capacity [18]. In this regard, Slant Transform (ST) is adopted by Zhao et al. [19] in designing two authentication algorithms—active and passive which could detect 98% of tampered area.

The problem of the low capacity payload of the derived authentication code was addressed in [15]. They used the two-dimensional Hartley Transformation [20] in decomposing the original image, as a result of which they succeeded in extending the derived watermark to 128 bits. In general, the authentication watermarking algorithms that are based on frequency transforms are semi-fragile since all the transform domains are almost designed to be robust against lossy compression; however, authentication code verification watermarking methods are very weak in a functional sense [21]. Thus, in this paper, the Bi Empirical Mode Decomposition algorithm (BEMD) transform is selected amongst other transforms because it decomposes the image into a sequence of high through to low frequency subbands. Hence, the most sensitive components of the image are used to generate the authentication code bits due to its ability to detect minor tampering area [22]. In addition, BEMD is an adaptive transform and suitable for nonlinear, nonstationary data analysis [23]. It is a fully data-driven method [24], and more accurate because it does not depend on pre-determined filter like other transforms [25].

3. Bi Empirical Mode Decomposition

The BEMD [26] is derived from the Empirical Mode Decomposition (EMD) [27]. It decomposes the signal into oscillatory components called intrinsic mode functions (*IMFs*) and the coarsest component termed as mean trend or residue (r). By this transform, a signal is projected on to basic functions which are directly derived from the signal itself, unlike other transforms that depend on predefined basis functions, such as Gabor analysis [25] and Wavelet analysis [28]. The coarsest component of BEMD is highly robust for attacks such as noise and JPEG compression, while the *IMFs* contain the less fragile frequency components. Accordingly, in the proposed algorithm, the properties of *IMFs* are exploited to derive the most sensitive authentication code bits that can detect any tampering on a pixel level with high detection rate.

3.1. Empirical Mode Decomposition (EMD) Algorithm

EMD was first introduced by Huang et al. [29] for the non-stationary function decomposition. It can decompose any complicated signal adaptively into finite and a small number of intrinsic mode functions (*IMFs*). *IMFs* are extracted from the signal using the sifting algorithm.

3.1.1 Sifting procedure: The sifting procedure extracts locally for each mode the highest frequency oscillations out of the original signal. For a sampled signal $s(k)$, there are two constraints that should be satisfied during the procedure:

- Each *IMF* has the same number of zero crossings and extrema;
- Each *IMF* is symmetric with respect to the local mean. Furthermore, it assumes that s (k) has at least two extrema.

For a signal $s \in l^2(z)$, the procedure of sifting algorithm is as follows:

1. Initialization: put the values of the residue r_0 and index number j of *IMF* to s and 1 respectively.
2. The j^{th} *IMF* is extracted.
3. (a) Initialization: $h_o = r_{j-1}$, $i = 1$

   (b) Local minima/maxima of h_{i-1} are extracted,

   (c) The upper and lower envelope functions x_{i-1} and y_{i-1} are computed by interpolating, local minima and local maxima of h_{i-1} ;

   (d) The mean envelope is computed as follows : $m_{i-1} = (x_{i-1} + y_{i-1})/2$

   (e) Updating: $h_i = h_{i-1} - m_{i-1}$ and $i = i + 1$.

   (f) To stop the sifting process, a certain criterion *SD* is computed from two successive sifting results as:

   $$SD = \frac{1}{N} \sum_{k=0}^{k} \left[\frac{(h(j-1)^{(k)} - hij^{(k)})^2}{h^2(j-1)^{(k)}} \right] \qquad (1)$$

   (g) Decision: steps (b) to (f) are repeated until $SD_i \leq \xi$, and set the value of $d_i = h_i$ (j^{th} *IMF).*
4. Updating residual value $r_{ij} = r_{j-1} - m_j$
5. Steps (1-3) are repeated with $j := j + 1$ and stopped when the number of r_j is less than 2.

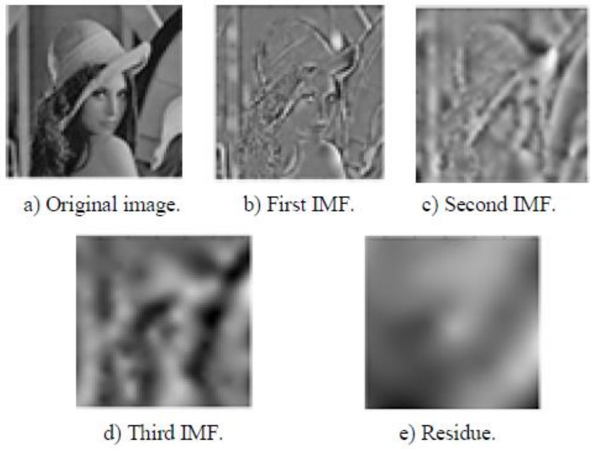
   The signal is reconstructed by superposition of all the *IMFs*:

   $$m(n) = \sum_{k=1}^{k} d_k(n) + r(n), \; K \in N^* \qquad (2)$$

where $d_k(n)$ is the *IMFs* and $r(n)$ is the residue.

For two dimensional signal such as image, every *IMF* is conveniently separated as containing information of a specific scale. For image authentication algorithms, the advantages of using *IMFs* are they can cover a wide frequency range and hence the watermark capacity payload will increase.

To illustrate the decomposition process in this research, 3*IMFs*, considered by repeating the process were satisfactory to obtain a suitable residue. Figure 1 depicts the decomposition process for Lena image as reported in [30].



a) Original image. b) First IMF. c) Second IMF.

d) Third IMF. e) Residue.

Fig. 1. Example of applying BEMD on Lena image.

It has been backed with credible research that all decomposition techniques including Fourier and wavelet are inferior to EMD because of the following characteristics [31].

i. It is a better method because of its efficiency and adaptability which makes it better suited for both linear and non-linear applications.

ii. Unlike Gabor [32] and wavelet analysis [23] that relies on predefined basis functions, this transform does not use a pre-determine basis function. Rather, a signal is directed to the basis function which can be derived from the data itself.

iii. An *IMF* is an AM–FM segment that has been utilised effectively as part of an assortment of utilisations including nonstationary investigation, image enhancement, edge detection; 3D shapes recovery from texture, computational stereopsis, segmentation of texture, and classification [29].

iv. With respect to design efficient authentication code that can detect minor manipulations to any image pixel, the properties of *IMFs* can be exploited in deriving a very sensitive watermark bits as early mentioned [23] .

v. Adopting BEMD in digital image watermarking gives the capacity to build the payload limit without influencing the imperceptibility of the watermarked image [29]. As specified before, watermark limit is a noteworthy issue on account of image verification watermarking as most existing algorithms have restricted limit.

4. Proposed watermarking algorithm

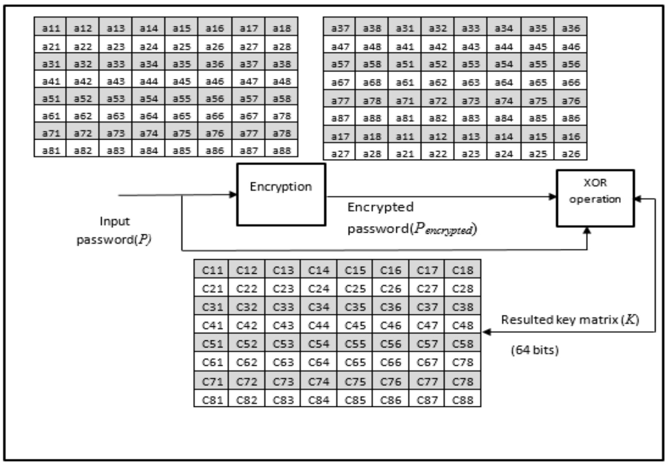
All the drawbacks highlighted in the above discussion of the existing image authentication algorithms which have adopted block-wise, pixel-wise, or frequency domains are taken into consideration in the design of the proposed algorithm. The watermarking algorithm consists of two stages: generating and embedding the authentication code bits and detecting the authentication code bits. The details of the fragile watermarks generation and embedding procedure are explained in the following subsections:

4.1. Watermark generation procedure

The watermark generation consists of three stages: key encryption, spatial domain watermark generating and frequency domain watermark generation.

4.1.1 Key generation procedure: In order to increase the security of the proposed system, a secret key is used before generating the watermark. It is devised by the user in order to prevent attackers from breaking the authentication algorithm. The secret key is produced and encrypted using the block encryption method which is simple but creates sufficient confusion for the attacker. The key is used as a password and is embedded in chosen blocks of the original image. The procedure for encrypting the secret key is shown in Figure 2 and as follows,

1. The password (P) of length 64-bit binary is chosen by the user.
2. The rows are shifted by two positions.
3. The columns are shifted by two positions.
4. The original password matrix (P) is XOR-ed with the newly converted one ($P_{encrypted}$) to produce the resulted encrypted key matrix (K).
5. The first four blocks of 4×4 size are selected from the original image for embedding the secret key.



a11	a12	a13	a14	a15	a16	a17	a18
a21	a22	a23	a24	a25	a26	a27	a28
a31	a32	a33	a34	a35	a36	a37	a38
a41	a42	a43	a44	a45	a46	a47	a48
a51	a52	a53	a54	a55	a56	a57	a58
a61	a62	a63	a64	a65	a66	a67	a78
a71	a72	a73	a74	a75	a76	a77	a78
a81	a82	a83	a84	a85	a86	a87	a88

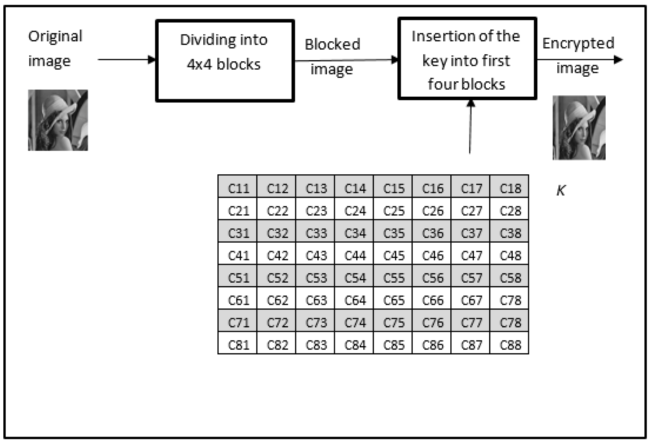
a37	a38	a31	a32	a33	a34	a35	a36
a47	a48	a41	a42	a43	a44	a45	a46
a57	a58	a51	a52	a53	a54	a55	a56
a67	a68	a61	a62	a63	a64	a65	a66
a77	a78	a71	a72	a73	a74	a75	a76
a87	a88	a81	a82	a83	a84	a85	a86
a17	a18	a11	a12	a13	a14	a15	a16
a27	a28	a21	a22	a23	a24	a25	a26

Encryption
XOR operation
Input password(P)
Encrypted password($P_{encrypted}$)
Resulted key matrix (K)
(64 bits)

C11	C12	C13	C14	C15	C16	C17	C18
C21	C22	C23	C24	C25	C26	C27	C28
C31	C32	C33	C34	C35	C36	C37	C38
C41	C42	C43	C44	C45	C46	C47	C48
C51	C52	C53	C54	C55	C56	C57	C58
C61	C62	C63	C64	C65	C66	C67	C78
C71	C72	C73	C74	C75	C76	C77	C78
C81	C82	C83	C84	C85	C86	C87	C88

Fig. 2. Secret key generation

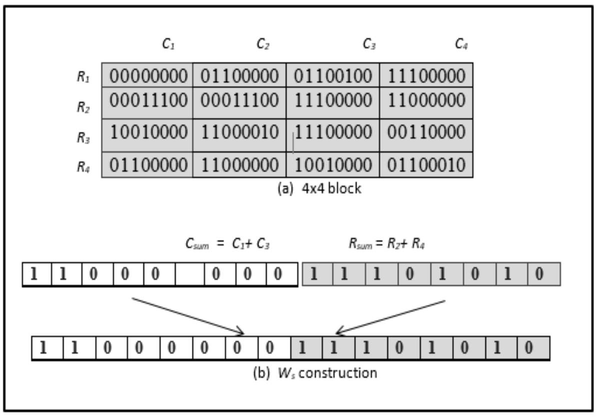
6. The resulted encrypted matrix (K) is changed into a 4×4 array to be embedded in the LSB of the selected blocks as shown in Figure 3.



C11	C12	C13	C14	C15	C16	C17	C18
C21	C22	C23	C24	C25	C26	C27	C28
C31	C32	C33	C34	C35	C36	C37	C38
C41	C42	C43	C44	C45	C46	C47	C48
C51	C52	C53	C54	C55	C56	C57	C58
C61	C62	C63	C64	C65	C66	C67	C78
C71	C72	C73	C74	C75	C76	C77	C78
C81	C82	C83	C84	C85	C86	C87	C88

Fig 1. The key embedding process

4.1.2 Spatial Domain Fragile Watermark Generation Procedure: After inserting the secret key, the fragile watermark (*Ws*) is generated in the time domain by dividing the image into blocks of 4×4 pixels. In order to create standard template in both the embedding and tamper detection procedure, the LSB of each pixel in the two processes is initially set to zero [33]. Once the LSB of each pixel is set to zero, the process of constructions the fragile watermark (*Ws*) from each block begun. *Ws* consists of 16 bits length, the first 8 bits of them (*Csum*) are constructed by adding the values of the pixel in the first and third columns for each block. The next 8 bits (*Rsum*) are then constructed by repeating the same procedure but on the second and fourth rows for each block. Figure 4 shows example of constructing 16 bits' watermark for certain block. At this stage, the combination of *Csum* and *Rsum* represents the 16 bit watermarks; *Ws* which are then changed to binary form to be embedded into the LSB of the block itself.



	C_1	C_2	C_3	C_4
R_1	00000000	01100000	01100100	11100000
R_2	00011100	00011100	11100000	11000000
R_3	10010000	11000010	11100000	00110000
R_4	01100000	11000000	10010000	01100010

(a) 4x4 block

(b) W_s construction

Fig. 4. Example of spatial domain fragile watermark construction

The combination of authentication bits in this manner gives the algorithm a high level of security and the ability to detect minor alterations. Any alteration in the pixel values will be reflected on the value of summation, and hence, on the watermark bits. In addition, embedding the watermark in the LSB gives many advantages such as simplicity, efficiency, less alteration in the value of the pixel which yields in an improved image perceptibility and high delicate to any manipulation or tampering [34]. After all the image blocks are watermarked, the image is rebuilt to its original dimensions from all the watermarked blocks.

4.1.3 Frequency Domain Fragile Watermark Generation Procedure: The second watermark (W_f) is then generated and embedded in the frequency domain by applying BEMD to the block watermarked image as shown in Figure 5.

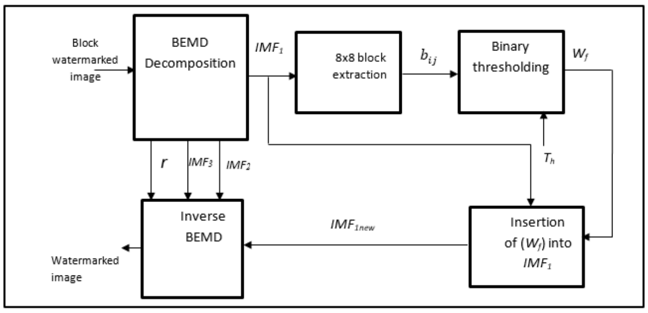


Fig. 5. frequency fragile watermark embedding

The BEMD decomposed the image into three *IMF*s and one residual component *r* according to the following equation:

$$I = \sum_{i=1}^{3}(IMFi) + r \qquad (3)$$

where $i \epsilon \{1,2,3\}$.

IMFs is extracted to be used as authentication codes, as they have been proven to produce effective feature in applications such as medical image analysis, scene analysis, and remote sensing [23] while the residue *r* component is untouched as it is the robust part of the image [27]. In addition, to make the algorithm more secure against counterfeiting attacks, the watermark is made dependent on the original image and derived from the image features [35]. In this regard, authentication code bits are derived from IMF_1 band and embedded into the same band by taking the first right - top corner blocks of size 4x4 from IMF_1. The watermark W_f is constructed by mapping the pixel values of the selected block into binary values according to a certain threshold (T_h) which is chosen empirically according to the following expression:

$$w_f \begin{cases} 1, & b_{ij} \geq T_h \\ 0, & othewise \end{cases}$$

where b_{ij} is the first right - top corner block , $i \epsilon \{1,\ldots,4\}$ and $j \epsilon \{1,\ldots.,4\}$

This process can identify pixel tampering effortlessly since any modification on the pixel stage would cause a disparity in the values of the watermark [12]. The binary watermark (W_f) is then iterated to be same dimension as the original IMF_1 and embedded into IMF_1 according to following equation:

$$IMF_{1new} = IMF_1 + W_f \qquad \underline{(5)}$$

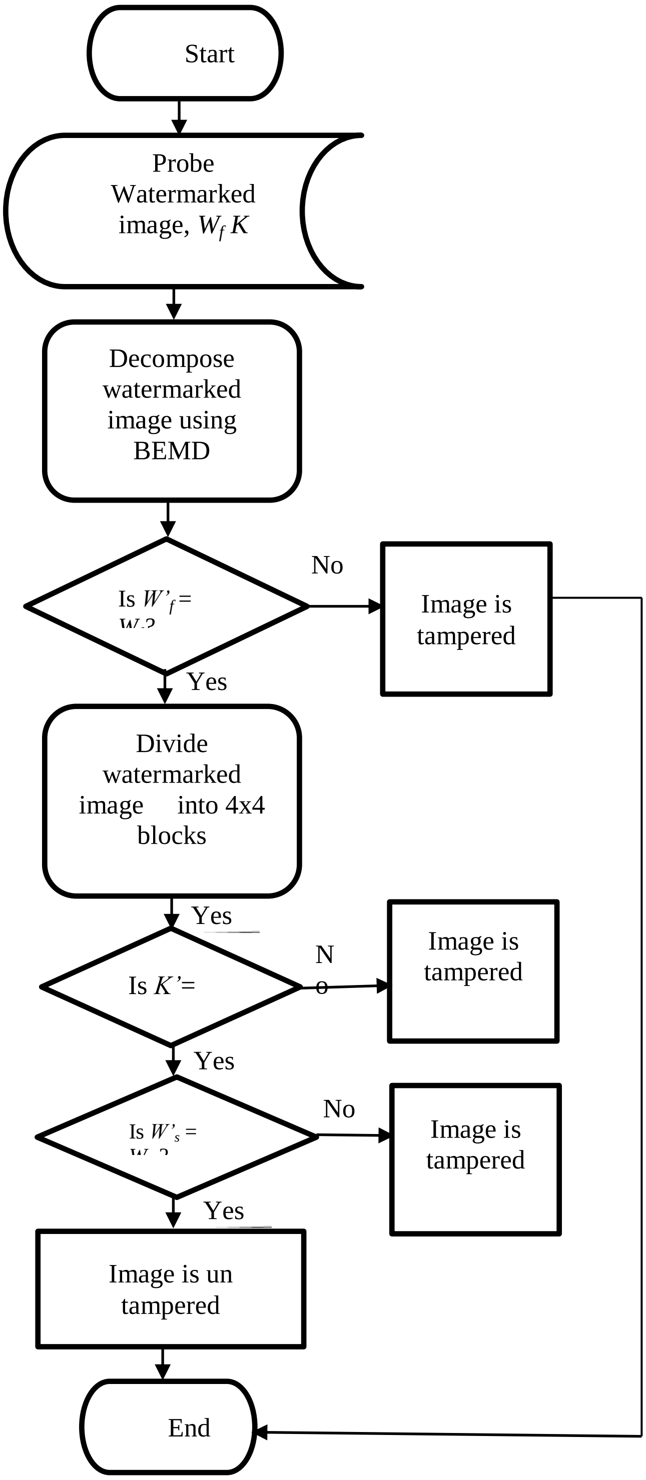


Fig. 6. Proposed authentication flowchar

where IMF_{1new} is the watermarked version of IMF_1.

After embedding the watermark, the watermarked image is built to spatial domain by applying inverse BEMD as follows:

$$I_{watermarked} = IMF_{1new} + IMF_2 + IMF_3 + r \quad (6)$$

4.2. Watermark generation procedure

In the proposed scheme, there are three levels of image verification as depicted in Figure 6. The first level of image authentication is performed in the frequency domain while the second and third levels are performed in the spatial domain. The authentication procedure is explained in the following subsections:

4.2.1 Frequency Domain Fragile Watermark Generation Procedure: The first procedure for authentication begins by decomposing the probe watermarked image using the BEMD into three *IMFs* and one residue *r.* The embedded $W'f$ is extracted from IMF_1 for the comparison with the original watermark bits W_f that provided by the user. To achieve image authentication, a binary error matrix (E_f) is generated by computing the bit error between the original and the extracted watermarks according to Equation (7).

$$E_f = (W_f \oplus W'_f) \quad (7)$$

where ⊕ is the XOR logical operation.

The mismatch and match between the two watermarks yields 1's and 0's in E_f, respectively. The tampered area can be allocated in terms of the bit error matrix as the most error pixels would cluster in distorted regions if tampering attacks were made on the watermarked image. On the other hand, the isolated error pixel is not considered as tampered because it is caused by unintentional attack [36].

.2.2 Secret Key Verifying Procedure: The second level of the image verification begins when the two watermarks; W_f and W'_f coincide and the image is rebuilt to the spatial domain by combining all *IMFs* and the residue component according to equation 4. Prior detecting the tampered bits, the secret key is verified by the receiver. First the user provides the receiver with the password *P*

4 which is then encrypted using the encryption procedure as described in section (4.1.1) to produce the encrypted key (*K*). The encrypted key is then compared with the referenced key (*K*') that embedded previously in the first four 4x4 blocks. The comparison is conducted for every single bit in each string. If the attacker modifies one bit of the message, then the key calculated by the receiver will vary from the received key. In this case, the corresponding watermarked image is considered as tampered, and the detection process is stopped. Otherwise, the image is considered as authentic and the third procedure of the authentication begins.

4.2.3 Spatial Domain Fragile Watermark Detection Procedure : The image authentication procedure continuous in the spatial domain by dividing the image into blocks of size 4×4 pixel and extracting the embedded watermark Ws' by getting the LSB of each pixel. The image verification procedure in this stage skips the four blocks that contain the secret key. As stated in section (4.1.2), the LSB of each pixel should be set to zero for creating the same template in both the embedding and extracting the watermarks. Another 16 bits watermark Ws is generated for each block by applying the same procedure as described in section (3.2.1.2). The first 8 bits of the watermark are produced by calculating the sum of the first and third columns of each block, then the second and fourth rows are summed to produce the second 8 bits of the watermark. once the 16 bits watermark is generated, the third level of authentication starts by comparing the newly produced watermark W_s and the referenced watermark W_s*'* for each single block using Equation 7.

5. Experimental results

Based on the watermark image embedding and the detection methods, as explained in detail in Section 4, different standard images which were downloaded from a trusted website [37] with dimension of (512 × 512 pixels), were tested. For objectively assessing the watermarked image's perceptual quality. The peak signal-to-noise ratio (*PSNR*) is utilised., The *PSNR* between the original image (*I*) and the watermarked image (*I'*) is given by:

$$PSNR = 10log_{10}\left[\frac{M\times N\times max(I)^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(I-I')^2}\right] \qquad (8)$$

where *M* and *N* are the dimensions of the image. The *PSNR* results for greyscale images are summarised in Table 1.

Table 1 shows the effect of adding the fragile watermarks on the quality of the watermarked image. The average value of *PSNR* value is 50.32. Generally, for ensuring that the inserted watermark is imperceptible, the image *PSNR* value must be higher than 35 dB [38]. Thus, the value of the achieved *PSNR* can be considered as tolerable and high.

For further evaluation, the embedding and extraction process was implemented for three cases; without attacks, under deletion attacks and under copy paste attacks as explained in the subsequent subsections.

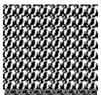
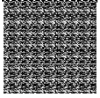
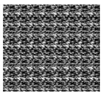
Table 1 *PSNR* values of proposed algorithm

Image	*PSNR*
Lena	50.65
woman dark hair	51.14
Peppers	50.23
Pirate	50.22
Womn_blonde	49.27
Jet Plane	50.28
Lake	50.64
mandrill	49.75
Elaine	50.73
Average	50.32
woman dark hair	51.14
Peppers	50.23
Average	50.32

5.1. *Fragile watermarks extraction without attacks*

To prove the ability of the algorithm in image authentication, first the fragile watermarks are extracted from the watermarked image without any attack and the bit error rate matrix according to equation (7) is computed between the original and extracted watermarks as shown in Table 2.

Table 2 Fragile watermarking Extraction

(a)Original image	(b)Generated watermark	(c) Watermarked image	(d)Extracted watermark
			
			

As shown in Table 2, when the watermarked image was untampered, the bit error rate contained all 0's. The 0's values refer to the matching between the embedded and the extracted watermarks [36]. Hence the watermark bits were all extracted successfully and the probe image was seen to be authentic.

5.2. *Fragile Watermarks Extraction under Deletion Attack*

The tamper detection capability of the fragile watermark is evaluated through the application of several tampering attacks such as deletion attack, which focuses on deleting part of the watermarked image. Different scales of this attack with different target locations are performed to evaluate the detection efficiency of the proposed scheme, as shown in Table 3.

Table 3 Tampered watermarked images

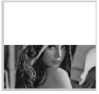
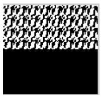
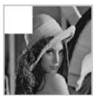
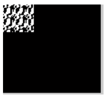
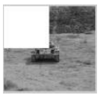
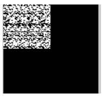
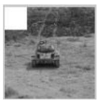
(a)Tampered Watermarked Image	(b)Bit Error Rate Matrix
	
	
	
	

Table 3 explains the tampered watermarked images and the corresponding bit error matrix after carrying out many attacks by deleting 50% and 25% of Lena and Tank watermarked images. The clustered region in the error matrix corresponds to erroneous pixel in which the tampered has been made to the watermarked image, wherein any mismatch between the embedded and extracted watermarks yields 1's [36].

Based on the above table, the proposed scheme could identify the deletion attacks that took place on the image with respect to the bit error rate matrix as a majority of

Table 4 Tamper detection rate (AV) for deletion attacks

Image	10% deletion	20% deletion	30% deletion	40% deletion	50% deletion
Lena	99.766	99.77	99.82	100	100
Cameraman	99.75	99.751	99.65	99.80	99.84
Peppers	99.75	99.751	99.78	99.80	99.84
Pirate	99.75	99.77	99.78	99.80	99.84
Womn_blonde	99.766	99.77	99.78	99.84	100
Jet Plane	99.75	99.766	99.80	99.84	100
Lake	99.766	99.75	99.77	99.84	100
Elaine	99.77	99.766	99.75	99.80	100
woman_darkhair	99.766	99.77	99.84	100	100
mandrill	99.77	99.80	99.84	100	100

the error pixels were seen to cluster in the distorted regions of the image if any size of tampering attacks were carried out on the watermarked image.

False positive (F_P), false negative (F_N) and average of detection rate (A_V) are the optimal methods used to evaluate the tampering detection capability of an authentication algorithm. In the tampering authentication procedure, the false positive value refers to the number of pixels detected as tampered pixels although they are untampered. By contrast, the false negative value refers to the number of tampered pixels that are detected as untampered.

The values of F_P, F_N and A_V are calculated according to the following equations:

$$fp = \frac{number\ of\ pixels\ in\ untamperd\ region,detected\ as\ tampered}{total\ number\ \ of\ pixels\ in\ untampered\ region}$$
(9)

$$fN = \frac{number\ of\ pixels\ in\ tamperd\ region,detected\ as\ untampered}{total\ number\ \ of\ pixels\ in\ tampered\ region}$$
(10)

$$AV = 1 - \left(\frac{fN + fP}{1+R}\right) \times 100$$
(11)

where R stands for the number of tampered regions. To evaluate the algorithm fairly against a deletion attack, another factor that should be considered is the size of the tampering area: 10%, 20%, 30%, 40%, and 50% are selected for the tampering attacks. In Table 4, the average detection rate values are calculated for standard greyscale images with dimensions of 512 × 512 according to equation 11.

According to all the observations noted in Table 4, it can be seen that the attained detection rate (AV) for the proposed algorithm is greater than 99.75% for the case of 10% tampering. Thus, the proposed algorithm could efficiently detect the tampering attacks with high detection rate even when the image was tampered slightly (10%).

The above table also shows that when the tampering region size decreases, the attained AV decreased from 99.84 (for 50%) to 99.75% (for 10%), in case of cameraman watermarked image. In general, with decreasing of the size of tampering, the detection rate will be decreased [33]. However, the decreasing value is not noticeable as the percentage degradation in AV is less than 9%, for cameraman watermarked image.

Table 5 Tampered Tank image

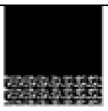
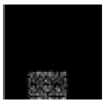
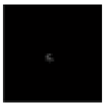
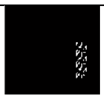
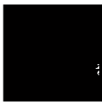
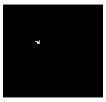
Tampered image	Error bit matrix
(a)	 (b)
(c)	 (d)
(e)	 (f)

Table 6 Tampered Lena image

Tampered image	Error bit matrix
(a)	 (b)
(c)	 (d)
(e)	 (f)

5.3. Fragile Watermarks Extraction under Copy-pate Attack

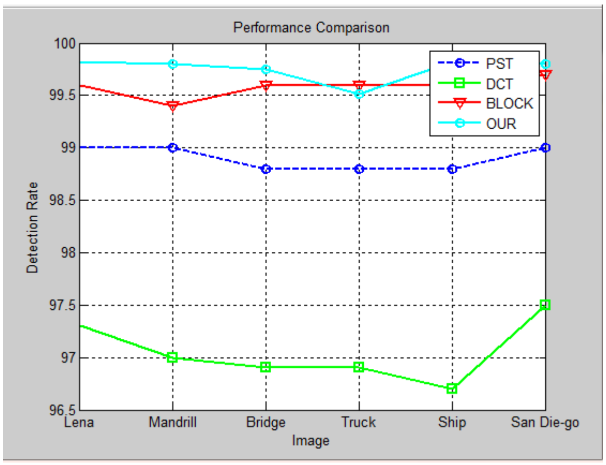


Fig. 7. Comparison of the proposed algorithm with others for 30% tampering rate

Another critical tampering attack that should be considered is the copy–paste attack, which is performed by either copying a certain region from the watermarked image itself or from outside the image and then pasting it somewhere in the image. Tables 5 and 6 depict the watermarked images which were altered by cutting certain areas from the images and then pasting the regions for constructing forged images. All the simulations and the respective results obtained have been described as follows:

Tank: In the case of the watermarked Tank image, the image was altered by inserting a different Tank having some surrounding area in the image (Table 5a). Another tampering was carried out by copying the Tank image without any surrounding area, as described in Table 5c. Table 5e describes the copying of a small region (a black rectangular area) in the Tank image. All the results for tamper detection have been described in Table 5b, 5d and 5f, respectively.

Table 7 Benchmarking the proposed scheme with others in terms of copy-paste attack

Image	**10% copy–paste/ *AV***				**20% copy–paste/ *AV***				**30% copy–paste/ *AV***			
	PST	**DCT**	**Block**	**Our**	**PST**	**DCT**	**Block**	**Our**	**PST**	**DCT**	**Block**	**OUR**
Lena	97.6	99.5	99.70	*99.77*	98.70	97.10	99.78	**99.77**	99.0	97.3	99.6	99.82
Mandrill	97.3	96.3	99.80	**99.70**	98.80	96.50	99.57	99.80	99.0	97.0	99.4	99.80
Bridge	97.5	95.4	99.70	99.77	98.60	97.00	99.78	99.81	98.8	96.9	99.6	99.75
Trucks	97.6	95.1	99.80	99.83	98.70	96.70	99.55	***99.65***	98.8	96.9	99.6	**99.51**
Ship	97.6	95.2	99.70	99.81	98.80	96.20	99.78	99.80	98.8	96.7	99.6	99.82
San Die-go	97.6	95.4	99.80	***99.77***	98.70	96.60	99.78	99.81	99.0	97.5	99.7	99.80
PSNR(dB)	39	-	51.0042	50.32								

Lena: Table 6a describes a modification of the Lena image, wherein the feather present in her hat was copied and pasted to a different region of the woman's face. Another modification included the copying of a black rectangular region which was then pasted in her hat's shadow Table 6c. Table 6e describes another modification that was carried out where the clasp was pasted above her hat (a very small region). Tables 6b, 6d and 6f describe all the results for tampering detection.

Tables 5 and 6 depict the tampered watermarked images and the corresponding bit error matrix after performing many copy paste attacks on Lena and Tank watermarked images. The clustered region in the error matrix corresponds to erroneous pixel in which the tampering has been made to the watermarked image, wherein any mismatch between the embedded and extracted watermarks yields 1's [36].

Based on the above table, the proposed algorithm was able to detect the different tampering even when small tampering areas were carried out which involved pasting a small clasp on the Lena hat or adding a small black rectangular region to the Tank image

5.4. Performance Benchmarking with Related Algorithms

For benchmarking of the proposed algorithm performance, three separate percentages of the copy-paste attacks were carried out. The algorithm was benchmarked with three different image authentication algorithms; Ho et al. [39] employed the Pinned Sine Transform (PST), Lin et al. [17] used the Discreet Cosine Transform (DCT), while Dadkhah et al. [33] used the Block wise-based fragile watermarking scheme while the proposed scheme employs Block wise and BEMD. In this study, standard images size of 512 × 512 pixels were used, which were similar to the earlier studies, as described in Table 7. In addition, Figure 7 compared the algorithm performance with the three algorithms for a tampering rate of 30%.

The comparative results show that the proposed scheme could display a better performance as compared to the existing schemes with respect to the tamper detection rate

(*AV*), except limited cases which are underlined. It is known that the most effective system is the one which could detect

tampering even in a low-tampering region [33]. The low tampering rate of 10% proved that our algorithm was very accurate and more effective as compared to the three related algorithms since the corresponding tamper detection rate was 99.77% while preserving high *PSNR*, 50.32 dB. In addition, the results of copy paste attack of the proposed algorithm are compared with that suggested by Hsu and Tu [40] in terms of *PSNR* and false positive fp values as depicted in Table 8.

The comparative results displayed that the PSNR *of the proposed scheme was 50.32, on average, are higher than that of the scheme proposed by Hsu and Tu [40] by 10.995 dB (Table 8). Moreover, the false positive rate of the proposed system was 0.00253, on average, around 0.00187 dB times lower than that of the system proposed by Hsu and Tu [40]. It can be summed up that the proposed system in this paper has a better implementation as compared to the related woks with respect to the false positive rates and, thus, is extremely applicable to image tampering detection.*

6. CONCLUSION

This study introduced an efficient authentication verification algorithm that employs BEMD and implemented the algorithm for grey scale images. The proposed algorithm can detect tampering at the pixel level with a high detection rate value and a high quality of the watermarked image at approximately 50 dB. This good performance is achieved because the embedding process is implemented in the time and frequency domains. In addition, algorithm security is improved by employing a secret key to prevent an unauthorised user from tampering with the image even if the person has complete knowledge of the embedding algorithm.

REFERENCES

[1] X. Qi and X. Xin, “A quantization-based semi-fragile watermarking scheme for image content authentication,” *J. Vis. Commun. Image Represent.*, vol. 22, no. 2, pp. 187–200, 2011.

[2] M. M. Yeun and F. Mintzer, “An invisible watermarking technique for image verification,” in *Proceedings of IEEE International Conference on Image Processing*, 1997, vol. 2, pp. 680–683.

[3] H. Yuan and X.-P. Zhang, “Multiscale fragile watermarking based on the Gaussian mixture model.,” *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3189–200, Oct. 2006.

[4] H. Yuan and X.-P. Zhang, “Multiscale fragile watermarking based on the Gaussian mixture model,” *Image Process. IEEE Trans.*, vol. 15, no. 10, pp. 3189–3200, 2006.

[5] J. Fridrich, “Security of fragile authentication watermarks with localization,” *Electron. Imaging*, vol. 10, no. 2, pp. 691–700, 2002.

[6] S. H. Liu, H. X. Yao, W. Gao, and Y. L. Liu, “An image fragile watermark scheme based on chaotic image pattern and pixel-pairs,” *Appl. Math. Comput.*, vol. 185, no. 2, pp. 869–882, 2007.

[7] S. Che, B. Ma, and Z. Che, “An adaptive and fragile image watermarking algorithm based on composite chaotic iterative dynamic system,” in *proceeding of International Conference on Intelligent Information Hiding and Multimedia Signal Processing.*, 2008, pp. 159–162.

[8] E. T. Lin, E. J. Delp, and W. Lafayette, “A Review of Fragile Image Watermarks,” in *proceeding of 1st Int. Conf. Innovative Computing, Information and Control*, 2001.

[9] A. T. S. Ho, X. Zhu, J. Shen, and P. Marziliano, “Fragile Watermarking Based on Encoding of the Zeroes of the-Transform,” *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 3, pp. 567–569, 2008.

[10] X. Zhang, S. Wang, and G. Feng, “Fragile watermarking scheme with extensive content restoration capability,” in *Digital Watermarking*, Springer, 2009, pp. 268–278.

[11] X. Zhang and S. Wang, “Statistical fragile watermarking capable of locating individual tampered pixels,” *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 727–730, 2007.

[12] Y. Lim, C. Xu, and D. D. Feng, “Web based image authentication using invisible fragile watermark,” in *Proceedings of the Pan-Sydney area workshop on Visual information processing-Volume 11*, 2001, pp. 31–34.

[13] P. W. Wong, “A public key watermark for image verification and authentication,” in *Proceedings of IEEE International Conference on Image Processing*, 1998, vol. 1, pp. 455–459.

[14] S. Suthaharan, “Logistic map-based fragile watermarking for pixel level tamper detection and resistance,” *EURASIP J. Inf. Secur.*, vol. 2010, no. 1, pp. 1–7, Oct. 2010.

[15] J. K. Mandal and S. K. Ghosal, “A Fragile Watermarking Based on Separable Discrete Hartly Transform for Color Image Authentication (FWSDHTCIA),” vol. 3, no. 6, pp. 23–33, 2012.

[16] N. Ahmidi and R. Safabakhsh, “A novel DCT-based approach for secure color image watermarking,” in *Proceeding of International Conference on Information Technology: Coding and Computing*, 2004, vol. 2, pp. 709–713.

[17] E. T. Lin, C. I. Podilchuk, and E. J. Delp III, “Detection of image alterations using semifragile watermarks,” in *proceeding of IEEE international conference society for optics and phonics*, 2000, pp. 152–163.

[18] C.-H. Chou and Y.-C. Li, “A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 6, pp. 467–476, 1995.

[19] X. Zhao, P. Bateman, and A. T. S. Ho, “Image authentication using active watermarking and passive forensics techniques,” in *Multimedia Analysis, Processing and Communications*, Springer, 2011, pp. 139–183.

[20] A. B. Watson and A. Poirson, “Separable two-dimensional discrete Hartley transform,” *JOSA A*, vol. 3, no. 12, pp. 2001–2004, 1986.

[21] R. S. Alomari and A. Al-jaber, “A Fragile Watermarking Algorithm for Content Authentication,” *Int. J. Comput. Inf. Sci.*, vol. 2, no. 1, pp. 27–37, 2004.

[22] N. Bi, Q. Sun, D. Huang, Z. Yang, and J. Huang, “Robust image watermarking based on multiband wavelets and empirical mode decomposition,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 1956–1966, 2007.

[23] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, “Image analysis by bidimensional empirical mode decomposition,” *Image Vis. Comput.*, vol. 21, no. 12, pp. 1019–1026, 2003.

[24] R. J. Oonincx and P. J. Oonincx, “Empirical mode decomposition: a new tool for S-wave detection,” 2002.

[25] D. Dunn, W. E. Higgins, and J. Wakeley, “Texture segmentation using 2-D Gabor elementary functions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 2, pp. 130–149, 1994.

[26] W. H. W. Huang and Y. S. Y. Sun, “A New Image Watermarking Algorithm Using BEMD Method,” in *proceeding of IEEE International Conference on Communications, Circuits and Systems*, 2007, pp. 588–592.

[27] P. Kovesi, “Phase congruency detects corners and edges,” in *proceeding of australian conference on pattern recognition society*, 2003.

[28] S. Mallat, “Wavelets for a vision,” *Proc. IEEE*, vol. 84, no. 4, pp. 604–614, 1996.

[29] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. R. Soc. London. Ser. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[30] J. Wan, L. Ren, and C. Zhao, “Image feature extraction based on the two-dimensional empirical mode decomposition,” in *the International Congress on Image and Signal Processing, Hainan, China*, 2008, vol. 1, pp. 627–631.

[31] A. SABRI, M. KAROUD, H. TAIRI, and A. AARAB, “Image Watermarking Using the Empirical Mode Decomposition,” 1796.

[32] R. G. Van Schyndel, A. Z. Tirkel, and C. F. Osborne, “A digital watermark,” in *Proceedings of IEEE International Conference Image Processing*, 1994, vol. 2, pp. 86–90.

[33] S. Dadkhah, A. A. Manaf, and S. Sadeghi, “Efficient image authentication and tamper localization algorithm using active watermarking,” in *Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations*, Springer, 2014, pp. 115–148.

[34] F. Mintzer and G. W. Braudaway, “If one watermark is good, are more better?,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, vol. 4, pp. 2067–2069.

[35] F. Deguillaume, S. Voloshynovskiy, and T. Pun, “Secure hybrid robust watermarking resistant against tampering and copy attack,” *Signal Processing*, vol. 83, no. 10, pp. 2133–2170, Oct. 2003.

[36] X. Xin, “A Singular-Value-Based Semi-Fragile Watermarking Scheme for Image Content Authentication with Tampering Localization,” *J. Vis. Commun. Image Represent.*, vol. 30, pp. 312–327, 2010.

[37] G. Weber, “USC-SIPI report image database: Version 4,” 1993.

[38] Y.-P. Lee, J.-C. Lee, W.-K. Chen, K.-C. Chang, J. Su, and C.-P. Chang, “High-payload image hiding with quality recovery using tri-way pixel-value differencing,” *Inf. Sci. (Ny).*, vol. 191, pp. 214–225, 2012.

[39] A. T. S. Ho, X. Zhu, and W. M. Woon, “A semi-fragile pinned sine transform watermarking system for content authentication of satellite images,” in *proceeding of Symposium in International Geoscience and Remote Sensing (IGARSS)*, 2005, vol. 2, pp. 737–740.

[40] C. S. Hsu and S. F. Tu, “Image tamper detection and recovery using adaptive embedding rules,” *Meas. J. Int. Meas. Confed.*, vol. 88, pp. 287–296, 2016.