A TRANSPARENT WATERMARKING METHOD FOR COLOR IMAGES

S. Armeni, D. Christodoulakis, I. Kostopoulos, Y. Stamatiou, M. Xenos

Computer Engineering and Informatics Department, University of Patras &
Computer Technology Institute, Greece.

ABSTRACT

This paper presents a secure watermarking method, which provides robustness against a set of attacks. The watermark is embedded into the blue channel of the images in specific locations that guarantee the minimization of the number of bits modified, based on bit similarities that exist between the signature and the cover data. Visual quality measurements and empirical tests have shown the imperceptibility of the embedded watermark. Over 500 tests were performed on numerous images from our image collection, confirming that the watermark can be recovered, even after multiple attacks have been launched in sequence. The probability of false recovery, according to this method is extremely low. Experimental confirmation of the low probability is also provided in this paper.

1. INTRODUCTION

Nowadays the convenience of copying and distributing digital objects (e.g. images, video, audio) combined with the desire of organizations, museums and digital libraries to publish digital images from their collections has led to the necessity of protecting Intellectual Property Rights (IPR). The scientific community has invested serious effort in the past towards the solution of this problem using cryptographic methods such as encryption, digital signatures, etc.

During the past decade, a new approach called *digital watermarking*, based on the idea of information hiding that introduced for the first time in terms of steganography in the fifth century BC [1], seems to give a solution to the problem of IPR. Digital watermarking methods for images take advantage of the fact that the cover data contain an amount of information that does not affect directly the image content as perceived by the observer. Consequently, modifications of this information will not be perceived by the casual observer. Two basic directions have been followed: (a) modification of cover data in the frequency domain (see, for example [5]) and (b) modification of cover data in the spatial domain [2, 4, 6]. Further references are given in Katzenbeisser et al. [1] and in Voyatzis et al. [8].

Most watermarking techniques applied in the spatial domain are based on the properties of the Human Visual System (HVS). Kutter et al. [4] proposed a technique that performs amplitude modulation of the blue component of the images. Moreover, commercial organizations such as Digimarc and Kodak are using watermarking techniques to embed additional information into the blue channel in order to protect digital images.

In this paper, a new watermarking method is presented exploiting both the HVS and the statistical nature of an image, in order to watermark it. The robustness of the method was tested with over 250 images under several attacks.

The method is evaluated using the model of Kutter–Petitcolas [7]. PSNR was used for the measurement of the difference between the original and the watermarked image. This paper is structured as follows: Section 2 gives an overview of the proposed method. Section 3 presents a set of experiments and their results. Finally, Section 4 discusses the results and gives some future directions.

2. WATERMARKING MODEL

This section describes the model of our watermarking method and provides theoretical estimations of its robustness against a set of possible attacks.

The description of the watermarking process is given in 2.1, along with the basic terminology. In 2.2, the watermark recovery process is analyzed and in 2.3 a theoretical model is presented, based on probabilities and statistical properties of digital images. The theoretical robustness of the method against attacks is also mentioned in 2.3.

2.1 WATERMARK PROCEDURE

The watermarking process consists of two steps: (a) a preprocessing step in order to find the best places to embed multiple signatures and (b) the step of embedding the watermark in the image.

A signature s with length $l \in [32, 64]$ bits is selected. The signature s must satisfy the following constrains: (a) $|k_1 - k_0| \leq l/8$, where k_0 is the number of appearances of *0s* in the signature and k_1 is the number of *1s*, and (b) no sequences of more than 5 *0s* or *1s* are allowed in the signature.

The blue channel is composed of 8–bit levels starting from the most significant bit (Level 0) to the least significant bit (Level 7). Each level is considered as an $m \times n$ binary matrix.

A parameter called *similarity* (S_m) is introduced to our method. The basic idea is to embed the additional information into the image, reducing as much as possible the modifications caused to the pixel values. S_m represents the number of bits that are found to be identical, comparing the signature s to bit sequences of the cover data with length l.

A threshold t is calculated using the signature length l and the similarity S_m. The threshold represents the number of bits that have to be flipped whenever a signature is embedded into a sequence of l–bits in the cover data.

The preprocessing step starts from the absolute similarity, in which case the threshold is $t=0$ and involves scanning the 8 levels of the blue channel of the image to find preexisting signatures. Given that the probability of finding a 64–bit

signature in a bit level of a typical image (640x480=307,200 pixels), before the watermarking process, is extremely low (as shown in Section 2.3, this probability is approximately $1.7\ 10^{-14}$**)**, the threshold t is increased by 1 and the entire scanning process is repeated.

This process is repeated until several positions are found in the bit–level in which the signature s is similar to the bit sequences found in these positions at l-t bits. The number of these positions is $(m \times n) - l+1$ where m and n are the image dimensions in pixels.

The embedding process applied on the 8 levels of the blue channel, distributes the signatures unequally, embedding fewer signatures in the most significant levels and more in the least significant levels. The non-uniformity is achieved by choosing different t among the levels. Small t (experimentally measured $t<16$) allows few candidate positions for watermarking and large t (experimentally measured $t>22$) allows many candidate positions for watermarking.

This method succeeds in modifying imperceptibly the original image since for every possible watermark it checks the PSNR between the original image and the image modified by the candidate watermark. Our method backtracks and reduces the candidate t when PSNR measurements are lower than 33dB. PSNR is a typical error measurement, based on the sum of the square differences between corresponding pixels of two images. It should be pointed out that there is a trade off between image quality and watermarking robustness.

2.2 WATERMARK RECOVERY

The watermark recovery process is based on the same reasoning as the embedding process. A key k containing the signature s is used for the comparison with every candidate signature. Every level of the blue channel is scanned using the signature s in order to find similar signatures with threshold t. The process succeeds when the probability of false recovery of the number of signatures found is lower than 10^{-6}**.** This value is adopted as a recovery criterion. The experiments have shown that optimal results are achieved for t within the neighborhood of 10.

In the ideal case of no alteration of the watermarked image, the number of signatures recovered is equal to the inserted signatures, thus false recovery is practically impossible.

2.3 METHOD VALIDATION

Our basic assumption is that the probability of finding preexisting x random bit–sequences in an original image I_o with threshold t is bounded. Based on this fact, we embed a large number of signatures in the watermarked image I_w, excluding, this way, the possibility that this number of signatures preexisted in the original image.

$$p_t = \sum_{i=0}^{t} \frac{\binom{l}{i}}{2^l} \qquad (1)$$

Each level of the blue channel is considered as a random sequence of bit values. A typical image consists of more sequences of *1s* or *0s* in the most significant bit levels and fewer in the less significant. Consequently, when scanning a bit level, the probability of a bit being the same as the previous bit is greater in the most significant bit levels than in the less significant. Although it is not the case for most images, considering the 8 bit–levels as random is advantageous for our estimations, since the calculated probabilities are upper bounds to the actual probabilities (without the assumption of random bit levels) which are even lower.

For a random sequence of l bits, where l is the signature length, and a threshold t, the probability P_t of having at least l-t bits equal to the bits of the signature s is shown in Equation (1).

$$P(x=k) = \binom{n}{x} p_t^{\,x} (1-p_t)^{n-x} \qquad (2)$$

In general, the probability that a number of x signatures is found in a bit–level follows the Binomial distribution [3]. Using the Bernoulli trials, this probability is presented in Equation (2).

3. RESULTS

This section discusses the degree of modifications occurring in an image after embedding a watermark and the results derived when recovering a watermark after a specific attack or a set of attacks in sequence. Statistical data based on over 500 tests performed on numerous images from our image database, altering watermarked images through multiple attacks, as well as detailed examples of few typical images are also presented.

3.1 ORIGINAL IMAGE PRESERVATION

A watermark, as discussed in Section 2, modifies the original image. The main goal of any watermarking method is to minimize the modifications caused to the original image, so as not to be detected by the observer. In order to test the way in which this method affects the original image [7], we used three types of tests: (a) measurements of Peak Signal to Noise Ratio (PSNR) between the original and the watermarked image, (b) measurements of similarities in each one of the 24 m x n bit matrices and (c) tests using the HVS.

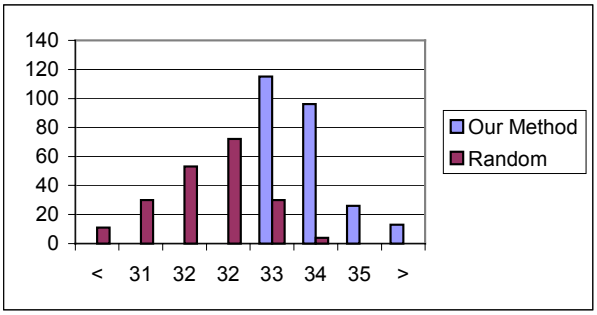

Figure 1. PSNR measurements

PSNR

Since our watermarking method determines the number of signatures to be embedded into an image based on PSNR measurements, PSNR is expected to range within the predetermined acceptable limits (i.e. PSNR > 33dB). Figure 1 shows the distribution of images for different PSNR measurements. As shown in Figure 1, the PSNR of most of the images modified by our watermark method is near 33dB. Moreover, the PSNR of all images modified by our method is above 33dB, as required by our method. As an indication of how

PSNR measurements are affected by image alterations (attacks), Jpeg Medium compression results in a typical PSNR of approximately 27dB and a 0.5° Rotation results in a typical PSNR measurement of approximately 20dB.

The application of the proposed method on a specific image and a specific number of signatures *n(i)*, where $i \in [0..7]$ (blue bit levels), succeeds in modifying the minimum number of bits due to the use of similarities. It is, therefore, expected that the PSNR measured for our method —for a specific *n(i)* number of signatures embedded— is higher or at least equal to the PSNR measured for any other method embedding the same *n(i)* number of signatures.

Figure 1 presents the results of the PSNR measurements for a set of images modified by the insertion of the same number of signatures, using our method and random insertion. Table 1 shows the actual figures from PSNR measurements between the original and the modified images for 5 images.

Images	*Dimensions*	*∑ n(i)*	*Our Method (dB)*	*Random (dB)*
Acropol	681 x 411	1109	35,404	32,322
Baboon	512 x 512	1404	33,753	30,402
F-15	732 x 500	2415	33,612	31,104
Kid	487 x 703	1852	34,877	31,601
Flower	640 x 480	1793	34,088	30,994

Table 1. Method comparison for 5 images

Bit level Similarities

Another technique to measure the degree of modification caused to an image by any watermarking method is to measure the similarities for each one of the aforementioned *m x n* matrices. Since our method modifies only the blue channel of the RGB (namely 8 of the *m x n* matrices), after the watermark insertion, it is evident that the remaining 16 matrices are 100% similar to the original.

Modifications are only measured for the blue channel where they applied. Our method affects mostly the least significant bits and barely modifies the most significant bits. The 2nd column of Table 2 shows the average percentage of the similarities measured in every bit level of the blue channel between the original and the watermarked images, for a total of 250 images watermarked using our method.

Bit Level	*Our Method*	*Random Embedding*
0	99.93%	99.91%
1	99.68%	99.27%
2	99.49%	99.03%
3	99.12%	98.48%
4	98.70%	98.06%
5	98.54%	97.82%
6	98.31%	97.51%
7	98.02%	97.12%

Table 2. Average percentage of similarities

The 3rd column of Table 2 shows the average percentage of the similarities measured in every bit level of the blue channel, for the same total of images, randomly watermarked by inserting the same number of signatures for each level. By using our method we modify the minimum number of bits for a specific number of signatures. Consequently, our similarity measurements are higher than those of the random method.

Human Visual System

Since one of the main goals of watermarking is to produce a watermarked image that is perceived by the HVS as identical to the original image, all images produced after the watermark insertion were tested by human observers. Since the proposed method modifies only a small number of bits in the blue channel, these changes remain unnoticed by the observer, as the human eye is less sensitive in modifications to the blue channel [4]. The watermarked images were tested by human observers and no watermark was perceived.

Figure 2 shows a typical image and the watermarked image after the insertion of the *n(i)= [9, 58, 204, 305, 410, 834, 919, 1556]* number of signatures (a total of 4295 signatures), where *i* is the bit level (from most significant to least significant). The signature used in this example is "Michalis" and the PSNR measurements between the original and the watermarked image indicate PSNR= 35.188dB.

Figure 2. Original Image – Watermarked image

Figure 3 shows another example of a typical image and the watermarked image after the insertion of the *n(i)= [8, 25, 98, 313, 655, 816, 1427, 1472]* number of signatures (a total of 4814 signatures). The signature used in this experiment is "Giannis0" and the PSNR measurements for this case indicate PSNR= 35.040dB.

Figure 3. Original Image– Watermarked image

3.2 ATTACKS AND RECOVERY

The robustness of the proposed method was tested experimentally by the conduction of more than 500 tests. The overall results of the experiments are presented in this section. In the original image I_o we embed watermarks producing a watermarked image I_w; subsequently this image is attacked producing image I_a. The attacks intend to remove the signatures embedded in I_w. The resulting image I_a may still contain a number of signatures. If this number is significantly higher compared to the probability of false recovery (i.e. the case in which this number of signatures incidentally preexisted in image I_a) then we consider that the watermark can be recovered.

The types of attacks on images I_w that we used are: rotation (up to 1°), rescale (up to 5%), blur, jpeg compression (low, medium and high) and cropping (part of image, or one horizontal and one vertical line). Furthermore, combinations of attacks were applied during our experiments (up to 4 subsequent attacks).

In order to consider a watermark as recovered, we found that the

best threshold, for the typical image size and for signature length *l=64,* is *t=10*. For a typical image size 307,200 pixels (640x480), the upper bound of the probability of incidentally finding 5 signatures (in all levels) is 1.8×10^{-13} (using Equation 2). Using the assumption discussed in Section 2.2, we consider every watermark recovered, with false recovery probability lower than 10^{-6}, as successful recovery. Tables 3 and 4 present the number of the images that were validated as watermarked after a single attack, while Table 5 presents the number of the images that were validated after a combination of multiple attacks in sequence.

Images	*Rotate*	*Rescale*	*Crop P*	*Crop L*
Recovered	100%	100%	100%	100%
Not Recovered	0%	0%	0%	0%

Table 3. Rotation, rescaling, cropping part and cropping lines

Images	*Blur*	*Jpeg L*	*Jpeg M*	*Jpeg H*
Recovered	100%	95%	99%	100%
Not Recovered	0%	5%	1%	0%

Table 4. Blur and Jpeg compression (low, medium and high)

According to the results presented in Tables 3 and 4, it is obvious that our method is 100% safe against image rotation, image resizing and blur. With regard to Jpeg compression, our method produces fine results even on low jpeg, but we cannot guarantee that all images will be validated after a low jpeg attack.

Images	*JM+Ro*	*Cr+Res*	*Cr+B*	*4-attack*
Recovered	100%	100%	100%	93%
Not Recovered	0%	0%	0%	7%

Table 5. Multiple attacks

The proposed method resists perfectly against two attacks in sequence (Table 5), such as Jpeg Medium and Rotation (2nd column), or cropping and rescaling (3rd column) and cropping and blur (4th column). It also produces good recovery rates against four attacks in sequence such as Jpeg high, followed by Rotation, Rescaling and Blur (4th column).

Image	*Dims*	*Rot*	*P*	*Res*	*P*
Kariat	672x405	*631*	10^{-3115}	*196*	10^{-869}
Maize	640x480	*319*	10^{-1464}	*59*	10^{-229}
Arch	400x594	*569*	10^{-2817}	*142*	10^{-619}
Pills	748x518	*372*	10^{-1693}	*58*	10^{-218}
Pepp	512x512	*142*	10^{-0612}	*128*	10^{-546}

Image	*Crop*	*P*	*All*	*P*
Kariat	*807*	10^{-4069}	*15*	10^{-51}
Maize	*611*	10^{-2975}	*22*	10^{-77}
Arch	*703*	10^{-3544}	*18*	10^{-64}
Pills	*572*	10^{-2710}	*14*	10^{-44}
Pepp	*299*	10^{-1384}	*10*	10^{-32}

Table 6. Recovery examples.

3.3 DETAILED EXAMPLES

Representative results from all types of attacks for 5 typical images are presented. In Table 6 we show the name of the images, their dimensions, the signatures recovered and the probability of a false recovery for every type of attack. This table depicts the examples and the probability of a false recovery for Rotation, Rescaling, Cropping and four combined attacks.

4. CONCLUSION AND FUTURE WORK

It is obvious by the entire set of examples that the proposed method modifies imperceptibly the original image. It is worth mentioning that PSNR measurements taken following the insertion of our watermark, were higher than those taken following all types of attacks. The method achieved very good results against most types of attacks, even multiple attacks in sequence. Various scan methods [2] can been used so as to increase the complexity of our key. In this way, the watermarked image is protected from malicious attacks by signature aware users, since both the signature and the scan pattern are parts of the key.

In the future, we intend to exploit image properties in order to embed signatures in a more sophisticated way and make publicly available our watermarking tools.

5. REFERENCES

[1] Stefan Katzenbeisser, Fabien A.P. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, 2000.

[2] S. Armeni, D. Christodoulakis, I. Kostopoulos, Y. Stamatiou, "Multiple Signature Embedding Using Consecutive Pixel Distribution", Workshop on Watermarking and Copyright Enforcement, October 22-23,1999, Paris, France.

[3] Edward R. Dougherty, Random Processes for Images and Signal Processing, SPIE/IEEE Series on Imaging Science & Engineering, 1999.

[4] Kutter, M., F. Jordan, and F. Bossen, "Digital Signature of Color Images Using Amplitude Modulation", in Proceedings of the SPIE 3022, Storage and Retrieval for Image and Video Databases V, 1997, pp. 518-526.

[5] I.J. Cox, J. Kilian, T. Leighton and T. Shamoon, "A Secure, Robust Watermark for Multimedia," Workshop on Information Hiding, Newton Institute, Univ. of Cambridge, May 1996.

[6] N. Nikolaidis and I. Pitas, Copyright protection of images using robust digital signatures, IEEE International Conference on Acoustics, Speech a nd Signal Processing (ICASSP-96), vol. 4, pp. 2168-2171, May 1996.

[7] M. Kutter, F.A.P. Petitcolas, "A fair Benchmark for Image Watermarking Systems", *Proceedings of SPIE: Security and Watermarking of Multimedia Contents,* Volume 3657, pp. 226-239, San Jose, California, January, 1999.

[8] G. Voyatzis, N. Nikolaidis and I. Pitas, "Digital Watermarking: An Overview", *Proc. of EUSIPCO'98*, September 8-11, Rhodes, Greece, 1998.