

Digital Watermarking of Images for Unique Source Identification

Arsalan Malik

*Department of Electrical, Electronics & Computer Engineering
School of Engineering & Physical Sciences
Heriot-Watt University, Edinburgh, UK
arsalan92@yahoo.com*

Abstract

This paper investigates various watermarking techniques available to identify a unique source of the image. The pros and cons of each technique are discussed. Two of the techniques are investigated in detail along with their implementation. The techniques are tested against pre-defined attacks. The results are discussed and compared.

Keywords: Watermarking, Digital Rights Management, Image Processing, Stagenography

Introduction

In recent years, an increase in inexpensive digital imaging equipment and storage devices together with the proliferation of the Internet has created an environment in which it has become easy to obtain, replicate and distribute digital images. This has prompted many issues with regard to rights management such as identification of the original producer of the image. Digital Watermarking is type of Digital Stagenography that can be used to tackle this problem [1]. A brief review of available watermarking techniques in literature is given first. Then a detailed description of the techniques being implemented is given along with pre-defined testing criterion. Finally results are discussed and compared with other techniques.

Survey of Watermarking Techniques

Many perceptual and non-perceptual watermarking techniques both in spatial and frequency domains have been proposed in literature to solve the unique source identification problem. The most straightforward method in spatial domain is to embed the watermark into the least significant bits (LSB) of the image. The feasibility of 'undetectable' digital water on a standard 512 x 512 intensity image was investigated by Schyndel et al [3]. They showed that a sequence with desirable cross correlation properties can be used to watermark

an image in spatial domain by manipulating least significant bits (LSB). They also showed that resulting image contained an invisible watermark that can be blindly extracted using simple bitwise logic operators suitable for hardware implementations. However this type of watermark was not robust to noise.

Another spatial domain technique was proposed by Walton [4]. He proposed that a checksum obtained by using most significant bits (MSB) of the selected pixels based on a secret key can be embedded in the LSB as watermark object that is invisible. Although he proposed this technique to detect tampering with image data, this technique was not dependant on image data itself. Hence it was possible to swap homologous blocks from two images that are protected with same key, thus allowing an undetectable tampering [5].

A watermark technique dependant on image contents was proposed by Fridrich and Goljan [6]. The technique involved embedding a compressed version of image into LSB's of its pixels. In addition to authentication, it also allows to recover tampered regions of the image. Similar to other LSB's based techniques the watermark produced by this technique was also invisible. However, major drawback of this technique was that embedded watermark was not robust to attacks such as filtering or lossy compression [5].

One of the early watermarking techniques in frequency domain was proposed by Cox et al [7]. They noted that in order to be watermark to be robust, it has to be embedded into perceptually significant areas of the image. Their technique was based on spread-spectrum communications, and involves changing of discrete cosine transform (DCT) coefficients. The watermark was a sequence of random numbers with unit variance and zero mean. The watermark was added to perceptually significant DCT coefficients using one of three equations proposed [7]. Once watermark was added, the watermarked image was obtained by taking

inverse DCT. This technique was non-blind and required original image to be used for extraction of the watermark. The extraction was done using same equation in DCT domain. Their method was robust to image scaling, compression, dithering, cropping and rescanning. Moreover, this method could be used to track multiple watermarked objects.

A watermarking scheme based on the Discrete Wavelet Transform (DWT) was proposed by Xia et al [8]. The watermark, modeled as Gaussian noise, was added to the middle and high frequency bands of the image. Sections of watermarks were extracted by taking the DWT of a potentially marked image and cross-correlated with original watermark. If the cross-correlation was above a threshold, then the watermark was detected. Otherwise, the image was decomposed into finer bands until the entire, extracted watermark was correlated with the entire, original watermark. This technique proved to be more robust than the DCT method [7].

Selected Watermarking Techniques

Two watermarking techniques are selected for the implementation and comparison, the spatial domain technique proposed by Schyndel et al [3] and the frequency domain technique proposed by Cox et al [7]. The techniques are selected because watermark detection process is simple and accurate. Hence this type of watermarking can be used for unique identification of the source. The robustness of these techniques is tested against pre-defined attacks.

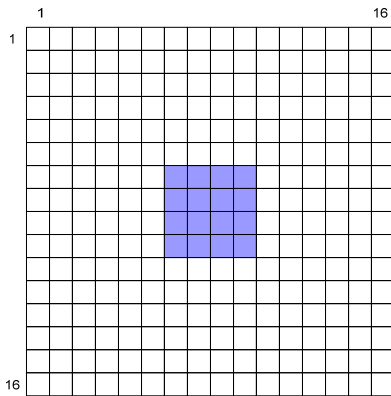


Figure 1: Shaded region represents middle band frequencies of DCT spectrum

The technique proposed by Schyndel et al [3] is implemented using a random Gaussian noise which resembles m-sequence shift register output [3]. The random Gaussian noise is generated using a unique key as random seed, and used as watermark object after segmenting into binary numbers. The

watermark is embedded into LSB of entire image using simple bitwise logical operators. Similarly, using simple bitwise operators, watermark is extracted. The detection process involves cross correlation with extracted watermark and the one generated by the unique key. If the cross correlation value is very high, the presence of watermark based on unique key is identified.

The DCT based technique proposed by Cox et al [7] is implemented using a watermark sequence based on Gaussian noise. The watermark sequence is embedded to the middle band frequencies in DCT spectrum using any of the three equations proposed [7]. The middle band is selected by a square region exactly in the middle of DCT spectrum as illustrated by the figure 1. The size of the square along with the strength of the watermark is the input parameter of this method.

The watermark is extracted using the inverse of the equation used for embedding. The first equation proposed [7] is always reversible, while other two are only reversible when DCT coefficient is non-zero [7]. The images used for testing contained only nonzero DCT coefficients. However it was found that during testing sometimes a cropped image does contain a zero coefficient.

Attacks and Testing System

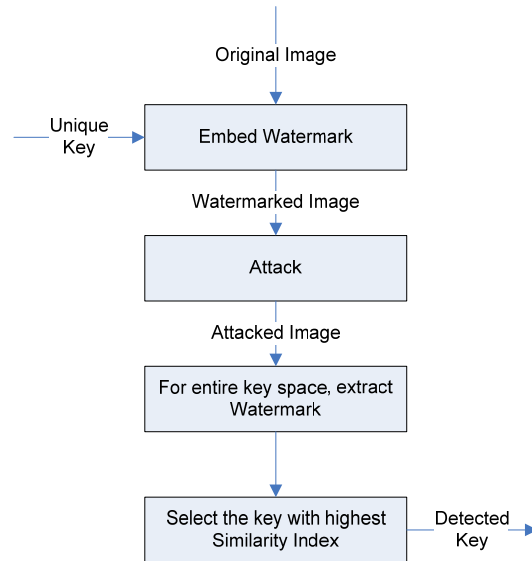


Figure 2: Watermarking Testing Flow Chart

The overall system employed for the testing is shown in figure 2. The original image is watermarked using a unique key as source identifier. The watermarked image is attacked using several

attack methods described below. The attacked image is used to extract watermark. The extracted watermark is compared against watermark generated by entire key space. For the purpose of testing, key space was assumed to be 6-bit, i.e. 64 unique keys. The comparison is represented by the Similarity Index. The key with highest Similarity Index is detected as original source identifier as shown by the key v similarity index graph in figure 3.

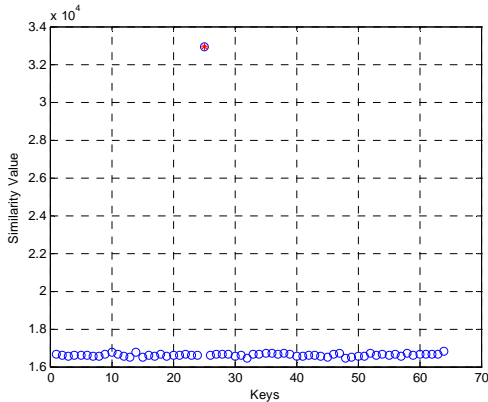


Figure 3: Key v Similarity Index graph

There are many attacks which can be performed to test the robustness of the watermark. Two watermarking techniques were tested against the following attacks:

- Gaussian Noise added in spatial domain. Different noise levels were tested
- JPEG Compression. Different compression levels were tested
- Image Cropping. Five different cropping were used, i.e. top-left, top-right, bottom-left, and bottom-right corner blocks and a central block. Note that the cropped image was resized to the dimensions of the original image in order to extract watermark.
- Image Rotation. Only 90, 180 and 270 degree angles were tested
- Image Scaling. Different scaling factors were tested from 5% to 200%. Similar to image cropping, the scaled image was resized in order to extract watermark.
- Histogram Equalization
- Embedding watermark again using different key

The watermark of size 32 x 32 was used in all of the DCT techniques. The watermark strength (alpha) 0.5 was used with equations 2 and 3, while strength of 2 was used with equation 1 during testing. These parameters were carefully selected after several

experiments in a way that image quality is perceptually maintained while watermark extraction is accurate.

Results and Discussion

The results of the tests are given in Appendix-A. Both spatial and frequency domain techniques were found accurate when no attack is performed on watermarked image. They both detected correct identifier from the extracted watermark.

It can be seen that spatial domain technique is more vulnerable to attacks than frequency domain. Spatial domain watermark was robust only against image scaling or JPEG compression at quality more than 90%. However, it successfully detected both watermarks when a 2nd watermark was embedded using different key.

The DCT based frequency domain technique was generally more robust against many attacks except image scaling in which LSB was more robust. Both techniques were highly vulnerable against image cropping or clipping attack. The DCT technique with equation 1 was more robust against additive Gaussian noise than equations 2 and 3. The equation 3 based embedding and extraction was more robust against Gaussian noise, JPEG Compression, rotation and scaling than equation 2. Both equations 2 and 3 were found robust against histogram equalization attack, while equation 3 was the only technique robust against rotation attack.

Conclusion

A large variety of digital watermarking techniques have been proposed in literature. Most of the techniques use properties of Human Visual System to embed a perceptually invisible watermark. It can be determined based on experiments that a particular watermark is only robust against some of the attacks. Generally frequency domain techniques are more robust than spatial domain techniques. Hence the application of the watermarking should dictate the choice of the technique.

References

- [1] Wenjun Zeng, Heather Yu and Ching-Yung Lin, Multimedia Security Technologies for Digital Rights Management, Elsevier Inc. 2006
- [2] Alan C. Bovik, Handbook of Image and Video Processing, Academic Press, 2000
- [3] R.G.van Schyndel, A.Z.Tirkel and C.F.Osborne, A Digital Watermark, IEEE

- Image Processing, Volume 2. Austin Texas, Nov. 1994, pp 86-90
- [4] S. Walton, "Information authentication for a slippery new age," *Dr. Dobbs Journal*, vol. 20, no. 4, pp. 18–26, 1995.
 - [5] Christian Rey and Jean-Luc Dugelay, A Survey of Watermarking Algorithms for Image Authentication, *EURASIP Journal on Applied Signal Processing* 2002:6, 613–621, 2002 Hindawi Publishing Corporation
 - [6] J. Fridrich and M. Goljan, "Protection of digital images using self embedding," in *Symposium on Content Security and Data Hiding in Digital Media*, New Jersey Institute of Technology, Newark, NJ, USA, May 1999
 - [7] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673-1687, Dec. 1997
 - [8] X. Xia, C. Bonchelet, and G. Arce, "A Multiresolution Watermark for Digital Images," *Proc. IEEE Int. Conf. on Image Processing*, Oct. 1997, vol. I, pp. 548-551.

Appendix A: Detection Results after an Attack

The detection results after different attacks with different parameters are given in the following tables. The parameters of the DCT watermarking technique are fixed for strength and size, while different embedding equations are tested. The result is 'Yes' if correct unique source is identified after the attack, otherwise it's 'No'.

Gaussian Noise Attack

Noise Level	LSB	DCT 1	DCT 2	DCT 3
1	No	Yes	Yes	Yes
10	No	Yes	No	Yes
20	No	Yes	No	Yes
50	No	Yes	No	No
80	No	Yes	No	No
90	No	Yes	No	No
100	No	No	No	No

JPEG Compression Attack

Compression Level	LSB	DCT 1	DCT 2	DCT 3
95	Yes	Yes	Yes	Yes
90	No	Yes	No	Yes
80	No	Yes	No	Yes
75	No	No	No	Yes
50	No	No	No	No
25	No	No	No	No

Cropping or Clipping Attack

Cropped Image	LSB	DCT 1	DCT 2	DCT 3
Top Left 128 x 128	No	No	No	No
Top Right 128 x 128	No	No	No	No
Centre 128 x 128	No	No	No	No
Bottom Left 128 x 128	No	No	No	No
Bottom Right 128 x 128	No	No	No	No

Rotation Attack

Angle	LSB	DCT 1	DCT 2	DCT 3
+90	No	No	No	No
+180	No	No	No	Yes
+270	No	No	No	No

Scaling Attack

Scale (%)	LSB	DCT 1	DCT 2	DCT 3
200	Yes	Yes	Yes	Yes
90	Yes	Yes	No	Yes
80	Yes	Yes	No	Yes
75	Yes	No	No	Yes
50	Yes	No	No	No
20	Yes	No	No	No
10	Yes	No	No	No
5	No	No	No	No

Other Attack

Method	LSB	DCT 1	DCT 2	DCT 3
Histogram Equalization	No	No	Yes	Yes
Watermark	2nd	2nd	1st	2nd

Note that in watermark attack, in case of DCT techniques, although either 1st or 2nd source was identified, both sources (10 and 27) were clearly identifiable on similarity plot as shown below:

