**CERIAS Tech Report 2001-67**

# A WATERMARKING TECHNIQUE FOR DIGITAL IMAGERY: FURTHER STUDIES

by Raymond B. Wolfgang and Edward J. Delp

Center for Education and Research in
Information Assurance and Security,
Purdue University, West Lafayette, IN 47907-2086

# A WATERMARK FOR DIGITAL IMAGES

*Raymond B. Wolfgang and Edward J. Delp*

Computer Vision and Image Processing Laboratory
Purdue Multimedia Testbed Laboratory
School of Electrical Engineering
Purdue University
West Lafayette, Indiana, 47907-1285
USA

**ABSTRACT**

The growth of networked multimedia systems has magnified the need for image copyright protection. One approach used to address this problem is to add an invisible structure to an image that can be used to seal or mark it. These structures are known as *digital watermarks*. In this paper we describe two techniques for the invisible marking of images. We analyze the robustness of the watermarks with respect to linear and nonlinear filtering, and JPEG compression. The results show that our watermarks detect all but the most minute changes to the image.

## 1. INTRODUCTION

The recent growth of networked multimedia systems has caused problems relative to the protection of intellectual property rights. This is particularly true for image and video data. The types of protection systems involve the use of both encryption and authentication techniques. In this paper we describe a form of authentication known as a *watermark*. These *digital watermarks* also offer forgery detection. Several watermarking techniques have been proposed. One uses a checksum on the image data which is embedded in the least significant bits of certain pixels [1]. Others add a maximal length linear shift register sequence to the pixel data and identify the watermark by computing the spatial crosscorrelation function of the sequence and the watermarked image [2]. Watermarks can be image dependent, using independent visual channels [3], or be generated by modulating JPEG coefficients [4]. These watermarks are designed to be invisible, or to blend in with natural camera or scanner noise. Visible watermarks also exist; IBM has developed a proprietary visible watermark to protect images that are part of the digital Vatican library project [5]. In this paper we present a watermark which is a two-dimensional extension of [2]. We describe a forgery detection scheme with a new approach to robustness. The watermark's robustness to mean and median filtering is investigated. We then introduce a second watermark that is robust relative to JPEG compression.

This work was partially supported by a grant from the AT&T foundation. Address all correspondence to E.J. Delp, ace@ecn.purdue.edu, http://dynamo.ecn.purdue.edu/~ace, or +1 317 494 1740.

## 2. ADDITION OF M-SEQUENCES

A linear feedback shift register with n stages can form pseudo-random binary sequences with periods as large as $2^n - 1$; m-sequences achieve this maximum period and have excellent randomness and autocorrelation properties [6]. To generate the watermark, a binary sequence is mapped from {0,1} to {-1,1}, arranged into a suitable block, and then added to the image pixel values. Advantages of this type of watermark include:

1. If an authorized user knows the watermark, the exact original image can be obtained. The LSB plane is not irrecoverably altered as it is with a checksum technique.
2. An attacker can only swap pixels with the same m-sequence bit without affecting the correlation properties. This requires knowledge of the private embedded sequence to successfully forge any reasonable area of the image.
3. Multiple watermarks can overlap each other and will not change the average value (brightness) of the image. Successive watermarks would treat the previously watermarked image as the original. This would also trace an image's chain of custody or audit history.

Some disadvantages include:

1. If the watermark covers the entire image, an attacker must merely guess if a given pixel has increased or decreased by one gray level to identify a particular bit in the watermark.
2. An attacker could compute an entire watermark block if *2n* consecutive bits are known. More secure non-

linear codes, such as the Gold or Kasami codes, address this problem [6].

3. This method does not specifically protect the DC value of the pixels covered by an individual block.

In [2], the watermark consists of extended m-sequences of length 512 bits added to each pixel in a row of the image. Extended m-sequences are m-sequences of order *n*, with a 0 inserted at the end of the *n - 1* run of zeros. The phase of the extended m-sequence carries the watermark information. The testing procedure filters the crosscorrelation function of the possibly forged, watermarked image row and the extended m-sequence. If a suitably large crosscorrelation peak is found, the row passes the watermark test.

Our watermark uses a much longer m-sequence, which is arranged row by row into a two-dimensional block. We append a 0 to the entire m-sequence, instead of using an extended m-sequence. Enough blocks are concatenated to cover the entire image. One advantage of a two-dimensional watermark is the ability to more effectively locate *where* an image has been changed. Forgeries made to only a small portion of the image would affect the respective block and not the entire row of the image. Our testing algorithm simply overlays the watermarked image block and the watermark block, computes an inner product, and compares the result to the ideal value. If the difference relative to the ideal value is larger than a defined threshold, the block fails the watermark test. This forgery detection algorithm eliminates the need to compute an entire crosscorrelation function. The details of this are described below.

One must perform several operations on each block of pixels in the image to test the new watermark. We first define the spatial crosscorrelation function of images X and Y as:

$$R_{XY}(\alpha,\beta) = \sum_i \sum_j X(i,j)Y(i-\alpha, j-\beta) \qquad (1)$$

Let X be the original image block, W be the watermark block, Y be the watermarked image block and Z be the watermarked image block that might be forged. The test statistic for the block, $\delta$, is defined as:

$$\delta = R_{YW}(0,0) - R_{ZW}(0,0) \qquad (2)$$

If the watermarked image is unchanged, $\delta = 0$. Note that $\delta$ does not depend on the entire crosscorrelation function. When $\delta$ is larger than a defined tolerance, the block fails the watermark test. A larger threshold provides more robustness, but increases the probability of missing a forgery. A threshold can be defined relative to the number of elements in the watermark block.

## 3. RESULTS OF FILTERING

One question that needs to be addressed is how robust is the watermark to typical image processing operations. The first experiment examines the effect of mean and median filtering on forgery detection. The test image consists of a 768 x 512 pixel grayscale image. The watermark block size was chosen to be 256 x 256 pixels. An m-sequence with a period of 65,535 with a single zero bit appended to the end of the sequence was used. It was segmented into 256 bit sections, then arranged row by row to form the watermark block. A 3 x 2 array of these blocks formed the watermark, which covered the entire image. Three different window sizes for each type of filter were applied to two regions in the image. The goal was to see if the watermark could be used to detect these alterations to the image.

The watermark test is able to detect every case of filtering. If the threshold for $\delta$ is set low enough, each image would fail the watermark test. Table 1 shows how each filter affected $\delta$. The filter sizes were: 3 x 3, 7 x 7, and 11 x 11. Region 1 is 11 x 20 pixels (0.34% of the block), and region 2 is 81 x 160 pixels (19.78%). The percentage change with respect to the number of elements of the watermark block (65,536) that $\delta$ represents is also shown. In each case the percentage change in $\delta$ was roughly equal to the percentage of the block affected by the filter. This indicates that the damage to the watermark was proportional to the area of the image block that was filtered. An example of a test scenario would be if an owner wanted to detect filtering of more than five percent of an image block. The threshold would be set to 3277. Changes in the first region would all pass the test, and changes in the larger, second region would cause the block to fail the watermark test. Even though this watermark is adequate, in the next section a new, more secure watermark is described.

## 4. AN IMPROVED WATERMARK

The previous watermarking technique was revised to improve security and localization. Localization is the ability to identify where in the image any changes have occurred. The block size is 8 x 8 pixels, and each block is formed as follows:

1. A large span m-sequence (n = 96) is generated with the first 128 bits skipped.
2. The next 64 bits are inserted in the first block of the watermark *column by column*. The next 32 bits are skipped.
3. The process repeats for the remaining blocks. These blocks make up the watermark row by row. This forms a 64 x 96 array of watermark blocks that cover the entire image (total of 6144 blocks).

Table 1. $\delta$ after mean and median filtering.

| *Filter size:* | *3 x 3* | *7 x 7* | *11 x 11* |
|---|---|---|---|
| | Mean Filter | | |
| $\delta$, Region 1 | 201 | 279 | 288 |
| % of block size | 0.31 % | 0.43 % | 0.44 % |
| $\delta$, Region 2 | 11562 | 12550 | 13055 |
| % of block size | 17.6 % | 19.15 % | 19.92 % |
| | Median Filter | | |
| $\delta$, Region 1 | 205 | 267 | 351 |
| % of block size | 0.31 % | 0.41 % | 0.54 % |
| $\delta$, Region 2 | 11297 | 12914 | 13208 |
| % of block size | 17.24 % | 19.71 % | 20.15 % |

## 5. JPEG COMPATIBILITY

This section describes how the revised watermark can be used in conjunction with JPEG compression. First, the image is watermarked with the revised scheme described above. JPEG compression is performed on the watermarked image. The image is then decompressed. The values of $\delta$ for each block were obtained as described in Equation 2.

Two different versions of the revised watermark have shown promise with JPEG. One version consists of 0 and 1, and the other consists of -1 and 1 (the bipolar watermark). Since the {0,1} watermark has greater low frequency energy than the bipolar one, it was thought that JPEG might destroy less of this watermark. This would mean that the average value of $\delta$ would be less after JPEG compression and decompression. Figure 1 shows the average value of $\delta$ using all 6,144 blocks of the watermark for various compression levels using both versions of the revised watermark.

The average value of $\delta$ for the {0,1} watermark was much lower than for the bipolar watermark. If each watermark should have zero mean (so as not to affect the image brightness), or if many watermarks will occupy the entire image, bipolar watermarks are more appropriate. This is especially crucial when building an audit or viewing history of an image, where many watermarks could occupy the same image. Because of this requirement, the remaining experiments use bipolar watermarks with quality factors of 75 and 85. For 24-bit RGB images, JPEG compresses each individual color plane as a monochrome image. JPEG's effect on the individual color planes was similar to that of the luminance image, with the red plane having slightly less of the watermark than the other two.

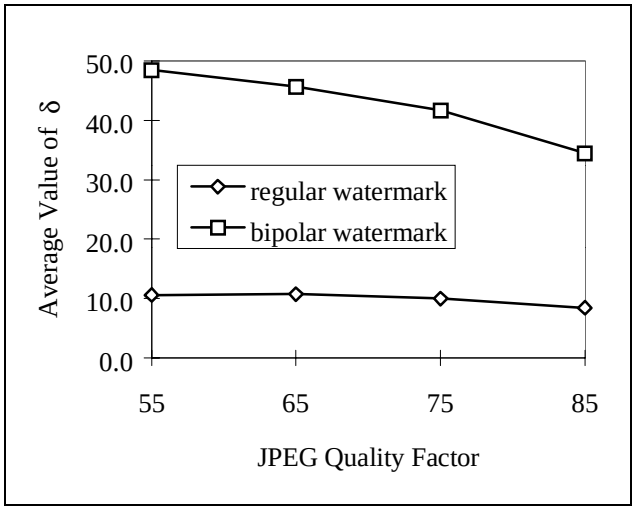


**Figure 1.** Average Value of $\delta$ vs. Quality Factor.

The changes in the individual values of $\delta$ determine *where* any image alterations occurred. Values of $\delta$ can vary widely after JPEG compression and decompression. Blocks with primarily low frequency energy usually had higher values of $\delta$ than blocks with a large portion of high frequency energy. Figure 2 shows histograms of $\delta$ for the image with the bipolar watermark for quality factors of 75 and 85.

The JPEG compressed and decompressed image would fail the watermark test with even the most generous thresholds on $\delta$. The large range in the values of $\delta$ motivates two changes to our previous forgery detection procedure. Let $Y_J$ be the watermarked image after JPEG compression and decompression, and $Z_J$ be a possibly forged watermarked image after JPEG processing. A new test statistic must be defined:

$$\delta_J = R_{Y_J W}(0,0) - R_{Z_J W}(0,0) \qquad (3)$$

The next section examines the performance of this new test statistic in the presence of small changes to $Y_J$.

## 6. DETECTION OF RANDOM BIT ERRORS

We would like to determine if image changes can be detected with our JPEG watermarking algorithm and new test statistic in Equation 3. The procedure is as follows:

1. The original image is first watermarked with the revised watermark, then JPEG compressed and decompressed.
2. Intensities of randomly selected pixels in the decompressed image were either raised or lowered by one bit. This is to approximate randomly occurring LSB transmission errors, or an attempted forgery. This formed the tampered image $Z_J$.
3. The $\delta_J$ for the tampered image were determined.

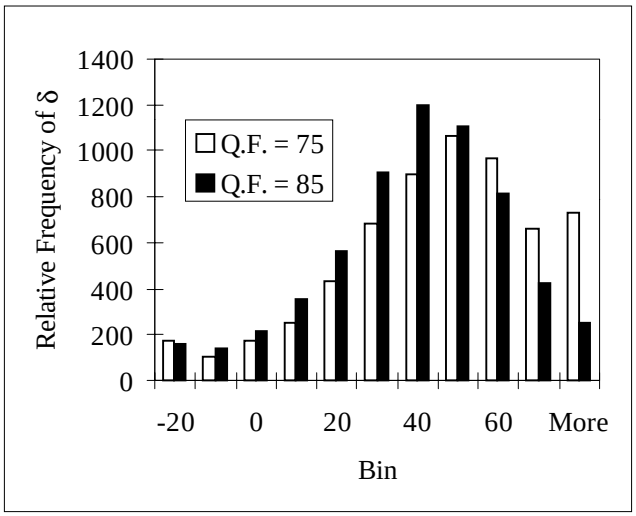


**Figure 2. Histogram of δ**

The watermark test can still detect changes to the image in most cases. Table 2 shows how random bit errors in the decompressed watermarked image affect $\delta_J$. Results will depend on where the bit errors occur, and on the image itself. One could also threshold $\delta_J$ for each block, in addition to thresholding the average value of $\delta_J$ using all blocks.

## 7. CONCLUSION

The proliferation of network multimedia systems dictates the need for copyright protection of digital property. This paper presents a visually undetectable, robust watermarking scheme. Our techniques can detect the change of a single pixel and can locate where the changes occur. The algorithms work for color images and can accommodate JPEG compression. Future research includes watermarking MPEG video sequences.

A postscript version of this paper is available via anonymous ftp to **skynet.ecn.purdue.edu** in the directory **/pub/dist/delp/icip96-secure**.

Table 2. $\delta_J$ after the introduction of random bit errors.

| *Pr. {bit error}* | *Avg.* $\delta_J$ | *Avg.* $\delta_J^2$ | *Max.* $\lvert\delta_J\rvert$ |
|---|---|---|---|
| | Quality Factor = 75 | | |
| 0.01 | 2.83 e-02 | 6.23 e-01 | 4 |
| 0.001 | 2.28 e-03 | 6.71 e-02 | 2 |
| 0.0001 | -1.30 e-03 | 7.81 e-03 | 1 |
| 0.00001 | -4.88 e-04 | 4.88 e-04 | 1 |
| | Quality Factor = 85 | | |
| 0.01 | 2.93 e-02 | 6.21 e-01 | 4 |
| 0.001 | 2.12 e-03 | 6.66 e-02 | 2 |
| 0.0001 | -1.30 e-03 | 7.81 e-03 | 1 |
| 0.00001 | -4.88 e-04 | 4.88 e-04 | 1 |

## 8. REFERENCES

[1] S. Walton, "Information authentication for a slippery new age," *Dr. Dobbs Journal*, vol. 20, no. 4, pp. 18-26, April, 1995.

[2] R. G. van Schyndel, A. Z. Tirkel, N. R. A. Mee, C. F. Osborne, "A digital watermark," *Proceedings of the International Conference on Image Processing*, November, 1994, Austin, Texas, vol. 2, pp. 86-90.

[3] J.-F. Delaigle, C. De Vleeschouwer, B. Macq, "Digital watermarking," accepted for publication, *Journal of Electronic Imaging*.

[4] F. M. Boland, J. J. K. Ó Ruanaidh and C. Dautzenberg, "Watermarking digital images for copyright protection," *Proceedings of the International Conference on Image Processing and its Applications*, July 1995, Edinburgh, Scotland, pp. 321-326.

[5] Fred Mintzer, Albert Cazes, Francis Giordano, Jack Lee, Karen Magerlein and Fabio Schiattarella, "Capturing and preparing images of Vatican library manuscripts for access via internet," *Proceedings of IS&T's 48th Annual Conference*, May, 1995, Washington, DC, pp. 74 - 77.

[6] J. Proakis, *Digital Communications*, McGraw-Hill, 1983.