Design and Analysis of a Fragile Watermarking Scheme Based on Block-Mapping

Munkhbaatar Doyoddorj, Kyung-Hyune Rhee

Design and Analysis of a Fragile Watermarking Scheme Based on Block-Mapping*

Munkhbaatar Doyoddorj and Kyung-Hyune Rhee**

Department of IT Convergence and Application Engineering,
Pukyong National University.
599-1, Daeyeon3-Dong, Nam-Gu, Busan 608-737, Republic of Korea
{d_mbtr, khrhee}@pknu.ac.kr

Abstract. Due to the wide variety of attacks and the difficulties of developing an accurate statistical model of host features, the structure of the watermark detector is derived by considering a simplified channel model. In this paper, we present a fragile watermarking based on block-mapping mechanism which can perfectly recover the host image from its tampered version by generating a reference data. By investigating characteristics of watermark detector, we make an effective analysis such as fragility against robustness measure and distinguish its property. In particular, we derive a watermark detector structure with simplified channel model which focuses on the error probability versus watermark-to-noise-ratio curve and describes a design by calculating the performance of technique, where attacks are either absent or as noise addition.

Keywords : Fragile Watermarking, Characterization, Block-Mapping, Tamper Localization

1 Introduction

In the past decades, the advent of versatile digital multimedia processing tools has made multimedia duplication and manipulations much easier. The availability of such powerful tools, however, has also provided opportunities for theft and misuse of intellectual properties. As a result, multimedia authentication and integrity verification have become a popular research area in recent years. To address both the authentication and integrity issues, a wide variety of schemes have been proposed for different applications. The authentication schemes can be divided into two categories: digital signature based [13] and digital watermark based [14] schemes. A digital signature can be either an encrypted or a signed hash value of image contents and image characteristics. The major drawback of signature based scheme has limitation to identify the modified regions, that is, it can detect whether the image has been modified or not, however, it cannot

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2012-0001331).

** Corresponding author

locate the regions where the image has been modified. To solve this problem, many researchers have proposed digital watermarking based schemes for image authentication.

Watermarking schemes are an alternatives to the concept of cryptographic signatures, specially designed to embed authentication and integrity data within media objects, thus eliminate the need for separate storage. They occur in different security scenarios:

- *Robust* watermarking [10] may be employed if for instance the origin of a media object needs to be determined to trace illicit reproduction. Robust watermarks withstand most digital processing operations in video clips and digital images and can be recognized even after several alterations. However, in order to provide such a tamper resistant method, straightforward usage of cryptographic signatures is all but impossible.
- *Fragile* watermarking [9] may be employed if the integrity of a media object needs to be proven to deem its content authentic. They designed to be instantly destroyed when the media object is tampered with. These schemes are commonly used for tamper detection (integrity proof). Modifications to an original work are clearly noticeable.

Generally, the digital watermarking for integrity verification is called fragile watermarking as compared to robust watermarking for copyright protection [1]. The fragile watermark can serve as an embedded signature to guarantee the authenticity of the data. Ideally, a fragile watermark might even reveal, through how it has been distorted, what processing the original data has undergone. To localize the tampered area, the fragile watermarking techniques for image authentication usually partition the image into blocks with the same size. For each block, watermark data are generated based on the secret key and then inserted into the least significant bits (LSBs) of the same block [5-8]. This way, the tampered block can be easily located by checking the consistency of the content and the embedded watermark in itself.

In watermarking techniques, there have many aspects and problems to be considered, such as embedding domains, characteristics, human perception system (auditory or visual), attacks, security and specific application requirements, so on. We consider the characteristics of watermark detector such that robustness and fragility are intimately connected to, respectively, the copyright protection and the integrity of the host data. Therefore, the investigation of these characteristics is of great interest in order to support a design of techniques for a wide application scenario. To understand the possible functions of watermark detector and in order to reduce complexity of measure, we utilize a simplified channel model with noise addition that is based on information theoretic considerations [4]. The creator develops a piece of watermarked variable content and sends it through a simplified channel and then analyze how the distortions during the processing are affected when simplified channel model is used to simulate distortions.

Our contribution. The aim of this paper is to demonstrate the characteristics of fragile watermarking scheme through an analysis of watermark detector and to design a scheme for tampering detection and recovery on the spatial domain. We derive the watermark detector structure a simple case, dealing with simplified channel model, where attacks are either absent or modeled as noise addition. Then we evaluate the detection error probability and watermark-to-noise-ratio of our scheme for the simplified channel, being aware that a more accurate experimental analysis is needed to assess the performance of schemes in realistic situations. Hence, the fragility is measured by two types of authentication error probabilities. The result of analysis show that a fragile watermarking scheme is most appropriate to achieve the best trade-off between both error probabilities.

On the other hand, our scheme is based on the block-mapping mechanism which identifies the blocks containing tampered pixels whereas the authentication data allows individual localization of the tampered pixels [18]. Also, how the block-mapping can be used as the scheme to address fragile requirements and embedding distortions in a practical watermarking scheme. The regular structure of such mapping provides efficiency for encryption and verification algorithms that are straightforward to analyze.

The rest of this paper is organized as follows: Section 2 introduces some basic notations and definitions of watermarking schemes and concepts of block division and mapping strategy used in this paper. We formulate the fragile watermark scheme based on block-mapping and derive the embedding and detection (verification) structures in Section 3. In Section 4, we introduce the analysis of proposed scheme, and experimental results are shown in Section 5. Conclusion is drawn in Section 6.

2 Preliminaries

2.1 Notations and Definitions

In this section, we present basic notations and formal abstracted definitions used in the paper. Basically, we follow the terminology in [15]. We write $Y \leftarrow \texttt{Alg}(X)$ to denote running algorithm `Alg` on input X and assigning the output to variable Y. Optional inputs and outputs are set in squared brackets in $\texttt{Alg}(X_1, [X_2])$, the input of X_2 is optional.

To specify the probability, we use the notation $\mathbf{Prob}[\texttt{assign}(\upsilon_1, ..., \upsilon_n) :: \texttt{pred}(\upsilon_1, ..., \upsilon_n)]$. This denotes the probability that the predicate `pred` holds when the probability is taken over a probability space defined by the formula `assign` on the n variables υ_i of the predicate `pred`.

A negligible function $\epsilon(x)$ is a function where the inverse of any polynomial is asymptotically an upper bound, $\forall d > 0$, $\exists x_0, \forall x > x_0 : \epsilon(x) < 1/x^d$. We denote this by $\epsilon(x) < 1/poly(x)$. If $\epsilon(x)$ cannot be upper bounded in such a way, we say $\epsilon(x)$ is not negligible.

We define 1^n to denote the bit string consisting of n 1's. Finally, we define Boolean values $ind \in \{1, 0\}$, which the presence and absence of the watermark, respectively.

A suitable similarity function or predicate is a key aspect in definition of watermarking schemes. We assume a suitable polynomial time computable *similarity* function as follows:

Definition 1 (Similarity Function) *The polynomial time computable ideal similarity function* $sim(Y^\star, Y^\circ)$ *for given two items* $Y^\star$ *and* Y°*, outputs* 1 *iff* $Y^\star$ *can be considered sufficiently similar to* Y° *(as a manner of usual, agreed semantics), and has been derived from* Y°*. Note that* $sim()$ *does not need to be symmetric.*

Definition 2 (Detecting Watermarking Scheme) *A detecting watermarking scheme* $W = (GenKey, Emb, Det)$ *consists of three probabilistic polynomial time algorithms. We define the intactness or imperceptibility property as formally:*

- **Key generation algorithm:** On input of the security parameter 1^n, the key generation algorithm $GenKey(1^n)$ generates the matching keys (K_E, K_D) required for watermark embedding and detection, respectively.
- **Embedding algorithm:** On input of the host data (cover-data) X, the watermark W_R to be embedded with the key K_E, the probabilistic embedding algorithm $Emb(X, W_R, K_E)$ outputs the watermarked data (stego-data) Y, which is required to be perceptibly similar to the host data X.

$$Y \leftarrow Emb(X, W_R, K_E) \text{ then } sim(X, Y) = 1$$

- **Detection algorithm:** On input of (possibly modified) watermarked data Y', the watermark W_R, the *optional* data (sometimes also referred to as reference data in this context), detection key K_D, the probabilistic detection algorithm $Det(Y', W_R, [optional], K_D)$ outputs a Boolean value $\{0, 1\}$, it is commonly referred to effectiveness of the watermarking scheme.

$$Y' \leftarrow Emb(X, W_R, K_E) \wedge Det(Y', W_R, [optional], K_D) = 1$$

Remark 1. We refer to a watermarking scheme as being symmetric iff $K_E = K_D$ and in this case, we usually denote both keys as K_W. But otherwise, the scheme with $K_E \neq K_D$ is called asymmetric.

Definition 3 (Extracting Watermark Scheme) *An extracting watermark scheme is similarly defined, where a probabilistic extraction algorithm* $Ext()$ *instead of the detection algorithm that on input of watermarked data* Y *and the extraction key* K_{EX} *outputs the watermark contained in this data or the Boolean value* 0 *if it cannot extract any watermark.*

- **Extraction algorithm:** On input of (modified) watermarked data Y', the optional data (sometimes also referred to as reference data), and the extraction key K_{EX}, the probabilistic extraction algorithm $Ext(Y', [optional], K_{EX})$ either outputs the watermark W_R contained in Y' or fails with output 0.

$$Y' \leftarrow Emb(X, W_R, K_E) \wedge W'_R \leftarrow Ext(Y', [optional], K_{EX}) \text{ then } W'_R = W_R$$

When employing fragile watermarking schemes, the embedding process induces distortions into the original media object, thus inevitably altering the original. Although sophisticated embedding algorithms induce a barely visible distortion into the media object, a lossless reconstruction may be desirable. A fragile watermarking scheme can detect alterations even if the underlying digital work has been (maliciously) modified, as long as the scheme is very sensitive to the slight changes, more formally:

Definition 4 (Fragile watermarking) *A watermarking scheme is called fragile, iff it is computationally infeasible for a probabilistic polynomial-bounded adversary $\mathcal{A}$, given watermarked data Y and the watermark W_R, to produce perceptibly different and altered data Y'.*

$$\mathbf{Prob}[K_W \leftarrow GenKey(1^n); Y \leftarrow Emb(X, W_R, K_W); Y' \leftarrow \mathcal{A}(Y, 1^n); \\ :: Det(Y', W_R, [optional], K_W) = 1 \wedge sim(Y', Y) = 1] < \epsilon(x).$$

The main application of fragile watermarking is data authentication, where watermark loss or alteration is taken as an evidence that data has been tampered with, whereas the recovery of the information contained within the data is used to demonstrate origin.

2.2 The Block Division and Mapping Strategy

We will introduce two concepts of block mapping algorithm in this section. These strategies effectively break block-wise independency, and makes the self-recovery watermarking scheme invulnerable against the counterfeiting attacks [11]. An object X is partitioned into non-overlapping blocks $B_i (i = 1, ..., N)$ of 2×2 pixels by the block division as follows:

Definition 5 (Block Division) *A block division scheme is as a tuple of two probabilistic polynomial algorithms $\langle SEPARATE, JOIN \rangle$. On input X and a size of the block $S(m \times m)$, the algorithm $SEPARATE$ produces a tuple of blocks $B_1 \| \cdots \| B_N$. The algorithm $JOIN$ inverts the algorithm $SEPARATE$, on input $B_1 \| \cdots \| B_N$ with S, it outputs X.*

Except with negligible probability, we require that

$$JOIN(S, SEPARATE(X, S)) = X,$$

for object X and size S with $SEPARATE(X, S) \neq FAIL$.

Using secret key K_S, a pseudo-random sequence $ps = (ps_1, ps_2, ..., ps_N)$ is firstly produced, and then an ordered index sequence $(a_1, a_2, ..., a_N)$ such that $(ps_{a_1}, ps_{a_2}, ..., ps_{a_N})$ is obtained by sorting out the pseudo-random sequence ps. For each block, assign the index of block B_i to be $B_i = B_{a_i}$ such that $i = a_i$.

Here, the watermark information of the block B_1 is embedded in the block B_2, the watermark of block B_2 is embedded in the block B_3, and the watermark of block B_3 is embedded in the block B_4, and so on.

A random indexed block sequence (RIBS) algorithm generates a pairs of randomly distributed block sequences, such that on input $(\{B_i\}, K_S)$, where block sequence $(\{B_i\}|i = 1, ..., N)$ and a key K_S is generated by key generation function. It outputs a generated pairs of random indexed block sequence, as follows:

$$B_{pair} = \{(B_i, B_{i+1})|i = 1, ..., N\}.$$

The B_{pair} is the block pairs of random indexed block sequence, B_{i+1} is the next block of B_i. According to the security and tamper localization [12], the next block B_{i+1} of each block B_i should be randomly distributed in the whole image.

3 The Proposed Scheme

In this section, we introduce a formal definition of proposed fragile watermark scheme. This scheme provides the block-mapping strategy on the image content. Essentially, we apply the formal definition of watermarking scheme described in the previous section on each block B_i in X, with the exception that the there is some linkage (computed by a reference data extraction function) between the blocks. Technically, we rely on the concept of block-mapping sequence. In particular, our scheme embeds a watermark consisting of authentication and reference data for each block B_i of the host image into the generated block pair B_{i+1} by using the block-mapping construction. On the watermark detector, one can identify the tampered blocks by comparing the extracted authentication data in B_{i+1} with the calculated authentication one in B_i. The reliable reference data, which extracted from block pair B_{i+1} is used to exactly reconstruct the host image, if the block B_i is tampered. Furthermore, our scheme is sensitive to any tiny changes in images so that it provides an ability of optimized tampering localization while it keeps robustness against incidental distortions.
In short, we consider the following requirements for fragile watermarking scheme.

- Robustness and fragility objectives should be simultaneously addressed. When both cannot be completely achieved, one must have a quantitative mechanism to tradeoff between these two objectives.
- The fragile authentication system must be secure against an intentional tampering. For security, it must be computationally infeasible for the opponent to devise a fraudulent message.
- If the watermark is an authenticator, then embedding must be imperceptible.
- The authentication embedding and verification algorithms must be computationally efficient, especially for real time applications.

3.1 General Descriptions

We incorporate a random indexed block sequence mechanism into a fragile watermarking scheme for constructing a fragile watermarking based on block-mapping scheme [18]. The proposed scheme is specified as follows:

Definition 6 (FWBM) *We say that a four triple of probabilistic polynomial time algorithms* $FWBM = (GenKey, Emb, Det, Rec)$ *is a fragile watermarking scheme based on block-mapping iff*

- **Key generation algorithm:** Algorithm $GenKey$ generates the necessary keys for the application. $GenKey$ runs (1^n) to generate a triple tuple of keys $\langle K_S, K_E, K_D \rangle$.
- **Embedding algorithm:** Algorithm Emb takes K_E, a size of block S and an object X. The reference data $(R_{i,j}|i = 1, ..., N$ and $j = 1, ..., 8)$ is extracted from each pixel of block B_i and then embedded into LSB_3 of corresponding pixel in pair B_{i+1}. The output of the algorithm consists of an embedded object Y.
- **Detection algorithm:** Embedded objects can be detected (verified) by the algorithm Det with a public key. Algorithm Det takes the verification key K_D, a size of block S and an embedded object Y. The authentication data is extracted from each pixel of block B_i, while compared with calculated authentication data in pair of block B_{i+1} and outputs a boolean variable.
- **Reconstruction algorithm:** Finally, the algorithm Rec reverses the embedding mechanism and losslessly reconstructs X out of modified object Y'. Rec extracts the reference data $R_{i,j}$ from a pair of tampered block B_{i+1} in modified object Y' and reconstruct the tampered block B_i by using extracted reference data that takes the keys K_S, a modified object Y' and output a recovered object X'.

Note that we have defined all algorithms as probabilistic, which implies that they can fail on certain instances (for example it may not be possible to embed a watermark in an invertible manner); in this case, the algorithms output a special symbol fail. We require that the scheme works for almost all objects that can be authenticated. In particular,

$$Det(Emb(X, K_E, [R]), K_D) = 1 \wedge Rec(Emb(X, K_E), [R]) = X$$

must hold except for a negligible fraction of all objects X with $Emb(X, K_E) \neq FAIL$.

3.2 Construction

The detailed description of the scheme are given as follows:

- **Key generation algorithm:** On input (1^n) the key generation algorithm $GenKey$ outputs a triple of keys $\langle K_S, K_E, K_D \rangle = KeyGen(1^n)$, respectively.

The key K_S will be used in to generate the block-mapping step, whereas K_E, K_D are used for embedding and detection(verification). The detection key K_D is a public, whereas keys K_S and K_E are private keys.

– **Embedding algorithm:** On input of Emb takes an object X, a size of block S and keys K_S, K_E. The algorithm produces the following steps:

1. Divide an object X into blocks: $B_1 \| \cdots \| B_N \leftarrow SEPARATE(X, S)$.
2. Generate a random indexed block sequence by using K_S:
$B_{pair} \leftarrow RIBS(B_1 \| \cdots \| B_N, K_S)$, where $B_{pair} \in \{(B_i, B_{i+1})\}$.
for $i = 1, ..., N$ **do**
3. Extract the reference data $R_{i,j}$ and authentication bits (parity $p_{i,j}$ and check $c_{i,j}$ bits) from each pixel of B_i :
$B_i \in \langle R_{i,j}, p_{i,j}, c_{i,j} \rangle$, where $R_{i,j} = (MSB_3 \oplus LSB_3)$.
4. Embed into $B_{i+1} \leftarrow B_{i+1} \in \langle LSB_3, p_{i+1,j}, c_{i+1,j} \rangle \oplus B_i \in \langle R_{i,j}, p_{i,j}, c_{i,j} \rangle \oplus K_E$.
end for
5. Each embedded blocks are joined to output: $Y \leftarrow JOIN(B_1 \| \cdots \| B_N)$.
output Watermarked object Y

– **Detection algorithm:** On input of Det takes an embedded object Y, a size of block S and keys K_S, K_D, as follows:

1. Divide an object Y into blocks: $B_1 \| \cdots \| B_N \leftarrow SEPARATE(Y, S)$.
2. Generate a random indexed block sequence by using key K_S:
$B_{pair} \leftarrow RIBS(B_1 \| \cdots \| B_N, K_S)$, where $B_{pair} \in \{(B_i, B_{i+1})\}$.
for $i = 1, ..., N$ **do**
3. In order to separate the reference data $R_{i,j}$ and authentication bits (parity $p_{i,j}$ and check $c_{i,j}$ bits) of each pixel in B_i from B_{i+1} :
$B_{i+1} \in (R_{i,j}) \leftarrow B_{i+1} \in (LSB_3) \oplus K_D$,
and $B_i \leftarrow B_{i+1} \in (LSB_3) \oplus B_{i+1} \in (R_{i,j}) \oplus K_D$.
4. Calculate the authentication bits $p_{i,j}$ and $c_{i,j}$ of block B_i.
5. Check the authentication bits between the blocks B_i and B_{i+1}:
if ($p_{i,j} == p_{i+1,j}$ and $c_{i,j} == c_{i+1,j}$) **exit with 1**
return Locations of tampered block l.
exit with 0
end for
6. $Y \leftarrow JOIN(B_1 \| \cdots \| B_N)$.
output Boolean value (1 or 0)

– **Reconstruction algorithm:** On input of Rec takes a modified object Y', a size of block S and a locations of tampered block l as follows:

1. Divide an object Y' into blocks: $B_1 \| \cdots \| B_N \leftarrow SEPARATE(Y', S)$.

2. Generate a random indexed block sequence by using key K_S:
$B_{pair} \leftarrow RIBS(B_1 \| \cdots \| B_N, K_S)$, where $B_{pair} \in \{(B_i, B_{i+1})\}$.
for $i = 1, ..., N$ **do**
3. Get the location of tampered block l.
4. Extract the reference data's of block $B_{l+1,j} : B_{l+1} \in R_{l+1,j} \leftarrow REF(B_{l+1})$.
5. Recover the modified blocks: $B_{l,j} = (B_{l,j} \| R_{l+1,j})$
end for
5. $X' \leftarrow JOIN(B_1 \| \cdots \| B_N)$.
output Recovered object X' and $sim(X', X) = 1$

Remark 2. A construction of pixels in a block is denoted as follows: $\langle MSB_3 \| LSB_3 \| p \| c \rangle_8$, where MSB_3 are three most significant bits, LSB_3 are three least significant bits and authentication bits p and c, which represent a parity and check bits.

4 Analysis of Proposed Scheme

A common way of modeling for watermarking schemes is to treat them as communication systems with appropriate transmission channels. A modeling for watermarking schemes as communication system has advantages for designing and analysing of concrete schemes, because it allows to draw from the established results in the field of communications and signal processing of certain data types. In this section, we only deal with a single aspect of the problem, that is, the analysis of the *robustness* and $fragility$ characteristics of the embedded watermark. Two key concepts are central to our discussion. The robustness refers to the ability to reliably extract the watermark information (keeping the watermark detection error probability low) even when the amount of noise introduced by the attacker is large. This robustness condition is mostly desirable for host copyright protection applications. On the other hand, a fragility refers to the ability to prevent the digital watermark from being detected even when the intensity of the attack is low. This fragility condition is mostly desirable for host authenticity and integrity verification applications.

4.1 Problem Formulation

We consider the watermarking as a communication model, where watermark communication channel is characterized by possible attacks against the embedded watermark. In other words, which is called a simplified channel model [19].

One specifically interesting attack is the addition of white Gaussian noise, which can be applied easily so that each watermarking scheme should show good robustness at least against this type of attacks. Since Gaussian noise is the most harmful power-limited additive noise with respect to mutual information, and many derivations can be performed analytically when a Gaussian distribution is employed. Figure 1 depicts the described simplified channel model of watermarking as communication scenario, where an attack by an Gaussian noise is assumed.

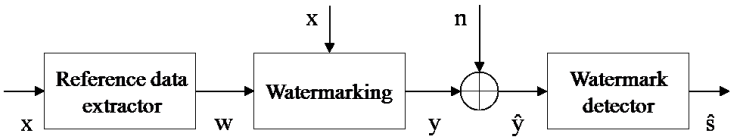

Fig. 1. Problem model for digital watermarking

The problem of simplified channel model is described as follows: The reference data extractor derives from the *host data* $\mathbf{x}$ and to generate the *watermark* $\mathbf{w}$, which is added to the host data to produce the *watermarked data* $\mathbf{y}$. An *embedding distortion* D_w in the watermarked data $\mathbf{y}$ in relation to $\mathbf{x}$ due to the modulation with the watermark $\mathbf{w}$. The $\mathbf{w}$ must be chosen such that the distortion between $\mathbf{x}$ and $\mathbf{y}$ is negligible, which is defined as:

$$D_w = \left[\frac{1}{N} \sum_{i=1}^{n} E\{(y_i - x_i)^2\} \right] < \epsilon(x),$$

where $E\{\cdot\}$ denotes expectation and N is the number of samples. Note that D_w also represents the watermark power ϕ^2_{wm}.

Next, the watermarked data $\mathbf{y}$ might be processed by Gaussian noise $\mathbf{n} \in \mathbb{R}^N$. Such processing potentially impairs watermark communication and thus is denoted as an attacks against the embedded watermark. In general, attacks against watermark are only constrained with respect to the distortion between $\mathbf{x}$ and $\widehat{\mathbf{y}}$. The watermark detector input $\widehat{\mathbf{y}} = \mathbf{y}+\mathbf{n}$ is presented, which represents the attacked watermarked data. Now, we can express the distortion of watermarked data $\widehat{\mathbf{y}}$ in relation to $\mathbf{x}$ as:

$$D_{\widehat{y}} = D_w + n = \phi^2_{wm} + \phi^2_n.$$

The objective of the watermark detector is to produce the best estimate of the watermark from the attacked watermarked data. So that, the receiver must be able to detect the watermark $\widehat{\mathbf{s}}$ from the received data $\widehat{\mathbf{y}}$.

In following, we introduce some parameters for future analysis. The robustness and fragility behavior analysis will be focused on their dependency to the performance of the watermarking technique. The performance of technique is taken here as the detection error probability p_e. In practice, p_e is estimated by the bit error rate (BER) measurement, corresponding to the number of wrongly estimated bits, from the attacked signal, over the total number of embedded bits. The strength of attacks is also measured by the watermark to noise ratio (WNR), giving the ratio between the power of $\widehat{\mathbf{y}}$ and that of the noise $\mathbf{n} = \widehat{\mathbf{y}} - \mathbf{y}$, as follows:

$$WNR = \frac{\sum_{i=1}^{n} E\{\widehat{y}^2\}}{\sum_{i=1}^{n} E\{(D_{\widehat{y}} - D_w)^2\}} = \frac{\sum_{i=1}^{n} E\{\widehat{y}^2\}}{\sum_{i=1}^{n} E\{n^2\}} = \frac{E\{\widehat{y}^2\}}{E\{n^2\}} = \frac{\phi^2_{\widehat{y}}}{\phi^2_n}$$

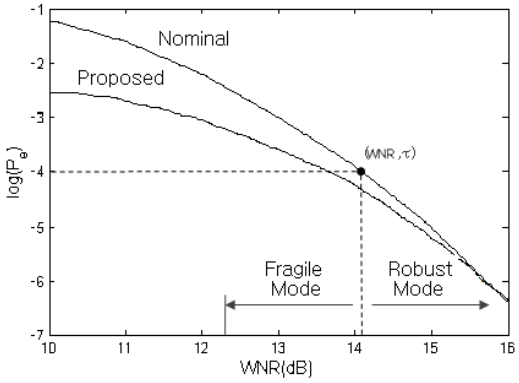

Fig. 2. The characterization curve for a nominal and our scheme

4.2 Characterization for Watermarking Schemes

We introduce some design parameters that take into account the robustness and fragility. The first parameter establishes an upper bound to the maximum allowed host distortion D_{max}. Above this distortion, for many copyright protection verification applications, the received signal is considered *useless*. In a covert communication application scenario, if this distortion is exceeded to some extent, the user will easily notice that a third part tried to jam the secret communication. A second parameter establishes a threshold τ for the detection error probability p_e. Above this threshold, the recovered watermark is no longer considered as a reliable one.

A characterization of fragile watermarking scheme in operation modes is classified as the following conditions:

$$OperationModes = \begin{cases} Robust \\ mode \quad \text{if } (p_e < \tau) \text{ and } (D_{\widehat{y}} > D_{max}); \\ Fragile \\ mode \quad \text{if } (p_e > \tau) \text{ and } (D_{\widehat{y}} < D_{max}). \end{cases}$$

We emphasize that inequalities of above described conditions in a definition are represent the robustness and fragility analysis. In the absence of an attack (no noise), the WNR is infinite and the detection error probability is negligible or zero. As the attack intensity increases, the WNR decreases, which is describing on the WNR versus p_e characterization curve of nominal value as shown in Figure 2. If the WNR decreases, the p_e is reached, the scheme is said to be a fragile, otherwise the scheme is said to be a robust.

The performance of our scheme was used to plot in Figure 2. The parameter p_e was chosen to be 10^{-4}, yielding to $WNR = 14.1$dB. The design can now be pursued by selecting appropriate values for the parameter τ and D_{max}. Our scheme is stated in fragile operation mode according to above definition, which is guaranteed by parameters $p_e = 10^{-3.9}$ and $WNR = 13.5$dB.

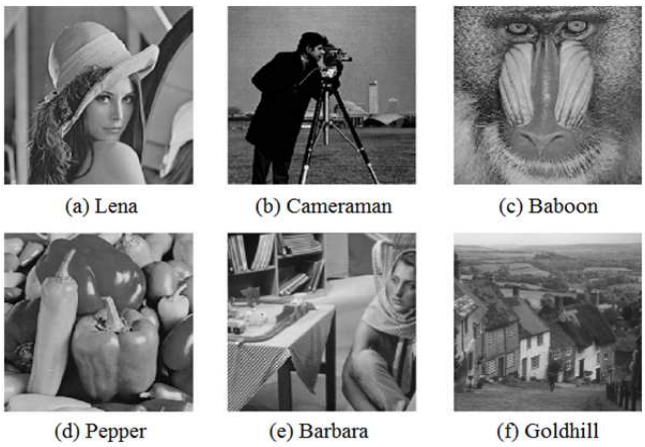

Fig. 3. Grayscale test images (512×512).

5 Experimental Results

In this section, we describe our experiments and discuss the results. We simulated our scheme under a PC with 1.8G Hz Dual CPU, 6G RAM, and Windows Vista platform. The simulation was carried out using Matlab version R2008a. In order to evaluate the performance of our proposed scheme, we considered six commonly used grayscale images with the size of 512×512 as shown in Fig. 3.

With respect to objective evaluation, the peak signal to noise ratio (PSNR) was used to measure the visual quality of fidelity for the host image and the watermarked image. Among the watermarked images, the image qualities measured by PSNR value were greater than 44dB, where an attack by a Gaussian noise signal $v \sim N(0, \sigma_v^2)$ is assumed. The PSNR value highly depends on the size of tampered regions and the accuracy of tampered block identification. The greater the PSNR, the better the performance of image recovery technique.

Table 1. The performance comparison

.	Proposed		Zhang et. al [16]		Zhu et. al [17]	
Test Images	Watermarked PSNR (dB)	Recovered PSNR (dB)	Wat.d PSNR (dB)	Rec.d PSNR (dB)	Wat.d PSNR (dB)	Rec.d PSNR (dB)
Lena	45.63	32.23				
Cameraman	44.95	32.81				
Baboon	45.32	30.65	Average	Average	Average	Average
Pepper	45.53	29.37	39.90	27.79	36.70	22.80
Barbara	44.06	30.62				
Goldhill	45.49	31.86				

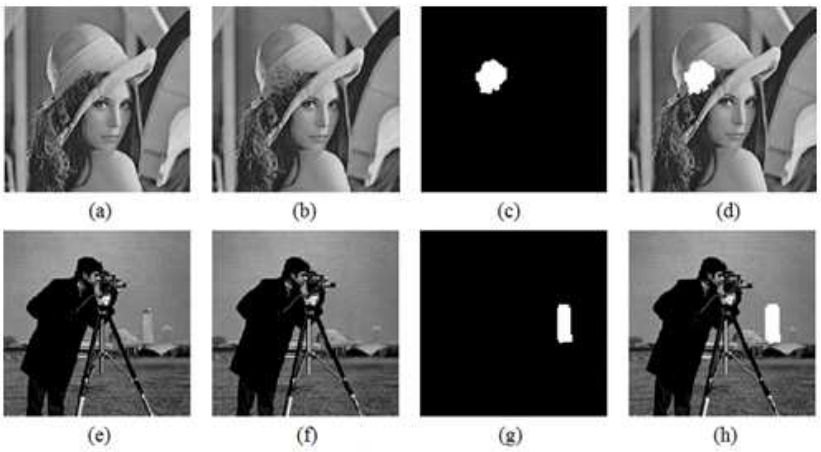

Fig. 4. Example of ordinary tampering detection. (a),(e) The host images, (b),(f) Watermarked images, (c),(g) Tampered block detection and (d),(h) Tamper localization.

Table 1 shows the performance of proposed scheme for all test images by comparing with related methods. From these comparisons, it is observed that our scheme has achieved the higher PSNR values of watermarked and recovered images.

We consider ordinary tampering. Two test images Lena and Cameraman sized 512×512 are used as the host image, in Fig. 4(a),(e). The PSNR values due to watermark embedding are 45.63dB and 44.95dB, respectively. We modify the watermarked images by extensively replacing the original content with fake information as shown in Fig. 4(b),(f), and the tampering rates are 8.15% and 6.36%. Fig. 4(c),(g) gives the result of tampered block detection, in which the blocks judged as valid and invalid are indicated by black and white areas. Here, all tampered blocks were correctly located in Fig. 4 (d),(h). Finally, we calculated the PSNR values of recovered images, which are 43.26dB and 41.09dB, respectively. These results indicate that after identifying the tampered blocks, our scheme exactly locate the tampered pixels and perfectly restore the watermarked version. The computational complexity of our scheme is light since it does not need to apply any transform such as discrete cosine transform (DCT) and Fourier transform (FFT). The required processing mainly lies on generating the RIBS, scanning pixels, and embedding and decryption using XOR operation in spatial domain. Hence, the execution time is rather short.

6 Conclusion

In this paper, we introduced design and analysis of a fragile watermarking scheme with tampering localization and recovery mechanism. The focus of our analysis is on characteristics of watermark detector and distinguish its property such as

fragility against robustness measure. In order to design effective watermarking scheme, we analyzed the characteristics of our scheme by defining the characterization mode. The proposed scheme utilizes the block-mapping strategy to identify the tampered region by generating the random indexing block sequence. By using this technique, we can detect any modifications made to the image and indicate the specific locations where the modification was made. As compared with some previous works, the proposed scheme based on block-mapping on spatial domain not only is as simple and as effective in tamper detection and localization, but also provides the capability of tamper recovery by trading off the quality of the watermarked images about 44dB. This implies that the proposed scheme can offer high embedding quality and low image degradation. The experimental results confirm the effectiveness of our scheme by demonstrating that the watermarked image with acceptable visual quality can be recovered as well as tampering detection and localization.

References

1. C. Rey, J.L. Dugelay, "A survey of watermarking algorithms for image authentication," EURASTP Appl Signal Process (6) 613-621, (2002).
2. S. Suthaharan, "Fragile image watermarking using a gradient image for improved localization and security," Pattern Recogn Lett 25 (16) 1893-1903 (2004).
3. P.-L. Lin, C.-K. Hsieh, P.-W. Huang, "A hierarchical digital watermarking method for image tamper detection and recovery," Pattern Recogn 38 (12) 2519-2529 (2005).
4. N.Merhav, E.Sabbag, "Optimal Watermark Embedding and Detection Strategies Under Limited Detection Resources,"IEEE Transactions on Information Theory, 54(1) (2008).
5. M. Celik, G. Sharma, E. Saber, A.M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," IEEE Transaction on Image Processessing, 11(6) 585- 595 (2002).
6. S. Suthaharan, "Fragile image watermarking using a gradient image for improved localization and security," Pattern Recognition Letters, 25, 1893-1903 (2004).
7. S. Liu, H. Yao, W. Gao, Y. Liu, "An image fragile watermark scheme based on chaotic image pattern and pixel-pairs," Applied Mathematics Computation, 185(2), 869-882 (2007).
8. Wang MS, Chen WC, "A majority-voting based watermarking scheme for color image tamper detection and recovery," Computer Standards and Interfaces (29), 561-570 (2007).
9. M. Yeung and F. Mintzer, "Invisible watermarking for image verification," Journal of Electronic Imaging (7), 578-591 (1998).
10. M.D. Swanson, Zhu Bin, and A.H. Tewfik, "Transparent robust image watermarking," IEEE In International Conference on Image Processing (3), 211-214 IEEE Computer Society Press, (1996).
11. M.Holliman, N.Melon, "Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes," IEEE Trans. Image Processing 9 (3), 432-441 (2000).
12. H.J.He, J.S.Zhang, and H.X.Wang, "Synchronous counterfeiting attacks on self-embedding watermarking schemes," Internetional Journal. Computer Science and Network Security 6 (1), 251-257 (2006)

13. M.Tagliasacchi, G.Valenzise, and S.Tubaro, "Hash-Based Identification of Sparse Image Tampering," IEEE Transactions on Image Processing 18(11) 2491-2504 (2009).
14. C-C.Lai, C-C,Tsai, "Digital Image Watermarking Using Discrete Wavelet Transform and Singular Value Decomposition," IEEE Transactions on Instrumentation and Measurement 59(11) 3060-3063 (2010).
15. A.Adelsbach, S.Katzenbeisser, and A.-R.Sadegni, "A Computational Model for Watermark Robustness," IH 2006, LNCS 4437, 145-160 (2007).
16. X. Zhang, S. Wang, "Fragile watermarking scheme using a hierarchical mechanism," Signal Processing, (89) 675-679 (2009).
17. X. Zhu, A. Ho, P. Marziliano, "A new semi-fragile image watermarking with robust tampering restoration using irregular sampling," Signal Processing and Image Communication (22) 515-528 (2009).
18. H.J.He, F.Chen, H-M.Tai, "Performance Analysis of a Block-Neighborhood-Based Self-Recovery Fragile Watermarking Scheme," IEEE Transactions on Information Forensics and Security, 7(1), 185-196 (2012).
19. R.Baeuml, J.J.Eggers, R.Tzschoppe, J.Huber, "Channel model for watermarks subject to desynchronization attacks," Proc. SPIE, Security and Watermarking of Multimedia Contents IV, 4675 281-292 (2002).