A Robust Fragile Watermarking Technique for Digital Image

Biswa Mohan Sahoo#1, Jyostnamayee Behera#2 ,Ranjeet Kumar Rout#3
Amity University, Greater Noida, GIFT,
Bhubaneswar, NIT Jalandhar

Abstract: **Watermarking differs from authentication or digit al signature that proves to a receiver that the message could only have come from one particular transmitter. Mostly, authentication messages in the form of conventional hash functions can easily be deleted by a pirate who wishes to use copyrighted material for illegal purposes. The goal is to give the copyright owner of a digit al image (or other piece of information) the possibility to attest technically the origin of the image.**

Keywords: Watermarking, Security, Image, communication channel, imperceptibility.

1. INTRODUCTION:

The enormous popularity of the World Wide Web in the early 1990's demonstrated the commercial potential of offering multimedia resources through the digital networks. Since commercial interests seek to use the digital networks to offer digital media for profit, they have a strong interest in protecting their ownership rights. Digital watermarking has been proposed as one way to accomplish this.

A digital watermark is a digital signal or pattern inserted into a digital image. Since this signal or pattern is present in each unaltered copy of the original image, the digital watermark may also serve as a digital signature for the copies. A given watermark may be unique to each copy (e.g. to identify the intended recipient), or be common to multiple copies (e.g. to identify the document source). In either case, the watermarking of the document involves the transformation of the original into another form. This distinguishes digital watermarking from digital fingerprinting, where the original file remains intact and a new created file 'describes' the original file's content.

Digital watermarking is also to be contrasted with public-key encryption, which also transform original files into another form. It is a common practice nowadays to encrypt digital documents so that they become un-viewable without the decryption key.

Unlike encryption, however, digital watermarking leaves the original image (or file) basically intact and recognizable. In addition, digital watermarks, as signatures, may not be validated without special software. Further, decrypted documents are free of any residual effects of encryption, whereas digital watermarks are designed to be persistent in viewing, printing, or subsequent re-transmission or dissemination

2. ELEMENTS OF WATER MARKING SYSTEM

A watermarking system can be viewed as a communication system consisting of three main elements: an embedder, a communication channel and a detector. Watermark information is embedded into the signal itself, instead of being placed in the header of a file or using encryption like in other security techniques, in such a way that it is extractable by the detector. To be more specific, the watermark information is embedded within the host signal before the watermarked signal is transmitted over the communication channel, so that the watermark can be detected at the receiving end, that is, at the detector [1].

A general watermarking system is illustrated in Fig. 1. The dotted lines represent the optional components, which may or may not be required according to the application. First of all, a watermark Wo is generated by the watermark generator possibly with a secret watermark generation key Kg [2]. The watermark Wo can be a logo, or be a pseudo-random signal.

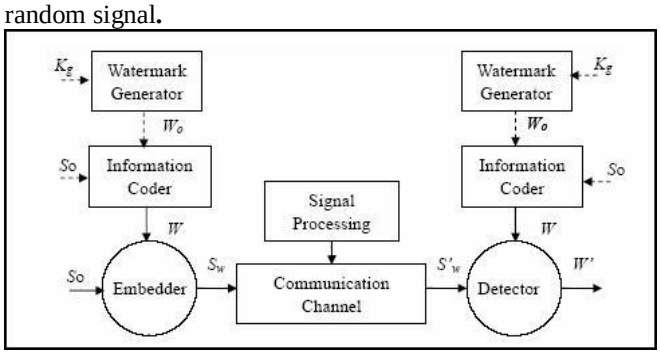

Fig 1. General Watermarking System

Instead of directly embedding it into the host signal, the watermark Wo can be pre-coded to optimize the embedding process, i.e. to increase robustness against possible signal processing operations or imperceptibility of the watermark. This is done by an information coder which may require the original signal so the outcome of the information coding component is denoted by symbol W that, together with the original signal So and possibly a secret key K, are taken as input of the embedder. The secret key K is intended to differentiate between authorized users and unauthorized users at the detector in the absence of key Kg. The embedder takes in W and K as to hide W in a most imperceptible way with the help of K, and produce the watermarked signal Sw. Afterwards, Sw enters into the communication channel where a series of unknown signal processing operations and attacks may take place. The outcome of the communication channel is denoted by the symbol Sw.

At the receiving end, the detector works in an inversely similar way as the embedder, and it may require the secret key Kg, K, and the original signal so Then the detector reads Sw and decides if the received signal has the legal watermark [3].

3. TYPES OF DIGITAL WATER MARKING

Watermarks and watermarking techniques can be divided into various categories in various ways. Watermarking techniques can be divided into four categories according to the type of document to be watermarked as follows: [5]

- Text Watermarking
- Image Watermarking
- Audio Watermarking
- Video Watermarking

In other way, the digital watermarks can be divided into three different types as follows: [4]

- Visible watermark
- Invisible-Robust watermark
- Invisible-Fragile watermark

Visible watermark is a secondary translucent overlaid into the primary image. The watermark appears visible to a casual viewer on a careful inspection. The invisible-robust watermark is embedded in such a way that alternations made to the pixel value are perceptually not noticed and it can be recovered only with appropriate decoding mechanism. The invisible- fragile watermark is embedded in such a way that any manipulation or modification of the image would alter or destroy the watermark.

Also, the digital watermarks can be divided into two different types according to the necessary data for extraction:

3.1 Informed (or private Watermarking):

In this watermarking the original non watermarked cover is required to perform the extraction process.

3.2 Blind (or public Watermarking):

In which the original non watermarked cover is not required to perform the extraction process.

4. WATERMARKING REQUIREMENTS

There are many different requirements of the digital watermarking. These all requirements are as in following sub sections.

4.1 Security

The security requirement of a watermarking system can differ slightly depending on the application. Watermarking security implies that the watermark should be difficult to remove or alter without damaging the host signal. As all watermarking systems seek to protect watermark information, without loss of generality, watermarking security can be regarded as the ability to assure secrecy and integrity of the watermark information, and resist malicious attacks [5].

4.2 Imperceptibility

The imperceptibility refers to the perceptual transparency of the watermark. Ideally, no perceptible difference between the watermarked and original signal should exist. A straightforward way to reduce distortion during watermarking process is embedding the watermark into the perceptually insignificant portion of the host signal. However, this makes it easy for an attacker to alter the watermark information without being noticed.

4.3 Capacity

Watermarking capacity normally refers to the amount of information that can be embedded into a host signal. Generally speaking, capacity requirement always struggle against two other important

requirements, that is, imperceptibility and robustness given under figure 2. A higher capacity is usually obtained at the expense of either robustness strength or imperceptibility, or both.

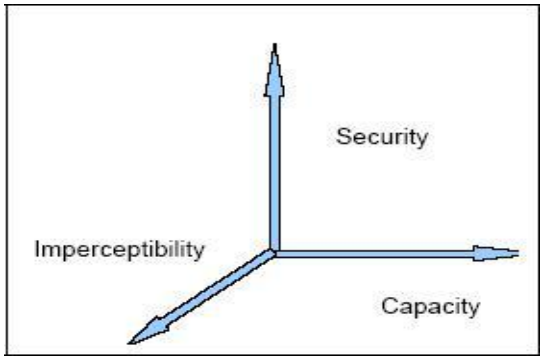

Fig 2. The tradeoffs among imperceptibility, Robustness, and capacity

4.4 Robustness

Watermark robustness accounts for the capability of the watermark to survive signal manipulations. Apart from malicious attacks, common signal processing operations can pose a threat to the detection of watermark, thus making it desirable to design a watermark that can survive those operations. For example, a good strategy to robustly embed a watermark into an image is to insert it into perceptually significant parts of the image. Therefore, robustness is guaranteed when we consider the case of lossy compression which usually discards perceptually insignificant data, thus data hidden in perceptual significant portions is likely to survive lossy compression operation. However, as this portion of the host signal is more sensitive to alterations, watermarking may produce visible distortions in the host signal. The exact level of robustness an algorithm must possess cannot be specified without considering the application scenario [6]. Not all watermarking applications require a watermark to be robust enough to survive all attacks and signal processing operations. Indeed, a watermark needs only to survive the attacks and those signal processing operations that are likely to occur during the period when the watermarked signal is in communication channel. In an extreme case, robustness may be completely irrelevant in some case where fragility is desirable.

5. WATER MARKING TECHNIQUE

Several different methods enable watermarking in the spatial domain. The simplest (too simple for many applications) is just to flip the lowest-order bit of chosen pixels. This works well only if the image is not subject to any modification. A more robust watermark can be embedded by superimposing a symbol over an area of the picture. The resulting mark may be visible or not, depending upon the intensity value. Picture cropping, e.g., (a common operation of image editors), can be used to eliminate the watermark.

5.1 Spatial watermarking

can also be applied using color separation. In this way, the watermark appears in only one of the color bands. This renders the watermark visibly subtle such that it is difficult to detect under regular viewing. However, the mark appears immediately when the colors are separated for printing. This renders the document useless for the printer unless the watermark can be removed from the color band. This approach is used commercially for journalists to inspect digital pictures from a photo-stock house before buying unmarked versions.

Watermarking can be applied in the frequency domain (and other transform domains) by first applying a transform like the Fast Fourier Transform (FFT). In a similar manner to spatial domain watermarking, the values of chosen frequencies can be altered from the original. Since high frequencies will be lost by compression or scaling, the watermark signal is applied to lower frequencies, or better yet, applied adaptively to frequencies that contain important information of the original picture. Since watermarks applied to the frequency domain will be dispersed over the entirety of the spatial image upon inverse transformation, this method is not as susceptible to defeat by cropping as the spatial technique. However, there is more a tradeoff here between invisibility and decidability, since the watermark is in effect applied indiscriminately across the spatial image.

6. FRAGILE WATER MARKING

A watermark is said to be fragile if the watermark hidden within the host signal is destroyed as soon as the watermarked signal undergoes any manipulation. When a fragile watermark is present in a signal, we can infer, with a high probability, that the signal has not been altered.

Fragile watermarking authentication has an interesting variety of functionalities including tampering localization and discrimination between malicious and non-malicious manipulations. Tampering localization is critical because knowledge

of where the image has been altered can be effectively used to indicate the valid region of the image, to infer the motive and the possible adversaries. Moreover, the type of alteration may be determined from the knowledge of tampering localization [7].

As to the fragile watermarks for authentication and proof of integrity, the attacker is no longer interested in making the watermarks unreadable. Actually, disturbing this type of watermark is easy because of its fragility. The goal of the attackers is, conversely, producing a fake but legally watermarked signal. This host media forgery can be reached by either making undetectable modifications on the watermarked signal or inserting a fake watermark into a desirable signal.

Now, it is necessary to formulate the unique features of fragile watermarking systems in order to demonstrate what features are well sought after. The features can also serve in theoretical analysis for making comparisons among algorithms:

1. High resolution tampering localization: This becomes an important merit of fragile watermarking systems as it is one of the features that makes watermarking outweighs cryptography in some applications. The outcome of a detector can be as simple as authentic/tampered, but a result indicating which portions in an image are tampered is more desirable.

2. Tampering detection with low false positive. A good fragile watermarking system should have a sound tamper indication stating both statistical tampering probability and tampering localization with a low false positive rate.

6.1 Fragile Watermarking Scheme Exploiting Non-Deterministic Block-Wise Dependency

The combination of contextual and non-deterministic information has been proved to be one of the most effective means to improve the security of a watermarking system [3, 7]. The discussed scheme [7] depends on both of the contextual information and non-deterministic information to perform watermarking [8]. Contextual information, or dependency information, refers to the information from other portions of the image. Non-deterministic information usually relies on some randomly chosen parameters so that it can produce a unique signature. In this way, the non-deterministic signatures of two identical blocks at the same position of two images are different, even when they have the same neighborhood. As it is proved later, dependency information is recognized as effective mean to counter basic forgery attacks while the addition of non-deterministic information can thwart more advanced ones.

6.2 Symbol Definitions

Symbols to be used in the scheme are defined as follows:

f: the original gray scale image

f(i): the gray scale of the ith pixel of f

fM(i): the seven most significant bits of f(i)

fl(i): the least significant bit of f(i)

f ′: the image received by the watermark detector. If not tampered with, f ′ is the watermarked version of f

Z: the size of the image f

w: the secret key generated binary watermark of the same size M as the image f

w(i): the ith bit of w

w′: the extracted binary watermark by the decoder

w′(i): the ith bit of w ′

D: the binary difference map between w and w ′ with its ith pixel denoted as D(i) (D(i) ∈ {0, 255}) indicating whether

w(i) and **w ′ (i)** are different. Wherever the watermarked image is manipulated,

noises are shown in the corresponding portion of the difference map D. We could also identify what type of manipulation has been done from D

k: the length of dependency neighbourhood**. L:** the size of the neighbourhood, L= k × k.
N(i): the square dependency neighbourhood centred at pixel i consisting of k × k pixels including pixel i itself

S(i): the secret non-deterministic dependency information of pixel i calculated.

6.3 The Embedding and Detection Algorithm

In this scheme, only the seven most significant bits are used, which remain intact during the embedding process, of the gray scales of the pixels in the neighbourhood to generate the non-deterministic dependency information. This is important because the embedding no longer has to follow a specific scan order.

The most important part in the embedding process is the method of generating the secret dependency information S. First of all, S should depend on the un- altered part of the neighbourhood but not the to-be- marked pixel itself. Secondly, the secret neighbourhood is chosen in such a way that it is as random as possible and the attacker should not be able to learn enough from the resulting signal to succeed in a counterfeiting attack. Therefore, we introduce non-deterministic information in creating the secret information.

The watermark embedding algorithm can now be described in figure 3 and watermark detection algorithm in figure 4.

1. Generate a binary watermark w with the secret key shared with the detector.

2. For each pixel **i** in image **f**

a. Calculate the secret dependency information **S(i)** according to

$$\sum_{M} w(j) \cdot f(j).$$

where ⊕ denotes the exclusive-or **(XOR)** operation.

b. Adjust $f_I(i)$ so that the following equation holds

PARITY(f(i)+s(i))=w(i)

Fig 3. Watermark Embedding Pseudo Code

The binary watermark w is used as a selection determinant. From the XOR operation in Eq. (1), we know that a pixel fM(j) in N(i) is chosen only when the w(i) and w(j) are different. The factor (-1)w(j) in Eq. (1) determines whether fM(j) is to be added to or subtracted from S(i). Involving the watermark bits unknown to the third party in Eq. (1) introduces non-deterministic information and, thus, allows the scheme to counter the aforementioned forgery attacks.

1. Generate a binary watermark w with the secret key shared with the embedder.

2. For each pixel i

a. Calculate the secret dependency information **S(i)** from the received image **f'** according to Eq **(1)**

b. Extract the watermark bit **w ' (i)** according to:

w ' (i)=PARITY(f'(i)+S(i))..... (3)

c. Calculate the Difference D(i) between w(i) and the extracted watermark w'(i) using:

D(i)=0 ,w(i)=w'(i) ..... (4) D(i)=255, otherwise ..... (4)

Fig 4. Watermark Detection Pseudo Code

Wherever the watermarked image is manipulated, noises are shown in the corresponding portion of the difference map D [9]. We can also possibly identify what type of manipulation has been done from D.

7. EXPERIMENTS ON ATTACKS

The main content-altering modifications must raise tamper alarm in the detector, so does the non- malicious signal processing manipulations. The common non-permissible alterations are listed as follows:

- Common signal processing such as lossless/lossy compression, low pass filtering.
- Image manipulations that modify the geometry of objects such as rotation, flipping, translation, scale and cropping
- Image forgeries intended to remove, substitute or insert objects in the scene

In the following experiments regarding to the above alterations, neighborhood size is set to be 9×9 unless stated.

7.1 Low pass filtering attack

A low pass filtering is done over the watermarked image and it results in a difference map composed of noise. The scenarios of low pass filter attack is given in figure 5.

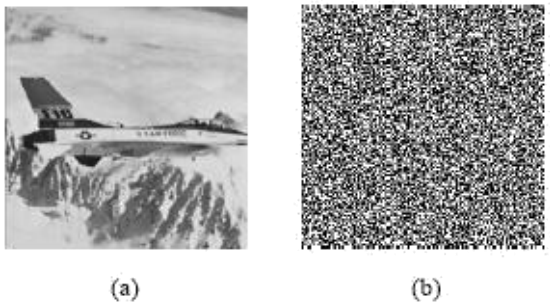

Fig 5. Detecting Low Pass Filter, (a) Watermarked Image, (b) Difference map after the watermarked image is low pass filtered.

7.2 Geometric attack

All manipulations that affect the geometry of the image such as flipping, rotation, cropping, etc. should be detectable. A cropping attack from the right-hand side and the bottom of the image is illustrated in Fig. 6.

Figure 6. Detecting cropping attack, (a) The cropped image, (b) The difference map showing the cropping trace.

In the difference map in Fig. 4, there are two obvious noisy stripes along the borders of the image denoting the tampering. One good feature of the proposed scheme demonstrated here is that it is able to differentiate cropping attack from global manipulations such as scaling, low-pass filtering, and histogram equalization, which, when mounted, would give rise to an entirely noisy difference map.

7.3 Forgery attack:

The forgery attacks that result in object insertion and deletion, scene background changes are all tantamount to substitution. It follows then that the variety of forgery attacks considered above can be collapsed to substitution attack only. Thus the following experiment is against substitution attack by replacing the watermarked image portion with another portion from the same image.

In Fig.7(a) a substitution attack is performed in an attempt to conceal the existence of the characters on the fuselage of the jet fighter. Fig. 7(b) highlights the exact corresponding region, which has actually been tampered with (i.e., the ground truth). The difference map D in Fig. 7(c) with the size of the neighbourhood equal to 9×9 indicates the authentication result before turning the missed/false alarms on/off. The difference map D in Fig. 7(d) indicates the authentication result with the size of the neighbourhood equal to 7×7. It clearly shows that the inauthentic region has been significantly shrunk closer to the actual tampered region while the interior of the cluster has become denser.

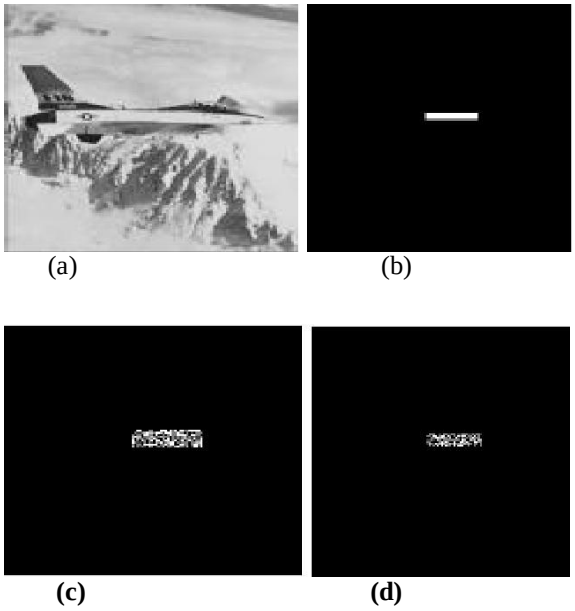

Figure 7. Detecting forgery attack, (a) The tampered image, (b) The actual region tampered, (c) The difference map with $k=9$, (d) the difference map with $k=7$

It is clear that the size of the neighbourhood has certain influence on the tampering detection result when Fig. 7(c) and Fig. 7(d) are compared. The larger the size of neighbourhood L is, the less precious the localization resolution is. However, from the security analysis, we can see that the value of L is tied with the security strength of the scheme. Therefore, a trade off needs to be obtained between localization resolution and security [10].

7.4 VQ attack

To effectively show how the new scheme thwarts the VQ attack, we present the following experiment. Four images are used as the original images as shown in Fig. 8, and they are watermarked by our proposed scheme.

Figure 8. Modified Images of lena, (a) The image lena, (b) The image with the LSB plane and the second LSB plane flipped (c) The image with the LSB plane flipped (d) the image with the second LSB plane flipped

Then an attack is performed by patching a part of the four images to form a fake work. To be specific, each watermarked image is divided into four blocks of size 50×50. Suppose the attack picks up one block from each of the image and makes a patched work, Fig. 8(a) illustrates the patched result, and Fig. 8(b) shows the difference map when the patched work goes past the detector.

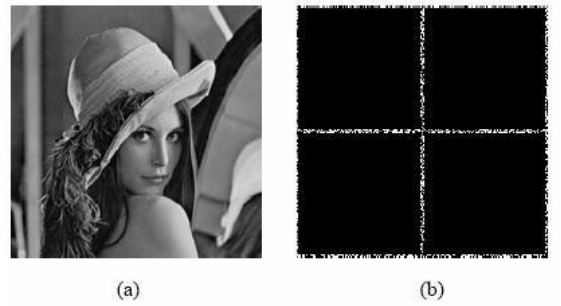

Figure 9. VQ Attack, (a) Patched image, (b) Difference map by detection scheme

8. CONCLUSION:

In this report, we discussed about watermarking techniques, especially the fragile authentication watermarking techniques. First we identified the watermarking system, types, and requirements. Further we provided a small introduction about fragile watermarking, and then we introduced a secure fragile watermarking system exploiting non-deterministic information and contextual information.

Although the field of digital watermarking is young and there are some foreseeable limitations, it has potential and some unique features which other alternatives lack. Also, it will be fruitful if the connection between data hiding and cryptography is further investigated. However, there are still number of questions needed to be answered: whether the combination of data hiding and cryptography can solve the limitations of each other? How strong the relation between them can be? Will there be any new problem generated from the combination?

REFERENCES:

[1] Watson, A.B., Yang, G.Y., Solo mon, J.A. and Villasenor, J. (1997): Visibility of wavelet quantization noise. IEEE Trans. Image Processing, 8(6):1164-1175.

[2] Yu, G.J., Lu, C.S., Liao, Y.M. and Sheu, J.P. (2000): Mean quantization blind watermarking for image authentication. Proc. IEEE Int. Conf. on Image Processing ,Vancouver, Canada, III:706-709.

[3] Zhong, G.J., Cheng, L.Z. and Chen, H.W. (2001): A simple 9/7-tap wavelet filter based on lifting scheme. Proc. of ICIP , 2:249-252.

[4] Alomari, Raja', and Al-Jaber, Ahmed. "A Robust Watermarking Algorithm for Copyright Protection." The 3rd ACS/IEEE Conference on Computer Systems and Applications, Cairo, Eqypt, Jan 2005.

[5] Barreto, P., and Kim H., "Pitfalls In Public Key

[6] Watermarking," Proceedings of Sibgrapi- Brazilian Symposium on Computer Graphics and Image Processing, pp. 241-242, 1999.

[7] Cox J., Miller L., Bloom A., Digital Watermarking, Morgan Kaufmann Publishers, USA, 2002.

[8] Cox J., Miller L., "The First 50 Years of Electronic Watermarking", EURASIP J. of Applied Signal Processing, vol. 2, pp. 126-132, 2002.

[9] Tseng Y., and Pan H., "Secure And Invisible Data Hiding In 2-Color Images," in Proceedings of IEEE INFOCOM 2001, pp. 887-896, 2001.

[10] Xie L., and Arce G., "Joint Wavelet Compression And Authentication Watermarking," Proceedings of the IEEE International Conference on Image Processing, vol. 2, pp. 427-431, 1998.

[11] Yeung C., "An Invisible Watermarking Technique For Image Verification," Proc. of ICIP, pp. 680-683, 1997.

[12] www.ewatrmark.com

[13] www.altavita.com

[14] www.digitalwatermarking.com