Robust Undetectable Interference Watermarks*

Ryszard Grząślewicz[1], Jarosław Kutyłowski[2], Mirosław Kutyłowski[1], and Wojciech Pietkiewicz[1]

[1] Institute of Mathematics, Wrocław University of Technology,
Ryszard.Grzaslewicz@pwr.wroc.pl
Miroslaw.Kutylowski@pwr.wroc.pl
pietkiew@im.pwr.wroc.pl
[2] International Graduate School of Dynamic Intelligent Systems, University of Paderborn
jarekk@upb.de

Abstract. We propose a digital watermarking method for gray-tone images in which each watermark consists of a collection of single points and each point is encoded in the spatial domain of the whole image. The method is somewhat related to physical digital holograms and interference images. Reconstruction of such watermarks is based on a similar principle as the reconstruction of physical holograms.
While encoding a watermark in the spatial domain one of the major problems is to avoid a textured appearance due to the encoding scheme. We avoid a recognizable pattern by creating pseudorandom keyed watermarks, which resemble random noise.
The method proposed yields robust watermarks that are resistant against many attacks which preserve the distance between points (filtering, rotation, JPEG compressing). The watermarking scheme provides means for detection and reversal of scaling transformations, thus making the watermark resistant to this attack.
The original picture is not required for reconstruction. The watermark is quite hard to detect, which prohibits easy violation of watermark protection.
Our method guarantees exact reconstruction provided that the watermark image consists of a limited number of white pixels on a black background.

1 Introduction

Digital watermarks are used for protecting digital images and contain information necessary for exercising the rights of the owner. They need to be robust in the sense that digital image operations occurring during normal use do not destroy the watermark.

Two types of watermarks can be distinguished – watermarks that can be easily detected by everyone and the information they carry can be extracted (e.g.

* the third author was partially supported in years 2003-2005 by KBN, project 0 T00A 003 23

the name of the copyright owner) and such that are undetectable. Our scheme constitutes an undetectable watermark, which cannot be easily recognized in an image without the proper key. In such a scenario it is not crucial to encode a lot of information in the watermark. The most important issue is whether an image carries a watermark created with a given key.

There has been a lot of research on watermark schemes in the last decade (see for instance [1]). Despite enormous efforts there is no ultimate solution of this problem. To the best of our knowledge, every scheme proposed so far has some weakness that can be exploited by an adversary attacking the watermark.

Geometric distortions in the image are particularly dangerous. A standard set of operations, called RST, consist of rotations, scaling and translations (shifting the pixels some number of positions). However, there are further geometric attacks such as cropping the image or nonlinear transformations. Each of these techniques destroys watermarks for which the pixel positions cannot be changed.

A general idea to resist RST attacks is to find image characteristics invariant to the RST operations. Then we may encode a watermark into these characteristics – obviously this requires some freedom to manipulate the image without influencing it so that the changes become detectable by a human eye. A solution of this kind based on Fourier-Mellin transform is proposed in [3]. The method requires the original image and implementation problems have been claimed [2]. Fourier-Mellin transform is used again in [4], a characteristic vector of a digital image is defined so that it remains unchanged during RST operations. Again, implementation problems due to inaccuracies during embedding and detection computations have been reported. A similar approach [2] based on Radon transform provides a scheme that is claimed to be practical. Nevertheless, it has a weak point: the characteristic vector changes substantially if the watermarked image is not "homogeneous" and we crop it.

Digital watermarks proposed in [5] are constructed in a way that mimics physical holograms. They have a remarkable property that the watermark image can be reconstructed from each reasonably large block of pixels. However, the scheme is based on Fourier transform, with the watermark reconstructed in the frequency domain. Our experiments have shown that the watermark can be easily removed by blurring the watermarked image slightly.

Features of the New Scheme. We propose a scheme that is based on a very simple idea, but surprisingly yields very good results. Let us point to its major features:

- Watermark retrieval does not require the original image.
- Watermarks are resistant to translations, cropping and symmetries.
- Watermarks are resistant to scaling after such an operation has been detected. The watermarking scheme provides means for detecting scaling operations.
- Watermarks are resistant to standard image processing operations, such as changing the contrast, changing the brightness, blurring, adding random noise and JPEG compression.

– Watermarks can be retrieved from image parts, without even knowing the original location of the part within the image.
– The robustness of the scheme has been tested against the Stir Mark benchmark suite and satisfactory results have been obtained.
– A watermark is composed of single white pixels on a black background with white point positions encoding the watermark information.
– Once a watermark is known, it can be easily removed from the image by superposing the same watermark but with the negative sign.
– Watermarks are created with a secret key. The key is necessary for detecting, restoring and removing the watermark image.
– Retrieving a watermark is computationally intensive. In most practical cases this prohibits examining the images for watermarks without knowledge of the key and without access to enormous computing resources.

The paper is organized as follows: in Section 2 we present the physical motivation for our scheme and give a theoretical background for the watermarking encoding and decoding algorithms, which are described in Section 3. The results of our experimental evaluation are presented in Section 4.

2 Interference Images

Let us describe the basic idea of our approach. Assume that we have to encode a watermark image which can be represented as a matrix $(a_{i,j})_{i,j\leq n}$, where $a_{i,j} = 1$ if pixel (i, j) is white, and $a_{i,j} = 0$ otherwise. We assume that the white pixels are used to encode the information, so the main issue is how to represent a single white pixel. For this purpose let us recall the phenomenon of interference images.

Physical Motivation. Let us recall the Young experiment: coherent monochromatic light is passing through two small slits H_1, H_2 lying close to each other on plane P. The light is diffracted when passing through the holes and it goes into all possible directions. Consider a single point A on plane Q parallel to P. The distances between A and H_1 and between A and H_2 are slightly different. Let λ be the length of the light wave. The waves passing through H_1 and H_2 are in the same phase, but they have different distances to reach A. So when they reach A, they are shifted in phase – the shift corresponds to the additional distance one of these waves has to go. If the difference equals $\lambda \cdot i$, for $i \in \mathbb{N}$, then the waves sum up. But if the difference is $\lambda \cdot i + \lambda/2$, for $i \in \mathbb{N}$, then the waves cancel themselves out. It follows that on screen Q we get an interference pattern consisting of dark and bright lines. This effect is called two-source light interference. If there is more than one pair of slits in P, the values corresponding to different pairs sum up.

General Framework. First we pick a certain function F (later we discuss the necessary properties of F). Let $F_{i,j}(x, y) = F(x - i, y - j)$. Then we represent a

watermark image $(w_{i,j})$ by the sum

$$\sum_{i,j} w_{i,j} \cdot F_{i,j} \ .$$

That is, in the resulting image the pixel with coordinates (a, b) has the value

$$h_{a,b} = \sum_{i,j} w_{i,j} \cdot F(a-i, b-j) \ . \tag{1}$$

Reconstruction of the watermark image from $H = (h_{i,j})$ will be performed by computing for every point of the watermark

$$w_{a,b} := \sum_{i,j} |h_{i,j} - F(a-i, b-j)| \ . \tag{2}$$

Let us explain informally the motivation for such a reconstruction rule. Let us assume that the watermark image consists of points (e_1, f_1), (e_2, f_2), (e_3, f_3). Then

$$h_{i,j} = F(i-e_1, j-f_1) + F(i-e_2, j-f_2) + F(i-e_3, j-f_3) \ . \tag{3}$$

So for the point (e_1, f_1) the reconstruction rule yields

$$\sum_{i,j} |F(i-e_2, j-f_2) + F(i-e_3, j-f_3)| \ . \tag{4}$$

On the other hand, if we perform reconstruction at a point (u, v) that is different from (e_1, f_1), (e_2, f_2), (e_3, f_3), then the negative term in Eq. 2 does not cancel any of $F(i-e_k, j-f_k)$ and we get the expression:

$$\sum_{i,j} |F(i-e_1, j-f_1) + F(i-e_2, j-f_2) + F(i-e_3, j-f_3) - F(i-u, j-v)| \tag{5}$$

For summation over all integer i, j and after removing the absolute values we would get exactly the same result for Eq. (4) and Eq. (5). However, it may happen that $F(i-u, j-v) > F(i-e_1, j-f_1) + F(i-e_2, j-f_2) + F(i-e_3, j-f_3)$. Then in Eq. (5) we get an extra positive value that does not cancel out due to the use of absolute values. If we assume that $F(i-e_l, j-f_l)$, $l = 1, 2, \ldots$ and $F(i-u, j-v)$ are independent random variables uniformly distributed over $[0, M]$, then one can prove that the difference between expected values of the corresponding additive factors in by Eq. (5) and Eq. (4) is of order $M/k!$, where k is the number of white points in the watermark image. Certainly, the assumption about stochastic independence is not valid mathematically, but for our choice of function F similar phenomena can be observed. An important point is that for large k we cannot hope for a good reconstruction due to the factor $M/k!$.

An alternative way of computing reconstruction values based on orthogonal functions could be

$$w_{a,b} := \sum_{i,j} h_{i,j} \cdot F(a-i, b-j) \ .$$

However, our experiments have shown that for the employed functions F and images occurring in practice, reconstruction with equality (2) yields significantly better results.

In order to get reasonable resistance against attacks on watermarks we need some properties of F. The first point is that an adversary could crop the image, for instance take only one quarter of it, that is, put value 0 at all points except the chosen quarter. The crucial issue then is how much of the value $w_{a,b}$ is contained in the chosen quarter. We need a property that each rectangular block of points B contributes a value into $w_{a,b}$ which is roughly proportional to the area of B, provided that B is not too small. Informally speaking, the "energy" of a white watermark pixel should be dispersed quite evenly on the whole transformed image. The last property would automatically yield resistance against local editions in image H.

In order to get resistance against operations like replacing a pixel value through the average in its closest neighborhood, we need the property that the image H does not consist of waves of high frequency only. This objective contradicts the previous objective, where high frequency waves are preferred. So we need to find a proper compromise.

Another point is that the objects that are likely to appear on the images to be watermarked, should be almost orthogonal to functions $F_{i,j}$ (in the sense that such objects do not contribute many small values in the sum from Eq. (2)). For instance lines, treated as functions with value 1 on the line points and zero elsewhere, should be orthogonal to functions $F_{i,j}$. This excludes the functions such as $F(x, y) = \cos(\max(x, y))$, since vertical and horizontal lines of the image would coincide with constant values of F. Ideally, if we take a curve at which the value of F is constant, then the pixel values on the image to be watermarked should form a quasi-random multiset of values. Of course, this should be true for images that can occur in practice.

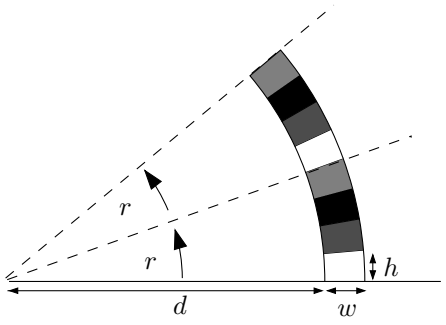

Fig. 1. Interference image values with function F_K

Artificial Pseudorandom Interference Images. We consider "interference images" created by changing the functions – from those describing physical reality to more handy ones suitable for watermarking. First we consider an interference image for a pair of holes located at point $(0,0)$ with interference value at point (x,y) described by a function

$$F(x,y) = \cos\left(\sqrt{x^2+y^2+z^2}\right)$$

where z is a parameter which can be thought of as the distance between the planes P and Q introduced in the description of the physical motivation. This image is not suitable for watermarking, since it contains visible circles. Thus, we introduce pseudorandomness to the interference images. Let H be a secure hash function and K be a secret key used for watermarking. We define

$$F_K(x,y) = P_K\left(\angle(x,y) \bmod r, \sqrt{x^2+y^2+z^2}\right) \cdot \left(\frac{\sqrt{x^2+y^2}}{f}+1\right)^{-1}$$

$$\text{where} \quad \angle(x,y) = 90\frac{2}{\pi}\arctan\left(\frac{x}{y}\right), \text{ and } P_K(a,b) = H(K,a,b)\ .$$

Thus the value of $F_K(x,y)$ depends both on the distance of point (x,y) from $(0,0)$ and on the angle $\angle(x,y)$ of the vector (x,y). Basing on the security of the hash function, it is impossible to predict the value of $F_K(x,y)$ without knowledge of key K.

Fig. 1 depicts a circle of points in distance d from point $(0,0)$. As the values of F_K depend on the residue of the angle $\angle(x,y)$ modulo r, the same pattern is repeated every r degrees. Thus when the interference image is rotated by a multiple of r degrees, reconstructing the watermark is possible. In order to deal with rotations by non-multiples of r, we have to try to reconstruct the watermark for r different rotations. (Note that it is not necessary to know around which point the image has been rotated.)

The "width" w of the circle (see Fig. 1) is determined by arithmetic precision of the computation of the distance from point $(0,0)$. The height h of each sector of the circle is determined by the arithmetic precision of the angle computation. This can be assumed to be one degree.

The term $\left(\frac{\sqrt{x^2+y^2}}{f}+1\right)^{-1}$ influences intensity of $F_K(x,y)$ based on the distance between $(0,0)$ and (x,y) and a constant parameter f.

In order to obtain integer values as arguments of P_K (and therefore of H), both input parameters are divided by, respectively, w and h and truncated to integer values.

3 Watermarking Algorithm

Creating a Watermark. The watermark is created by inserting the vertices of $z/3$ equilateral triangles with edge length T equally distributed on the watermark image. Thus the watermark image consists of z white pixels.

Inserting a Watermark into an Image. Let $h_{i,j}$ denote the value of pixel (i, j) of the interference image. In order to construct this image and embed it into the cover image we execute the following steps:

1. We start with a black interference image of size $n \times n$, which is equal to the size of the watermark.
2. For each white point (a, b) of the watermark and every point (i, j) within the interference image the value of $F(a-i, b-j)$ is computed and added to the current value of $h_{i,j}$.
3. The intensity of the interference image is normalized, so that the minimum intensity of a pixel equals $-\frac{\alpha}{2}\mathcal{M}$ and the maximum equals $+\frac{\alpha}{2}\mathcal{M}$, where $\mathcal{M}$ is the maximum intensity of the cover image. Values in between are scaled linearly. Parameter α determines the strength of the embedding of the interference image within the cover image. A bigger value of α results in a stronger watermark but lowers image quality.
4. The interference image and the cover image are superposed by adding the intensities of the pixels. The result represents the watermarked image.
5. The normalization procedure is applied to the watermarked image with the bounds 0 and $\mathcal{M}$.

In Point 2, we compute interference values as given by Eq. (1). Time complexity of inserting a watermark is proportional to n^2 times the number of white pixels.

Watermark Reconstruction. Our goal is to determine whether an image has been watermarked with a given key K. The reconstruction process consists of two phases. First, the scale factor of the image is detected and the image is re-scaled properly. In the second phase, the actual reconstruction takes place.

1. The watermarked image is rotated by $0, 1, 2, \ldots, r$ degrees (recall that r is the angular periodicity of F_K). The next algorithm steps are executed for each of these rotations until a watermark is found.
2. If the watermarked image has been scaled so that the distance between points is changed, this scaling must be detected and reversed. So, prior to the actual reconstruction the detection of the performed scaling is done in the following way:
   (a) An image part is chosen, such that there has been at least one watermark point encoded in this part. If the z points have been distributed uniformly over the whole watermark, then every image part of appropriate size should contain such a point.
   (b) A reconstruction for this part is computed for every reasonable scale factor. This reconstruction is performed as in step 3.
   (c) The value of each point is considered as a function of the scale factor. For each point the maximum peak is detected. The peaks among all points are ordered according to their size. The scale factor of the largest peak is chosen as the proper scaling factor and used for the reconstruction.

(a) original (b) watermarked

Fig. 2. Cover image prior to watermarking and afterwards

(d) The image is scaled by the discovered scale factor.

3. The actual reconstruction process is as follows:
   (a) For each position (a, b) of the watermark we compute

$$w_{a,b} := \sum_{i,j} |o_{i,j} - F_K(a-i, b-j)|, \tag{6}$$

   where $o_{i,j}$ is the pixel value of the watermarked image at point (i, j).
   (b) The difference between $o_{i,j}$ and $F_K(a-i, b-j)$ is near to zero for pixels which have been white in the watermark (since $o_{i,j}$ is closer to $F_K(a-i, b-j)$ in this case). Thus we invert the reconstructed watermark, so that the brightest pixels in the reconstructed watermark stand for those which have been white in the watermark.
   (c) We find the δz brightest pixels. The δ factor allows for some error-correction, if due to noise there are invalid bright pixels. Additionally, if more than one bright pixel is found in a small area, only one representative of them is chosen to be used in the computation. Such a situation can occur if due to scaling the intensity of a watermark pixel goes over to its neighboring pixels.
   (d) We look for equilateral triangles with vertices among the brightest pixels. If enough triangles with an edge length of T are found, the watermark is detected.

The presented algorithms work with low-precision arithmetic implemented using integer numbers, with a resolution of 10^{-2}.

4 Experimental results

We have implemented our scheme and checked the results to examine the practical relevance of the algorithm. We have looked for appropriate values of pa-

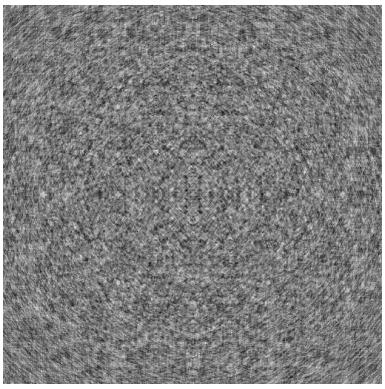

Fig. 3. Interference image for watermark

rameters giving good reconstruction results. The values has been chosen to be $d = 2000$, $k = 6$, $r = 10$, $\delta = 1.3$, $T = 100$, $z = 8$ and $f = 25\sqrt{2n^2}$. Embedding of watermarks into the cover image is performed with the factor $\alpha = 0.2$, which leads to a PSNR value larger than 32dB for all tested images. All images used in the test process had an original size of 1024×1024 pixels and were encoded using 8-bit grayscale values.

The watermarking process takes about one second on a modern-class PC whereas the reconstruction can last up to a few hours on a comparable computer.

Fig. 2(a) presents one of the cover images from the StirMark suite, whereas Fig. 2(b) is the cover image with the watermark embedded. The interference image is shown in Fig. 3.

Watermark robustness. We have examined the resistance of the watermark against several common attacks by invoking the Stir Mark Benchmark 3.1 [6] suite. We summarize the results

- Cropping – All tests (from 1% to 75%) passed. (9 of 9 successful)
- Remove rows/columns – Tests 1/1, 1/5 and 5/1 passed. (3 of 5 successful)
- Flip – Test passed. (1 of 1 successful)
- Scaling – All tests (from 0.5 to 2.0) passed. (6 of 6 successful)
- Change aspect ratio – All tests passed. (8 of 8 successful)
- Rotation with cropping – Tests with rotation smaller than 30% passed. (13 of 16 successful)
- Rotation with cropping and scaling – Tests with rotation smaller than 30% passed. (13 of 16 successful)
- Shearing – Tests with shearing of 1% on one axis passed. (2 of 6 successful)
- General linear transformation – No tests passed. (0 of 3 successful)
- StirMark – Test not passed. (0 of 1 successful)
- Gaussian filtering – Test passed. (1 of 1 successful)

– Sharpening – Test passed. (1 of 1 successful)
– Median filtering – All tests passed. (3 of 3 successful)
– LRAttack – Test passed. (1 of 1 successful)
– JPEG compression – All tests passed. (12 of 12 successful)

In overall 77 of 89 tests have been successful. In the present implementation there are no countermeasures against linear transformations applied, however a similar approach as for scaling may be used.

The StirMark test consists among others of distortions like shearing, stretching and rotating and nonlinear transformations like bending and random displacement. Its success against our scheme is mainly due to nonlinear transformations. However, there are techniques that help to trace which nonlinear transformations have been used. They might enable recovery of the image before transformations and in this way – recovery of the watermark. These countermeasures have not been included yet in the implementation tested.

5 Conclusion

The presented watermarking method is robust against several common attacks and provides an innovative technology for embedding undetectable watermarks in the spatial domain of images. A nice feature of the scheme is that the watermarks are reconstructed exactly – so it enables direct encoding of digital information.

Its asymmetric behavior – easy insertion and time-costly reconstruction even with a known key – might be a useful tool for copyright protection in the Web. Massive and automatic coping of digital images would require removing the watermarks – which is computationally intensive.

The main concern of the scheme remain nonlinear transformations. Future work should encompass the application of detection schemes for these transformations.

References

1. Hartung, F., Kutter, M.: Multimedia Watermarking Technique. Proc. IEEE 87, 1999, 1079-1107
2. Hyung-Shin Kim, Yunju Baek, Heung-Kyu Lee, Young-Ho Suh: Robust Image Watermark Using Radon Transform and Bispectrum Invariants. Information Hiding '2002, Lecture Notes in Computer Science 2578, Springer-Verlag, 145-159
3. O'Ruanaidh, J.J.K., Pun, K.: Rotation, Scale and Translation Invariant Spread Spectrum Digital Image Watermarking. Signal Processing 66, 1998, 303–317
4. Lin, C.Y., Wu, M., Bloom, J.A., Cox, I.J., Miller, M.L., Lui, Y.M.: Rotation, Scale, and Translation Resilient Watermarking for Images. IEEE Trans. Image Processing 10, 2001, 767–782
5. Takai, N., Mifune, Y.: Digital Watermarking by a Holographic Technique. Applied Optics 41(5), 2002, 865–873
6. Petitcolas, F., Anderson, R., Kuhn, M.: Attacks on Copyright Marking Systems. Information Hiding '1998. Lecture Notes in Computer Science 1525, Springer-Verlag, 219-239