Combining low-frequency and spread spectrum watermarking

Jiri Fridrich

Center for Intelligent System, SUNY Binghamton, Binghamton, NY 13902-6000
Mission Research Corporation, 1720 Randolph Rd. SE, Albuquerque, NM 87501

ABSTRACT

Low-frequency watermarks and watermarks generated using spread spectrum techniques have complementary robustness properties. In this paper, we combine both watermarking paradigms to design an oblivious watermark that is capable of surviving an extremely wide range of severe image distortions. An image is first watermarked with a low-frequency pattern and then a spread spectrum signal is added to the watermarked image. Since both watermarks are embedded in a different portion of the frequency space, they do not interfere. For the low-frequency watermark, we modify the NEC scheme so that the original image is not needed for watermark extraction. The image is first normalized and the watermark is embedded into the lowest frequency discrete cosine modes by encoding a binary pattern (a watermark) using a special quantization-like function. The watermark detector uses a weighted correlation and a simple one-dimensional search to adjust for proper normalization. The low-frequency watermark is combined with a spread spectrum signal added to the middle frequencies of a DCT. The resulting double watermarked image is extremely robust with respect to a very wide range of quite severe image distortions including low-pass filtering, pixel permutations, JPEG compression, noise adding, and nonlinear deformations of the signal, such as gamma correction, histogram manipulation, and dithering.

Keywords: image watermarking, spread spectrum, robust message hiding

1. INTRODUCTION

Digital watermark is a perceptually transparent pattern[(i)] embedded in an image using an embedding algorithm and a secret key. The purpose of the watermark is to supply some additional information about the image without visibly modifying the image (compare with date and time imprinting on negatives) or without the need to change the file format. Information appended in a visible form in the image or added to the header of a corresponding image format can be easily erased or replaced. Digital watermark is embedded in the image in an invisible form yet in a persistent, robust manner. The process of embedding a watermark depends on a secret key so that only those possessing the key can access the information hidden in the watermark. With the key, the information carried by the watermark can be read and decoded using a detection algorithm.

An important property of a watermark is robustness with respect to image distortions. This means that the watermark should be readable from images that underwent common image processing operations, such as filtering, scaling, noise adding, cropping, etc. Watermarks that are to be used for copyright protection, fingerprinting, or access control must also be embedded in a secure form. This means that an attacker who knows all the details of the embedding algorithm except the secret key should not be able to disrupt the watermark beyond detection. In such applications, the watermarking scheme is an example of a symmetric encryption scheme with private key[1]. In other applications when it is desirable that the watermark information be publicly accessed by a large number of people, such as adding additional captions to images or subtitles in several languages to movies, there is no motivation for intentional removal of the watermark, and the security of the watermark is not an issue. Although some candidates for a secure public detector have been proposed[2], almost all watermarking schemes that have been described in the literature so far have the property that the ability to read the watermark automatically implies the ability to remove the watermark[3,4]. The number of bits carried by the watermark could be as low as one bit or several hundred bits or more. Obviously, there is a trade-off between robustness and the capacity of the watermark.

Another important attribute of watermarking is the computational complexity of the embedding and extracting procedures. In some applications, it is important that the embedding process be as fast and simple as possible (e.g., embedding serial

Further author information –
E-mail: fridrich@binghamton.edu; WWW: http://ssie.binghamton.edu/~jirif
[(i)] Visible watermarks are not discussed in this paper.

numbers of digital cameras into images for the purpose of tamper detection) allowing the extraction to be more time consuming. In other applications, the speed of extraction is absolutely critical (e.g., extracting subtitles from movies). To summarize, the required properties of digital watermarks are:

- Robustness to common image processing operations – untargeted attacks
- Security – targeted attacks (application dependent)
- Perceptual invisibility
- Restrictions on computational complexity of embedding/extraction (application dependent)

Non-oblivious watermarking schemes[5,6,7] must access the original image in order to extract the watermark. The original image is usually subtracted from the suspected image before a detection algorithm is applied. The original image can also be used for registering the suspected image if it has been cropped, rotated, scaled, or transformed in some more general manner (e.g., as in StirMark[8]). Obviously, the availability of the original image makes non-oblivious watermarking schemes much more robust than oblivious schemes, which extract watermarks without accessing the original image. Non-oblivious watermarking is at present the only option for reliable copyright protection. Currently, there is no computationally efficient oblivious scheme that would be able to reliably extract watermarks from images that underwent general non-linear geometric transformations, such as those introduced by a general-purpose watermark removing software StirMark[8]. The quest for StirMark resistant oblivious watermarking scheme remains an active research topic. O' Ruanaidh et al.[9,10] have described an oblivious technique that uses a calibration pattern embedded into the amplitude of Fourier transform in log-polar coordinates. This enables them to register the suspected image after a combination of a shift, rotation, and change of scale. A second, spread-spectrum type of watermark embedded in the middle frequencies is used to carry a message of up to 100 bits or longer.

Most watermarking techniques can be roughly divided into two groups depending on whether the watermark is inserted by modulating the coefficients of some transform or directly the pixel values. In some techniques[6,7,11,12,23], the modulation is adjusted according to properties of the human visual system so that no perceptually visible distortions are introduced by the watermark. Transform-based techniques may use DCT[5,6,11,12], DFT[13], Hadamard transform[14], wavelets[15,16], or general, key-dependent transforms[2]. Typical representatives of techniques that embed watermark patterns directly into pixel values are[17,18]. The watermark pattern itself can have its energy mostly concentrated in low or high frequencies depending on the technique. Noise-like watermarks generated using spread spectrum methods in the spatial or frequency domains are statistically orthogonal to the original image, and can be extracted by performing a simple dot product with the watermarked image or a portion of its spectrum.

Low-frequency watermarks interfere with the image and it is thus necessary to have the original image for watermark extraction. On the other hand, the low-frequency character of the watermark does not increase the noise level of the image and increases the robustness with respect to image distortions that have low-pass character (filtering, nonlinear filtering such as median filter, lossy compression, adaptive Wiener filtering, etc.). Low-frequency watermarks also have fewer problems with synchronizing the watermark detector with the image and are less sensitive to small geometric distortions. On the other hand, oblivious schemes with low-frequency watermarks are more sensitive to modifications of the histogram, such as contrast/brightness adjustment, gamma correction, histogram equalization, and cropping.

Watermarks inserted mostly into middle and high frequencies are typically less robust to low-pass filtering and small geometric deformations of the image, but are extremely robust with respect to noise adding, nonlinear deformations of the gray scale, such as contrast/brightness adjustment, gamma correction, and histogram manipulations.

It is understandable that the advantages and disadvantages of low and middle-to-high frequency watermarks are complementary. It appears that by embedding both watermarks into one image, one could achieve extremely high robustness properties with respect to a large spectrum of image processing operations. Indeed, inserting a high-frequency spread spectrum signal on top of an image previously watermarked with a low-frequency watermark could lead to a scheme that enjoys the advantages of both watermarks. There will be very little interference between both watermarks since they will be inserted into two disjoint portions of the spectrum. However, it is not entirely clear how one would build an oblivious technique with low-frequency watermarks.

The purpose of this paper is to design a watermarking technique that is robust to a very wide range of extremely severe image distortions. To achieve this goal, we combine a low frequency watermark with a frequency-based spread spectrum scheme. In Section 2, we adapt the non-oblivious NEC watermarking scheme by Cox et al.[5] so that the need for the original image is removed. The NEC scheme has very impressive robustness properties and has been extensively analyzed in the past[5,19]. Since the watermark is spanned by low-frequency cosine modes, it is extremely robust with respect to image distortions that have low-pass character. On the other hand, the watermark interferes with the image in the sense that the watermark pattern is not orthogonal to the image. Because of this interference, it seems that it is necessary to subtract the original image before detection. To overcome this difficulty, we encode a binary pattern into low-frequency DCT coefficients. The detection function than simply checks for that pattern. After an image is watermarked with a low-frequency watermark, we add a noise-like signal to the middle frequencies of the watermarked image. In Section 3, we describe this frequency-based spread spectrum technique, which is a modification of the scheme proposed by Ó Ruanaidh[9]. It also bears some similarity to the technique introduced by Piva[11]. Section 4 is devoted to the robustness study of the combined watermarking technique. The paper is closed with Section 5 that discusses possible improvements and directions for future research.

2. OBLIVIOUS LOW-FREQUENCY WATERMARKING

There are many different methods how to make a low-frequency watermarking scheme oblivious. Koch and Zhao[20] encode a relationship into small 8×8 blocks of DCT coefficients by swapping selected DCT coefficients. However, the reported robustness is not too high and may introduce visible changes[21]. Schneider and Chang[22] describe a variation of this technique based on enforcing a relationship between the coefficients.

Kundur and Hatzinakos[15,16] embed message bits into disjoint triplets of wavelet coefficients chosen from the same resolution level. The middle coefficient is adjusted so that its relative position with respect to the other two coefficients falls into intervals of length $(c_{max}-c_{min})/(2Q-1)$, where c_{max} and c_{min} are the largest and the smallest wavelet coefficients from each triplet, and Q is a fixed integer. The number Q can be adjusted to obtain a good trade-off between robustness and watermark visibility. The choice of the triplets is based on a pseudo-random number generator initialized with a secret key. This method, however, cannot be directly used for discrete cosine basis because the standard deviation of the DCT coefficients corresponding to low frequencies is typically very large. To guarantee the watermark invisibility, one is forced to choose a high value for Q thus losing the robustness. Nevertheless, this method may work quite well for DCT frequencies in the middle range.

Swanson et al.[12] propose to use frequency and spatial masking to calculate the maximal allowable changes for each DCT coefficient in each 8×8 block. Each DCT coefficient is then quantized and adjusted so that its new value is either larger or smaller than the closest multiple of the largest allowable difference. In another paper[23] the same authors present a variation of this scheme in which the normalized DCT of author's signature is projected on the DCT of each image block and the value of the projection is modified using frequency masking values of that block. This technique enables insertion of multiple bits per block and can probably be made quite robust by sacrificing capacity for robustness. The fact that the quantization is dependent on the frequency masking values of small blocks indicates that the watermark will probably not survive large image distortions, such as mosaic filter or median filter with large kernel. Increasing the block size is not an option for this technique because the frequency masking model becomes very unreliable for blocks larger than 16×16.

In this paper, we took the following approach. The image is first converted to a signal with zero mean and certain standard deviation so that the DCT coefficients always fall into a prespecified, fixed range. The transformation

$$Y \to \frac{1024}{\sqrt{MN}} \frac{X - \overline{X}}{\sigma(X)}, \tag{1}$$

where $\sigma(X)$ and $\overline{X}$ denote the standard deviation and mean of X, maps the grayscale image X into a two-dimensional signal Y with zero mean such that the maximal DCT coefficients of Y are in absolute value close to 200–250. This transformation works quite well for a wide range of different images. We tested natural images with large uniform regions, highly textured images, as well as images containing man-made structures, and mixtures of all of the above.

For image Y, we calculate the DCT transformation and adjust the DCT coefficients so that a certain binary pattern is encoded. We start by defining a geometric sequence of real numbers,

$$x_{i+1} = \frac{1+\alpha}{1-\alpha} x_i, \qquad x_0 = 1,$$

parametrized by $\alpha \in (0,1)$. For $x > 1$, $x_i \le x < x_{i+1}$ we define an index function $ind(x) = (-1)^i$. It is easy to see that any real number $x > 1$ can be modified by adding or subtracting at most αx to change its index $ind(x)$. The index function for $\alpha = 0.1$ is depicted in Figure 1.

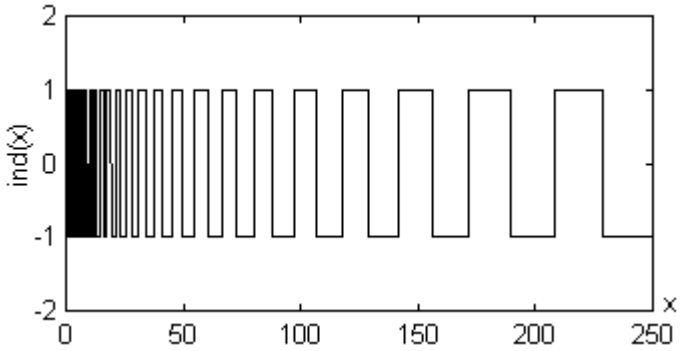


Figure 1 The index function $ind(x)$

To encode a watermark pattern $\{w_i\}_{i=1}^N$, $w_i \in \{-1,1\}$, we select N lowest frequency DCT coefficients d_i and adjust their values so that $ind(|d_i|) = w_i$. If $|d_i| < 1$, we leave the coefficient untouched. Because of the properties of the index function, each coefficient will be modified by at most 100α percent. It is also reasonable to expect that the changes will be practically random because there is no reason for DCT coefficients to initially follow any particular pattern.

The highest robustness with respect to image distortions will be achieved if we set the new DCT coefficients as the midpoints of the intervals $[x_i, x_{i+1}]$. However, this would cause easily identifiable clustering of DCT coefficients thus making the scheme insecure. It seems that the only way to introduce security into this scheme without noticeably decreasing the robustness is to use secret, key-dependent basis functions instead of discrete cosines. In fact, a version of this scheme with a smooth index function and a key-dependent random, smooth basis was previously introduced as a possible candidate for a secure public watermark detector[2]. By making the basis key-dependent, one could intentionally encode clusters into the projections on the new basis functions (counterparts of DCT coefficients). The detector would then check for presence of regularly separated clusters in the set of projections. This offers an interesting idea for a detection function that is not based on a correlation. In this paper, we concentrate on robustness rather than security and we do not discuss security issues further.

The value of α was adjusted so that the watermark is perceptually invisible. To this purpose, we used the linearized spatial masking model of Girod[24]. This model accurately predicts maximal image distortions in uniform areas and close to edges. To get a feeling about the right value of α, we performed experiments using the test image "Lenna" for N = 100, 150, 200, …, 1000 for progressively larger values of α. The results are summarized in Figure 2. For $\alpha < 0.03$, there are no visible changes induced by this watermarking technique. For $\alpha = 0.05$ and $N = 300$ the spatial masking model indicated roughly 6% portions of pixels containing visible changes.

Just for comparison, we used Girod's model for the NEC scheme[5] with watermark strength $\alpha = 0.1$ and found that this value of α caused visible changes for over 15% of pixels. Therefore, we feel that our choice of α, which introduces 6% of visible changes, is more on the conservative side.

The detection is obtained by converting the suspected image in the same manner as for inserting the watermark. A DCT is calculated for the converted, suspected image. Let us denote N lowest frequency DCT coefficients by d_i'. Since all N DCT coefficients were modified in the same manner by at most 200α percent, a simple correlation between $ind(|d_i'|)$ and w_i would produce a non-robust technique because the small, perceptually insignificant DCT coefficients contribute with the

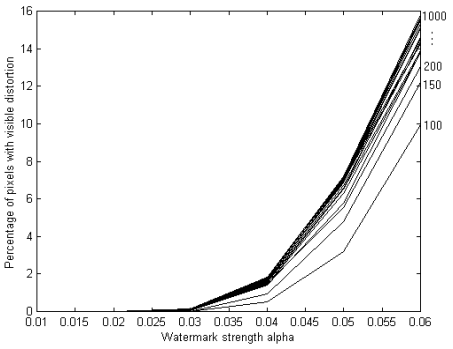


Figure 2 Percentage of pixels with visible changes as a function of watermark strength α. The curves correspond to increasing values of N = 100, 150, …, 1000.

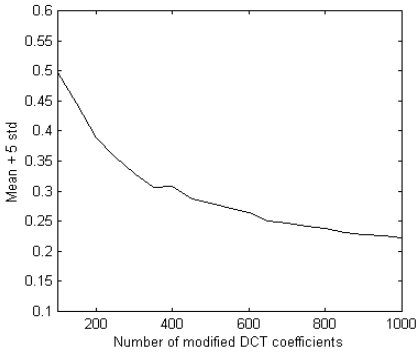


Figure 3 Detection threshold based on detecting 10000 random watermarks in one watermarked image.

same weight as the larger, perceptually more significant coefficients. Since we want our technique to be oblivious, we cannot constrain ourselves to only the largest N DCT coefficients because these could be different for modified images and their order might also be different. In our scheme, we decided to mark all coefficients, but let only the largest ones to contribute to the correlation. Therefore, we use a correlation weighted by the absolute values of the DCT coefficients

$$corr = \frac{\sum_{i=1}^{N} |d_i'|^{\beta} \; ind(|d_i'|) w_i}{\sum_{i=1}^{N} |d_i'|^{\beta}}.$$

The weights will automatically put more emphasis on large, perceptually significant modes while suppressing the insignificant modes that may be affected by image processing operations. Therefore, the correlation will only be affected by the largest DCT coefficients as in the NEC scheme[5]. The factor β can be used to adjust the importance of weighting. For β=0 a regular, non-weighted correlation is obtained. Large values of β would lead to a singular scheme whose detection function would depend on just one bit corresponding to the largest coefficient. In our studies, we found that $\beta \in (1/2,1)$ produced the best results.

It is possible to further increase the robustness of this technique by searching for maximum correlation with respect to standard deviation of the suspected image. The scaling (1) depends on the standard deviation of the suspected image Y. The standard deviation may be changed significantly, if some smoothing or noise adding operation is applied to the watermarked image. As a result, the DCT coefficients of the suspected image will be scaled by a fixed number (the ratio of the standard deviation of the original image and the suspected image). The pattern encoded into the DCT coefficients will, however, be unaffected by the linear change. This suggests that a simple one-dimensional search for the right scale that maximizes the correlation will help. Therefore, the complete detection function is as follows

$$corr = \max_{s \in (1-\delta, 1+\delta)} corr(s) = \frac{\sum_{i=1}^{N} |d_i'|^{\beta} \; ind(|sd_i'|) w_i}{\sum_{i=1}^{N} |d_i'|^{\beta}}.$$

The value of $\delta = 1/4$ seems to be adequate even for large image distortions. We note that the one-dimensional search can be done very quickly and does not noticeably increase the computational complexity of the detecting procedure. The most time-consuming procedure in both embedding and detection is the two-dimensional DCT. On a 333MHz Pentium II computer, the embedding takes less than 1 second and the detection less than ½ second for a gray scale image with 256×256 pixels. The increase in computational complexity is determined by the complexity of the DCT.

The weights in the detection decrease the information content of the watermark w. Since only the largest coefficients contribute to the detection, the information content of a watermark of length N is some fraction of N. Depending on the factor β, the capacity of the watermark is determined by the amplitudes of N lowest frequency DCT modes.

The one-dimensional search for the scaling factor that maximizes the correlation will clearly increase the percentage of false detections. Therefore, we carefully choose the threshold value so that the probability of false detections is below a user-defined limit. We watermarked a 256×256 test image “Lenna” with one watermark and then tested the presence of 10000

randomly generated watermarks. Then, we calculated the mean m and the standard deviation σ of the resulting correlation values and set the threshold to $Th = m + 5\sigma$. On the condition that the correlations are Gaussian distributed, the probability of a false detection is less than 5.7×10^{-7} = 1:1,744,300. The threshold is plotted as a function of N (for $\alpha = 0.05$) in Figure 3.

3. FREQUENCY-BASED SPREAD SPECTRUM

As explained in the introduction, to achieve high robustness properties with respect to as many image distortions as possible, we insert an additional watermark using a spread spectrum technique. We use the scheme proposed by Ó Ruanaidh[9] (a similar technique was proposed by Piva[11]). The watermark is inserted by adding a noise-like signal to the middle frequencies of its DCT. The DCT coefficients are converted to a vector and the middle 30% (N_m frequencies) is chosen for marking. The information carried by the watermark consists of M symbols and each symbol s_i is represented using r bits, $1 \le s_i \le 2^r$. For each i, a sequence $\xi^{(i)}$ of pseudo-random numbers of length N_m+2^r uniformly distributed in [0,1] is generated. Symbol s is represented using the segment $\eta^{(i)} = \xi_s^{(i)}, \ldots, \xi_{s+N_m-1}^{(i)}$ of consecutive N_m pseudo-random numbers. For each symbol a new sequence of pseudo-random numbers is generated. The seed for the PRNG serves as the secret key. The message of M symbols is then represented as a summation

$$S_p = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \eta^{(i)} .$$

The spread spectrum signal S_p is approximately Gaussian with zero mean and unit standard deviation even for moderate values of M (e.g., $M \approx 10$). The signal S_p is further multiplied by a parameter γ (watermark strength / visibility) and added to the middle N_m DCT coefficients d_j. Again, we used the spatial masking model of Girod[24] to adjust γ so that the double watermarked image is perceptually identical to the original image. The value of $\gamma = 13$ worked well for most images. The amplitude of the combined watermark was typically in the range [−20,20] with an average rms of 5 gray levels (see Figures 4 and 5). The watermark was repeatedly embedded in blocks of 128×128 pixels.

The detection of the message of M symbols proceeds by first transforming the image using a DCT and extracting the middle N_m DCT coefficients. The secret key is used to generate M pseudo-random sequences of length N_m+2^r needed for coding the message symbols. For each sequence, all 2^r segments of length N_m are correlated with the middle N_m DCT coefficients. The largest value of the correlation determines the encoded symbol.

Figure 4 Test image “Lenna”

Figure 5 Watermarked image

For a very wide range of contrast and brightness adjustments, histogram operations (stretching, equalization), and severe degradations caused by superimposing white Gaussian noise, we could actually recover a complete message of 10 6-bit symbols without any errors. To trade off robustness for capacity, we used a fixed message consisting of $M = 10$ identical symbols. The probability $P(k,M)$ that at least k r-bit symbols from the recovered message of length M are correct is

$$\frac{\binom{M}{k}}{2^{rk}}. \qquad (2)$$

For $M = 10$ and $r = 6$ we have $P(k,M) < 3\times10^{-7}$ when $k \ge 5$. Consequently, we set the decision threshold for the watermark presence to at least 5 correctly recovered symbols. The probability of a false detection is less than 3×10^{-7}.

4. TESTING THE ROBUSTNESS

The robustness properties reported in this section were achieved with $N = 300$ lowest frequency DCT coefficients, $\alpha = 0.05$, and $\gamma = 13$. The test image was the grayscale image of "Lenna" with 256×256 pixels (see Figure 4). The absolute amplitude of the low-frequency watermark pattern was within the interval [−6,9] with an average rms of about 5. Adding the second watermark increased the watermark depth to [−20,20] while the average rms did not change. The robustness to JPEG compression is shown in Figure 6. The low-frequency watermark could be extracted from JPEG compressed images with quality factor as low as 4%. The spread spectrum watermark was lost at the quality factor of 20%. Next, we tested the robustness with respect to low-pass filtering. We used the same kernel that is implemented for blurring in PaintShop Pro 4.12. The blurring kernel is shown in Figure 8. We could recover the low-frequency watermark after 11 repeated applications of the blurring filter (Figure 7).

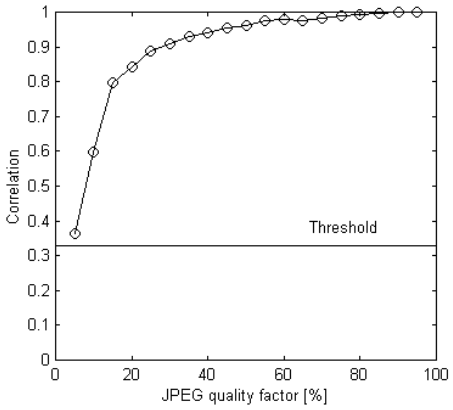


Figure 6 Robustness to JPEG compression

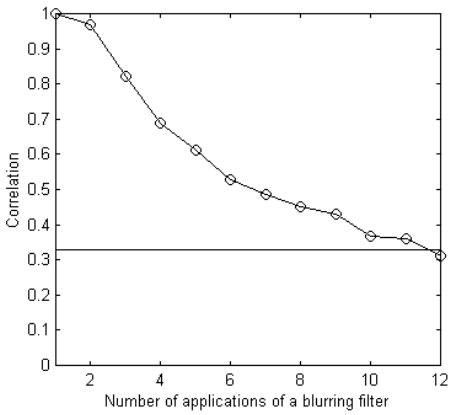


Figure 7 Robustness to repeated blurring

$1/44$

1	1	2	1	1
1	2	2	2	1
2	2	8	2	2
1	2	2	2	1
1	1	2	1	1

Figure 8 Blurring kernel

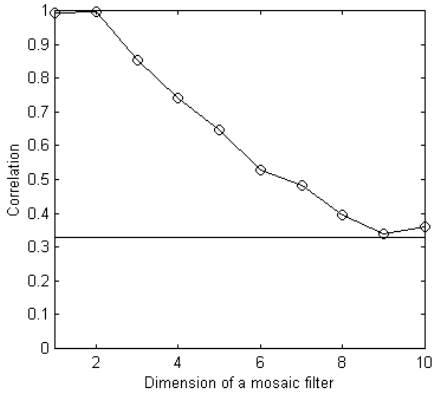


Figure 9 Robustness to mosaic filtering

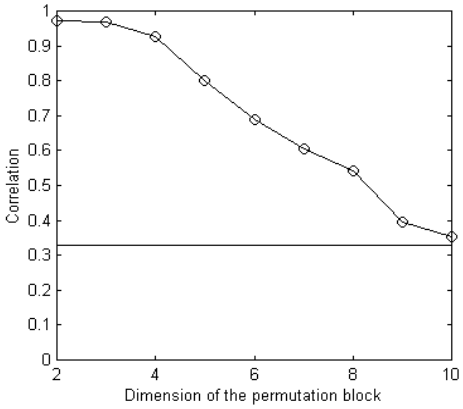


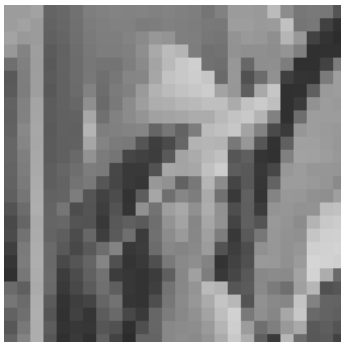
Figure 10 Robustness to permuting neighboring pixels

The low-frequency watermark was remarkably robust with respect to the mosaic filter. The mosaic filter divides the image into square regions of s×s pixels and replaces the whole square with an average gray value. We could recover the watermark after applying the mosaic filter with squares 10×10 pixels wide (see Figure 9). Probably the most impressive performance was achieved for image deformations due to permuting pixels. We wrote a simple routine in which the image was again divided into squares and the pixels in those squares were randomly permuted. The correlation shown in Figure 10 decreases

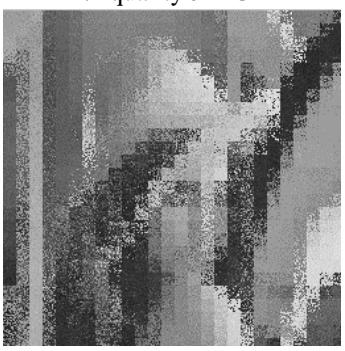
very gradually with the filter size and the low-frequency watermark can be detected even after permuting squares with side 10. The low-frequency watermark was also reasonably robust with respect to noise adding, however, the spread spectrum technique was clearly superior. We detected the watermark presence after adding white Gaussian noise with zero mean and standard deviation 90 gray levels. Histogram equalization successfully removed the low-frequency watermark, while the spread spectrum watermark completely retained its integrity (the whole 60-bit message was recovered). The message was also completely recovered after dithering to a black-and-white image. Below are examples of extreme image deformations from which watermark could be extracted. Even very large contrast / brightness adjustments (e.g., 80% or more in PaintShop Pro 4.12) did very little damage to the spread spectrum watermark, while the low-frequency watermark was lost at roughly ±25% brightness adjustment.

4% quality JPEG

11× blurred



Mosaic 10×10 filter



Permutation of 10×10 squares

Gaussian noise, σ =90

Dithering

-80% brightness adjustment

+80% brightness adjustment

The last two images due to extreme brightness adjustment render the image of Lenna almost unrecognizable. For the low-brightness image, we tested the presence of one fixed watermark (number 248 Figure 11) for 300 different secret keys (seeds for a PRNG). For each key, we calculated the number *k* of correctly recovered symbols and plotted the probability of obtaining *k* correct symbols in Figure 11. The figure shows the logarithm at the base 10 of the probability.

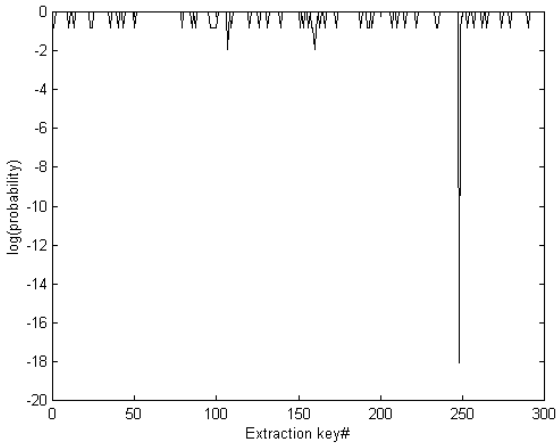


Figure 11 Logarithm of probability of extracting correct symbols for 300 randomly generated keys.

Other tested operations include median filtering (low-frequency watermark stays up to kernel size 7), gamma correction, high-pass filtering or sharpening. The spread spectrum signal was extremely robust with respect to histogram distortion, noise adding, and general nonlinear transformations of the signal. On the other hand, low-pass filtering, adaptive Wiener filtering, and other operations of low-pass character cause fast decline of correlation for spread spectrum techniques and the low-frequency watermark outperforms the spread spectrum watermark. While permutations of neighboring pixels does very little damage to the low-frequency watermark, it can make detection virtually impossible for spatial direct spread spectrum techniques[25,26].

5. SUMMARY AND FUTURE DIRECTIONS

In this paper, we proposed a technique that embeds two watermarks into an image: one low-frequency watermark and a medium frequency spread spectrum signal. The robustness properties of both watermarks combine into a scheme that can achieve robustness to a very wide range of severe image distortions. We believe that it would not be possible to achieve such a degree of robustness using a single scheme.

For the low-frequency watermark, we have modified the NEC scheme[5] so that the original image is not needed for watermark detection. First, the original image is transformed to a 2D signal with zero mean and a certain standard deviation, that depends on the image size, to guarantee that the DCT coefficients will fall within a fixed interval. A piece-wise constant step function with intervals of increasing length is used to convert the DCT coefficients into a sequence of 1s and –1s. The lowest N frequency DCT coefficients are modified so that a certain pattern of 1 and –1s is encoded. The design of the index function guarantees that any binary pattern can be encoded by modifying the coefficients by at most $100\alpha\%$ for a fixed number α (watermark strength and visibility measure). The spatial masking model of Girod[24] is used to adjust the watermark strengths to achieve the highest possible robustness without introducing perceptually visible artifacts. The detection function is a correlation weighted by the DCT coefficients. To adjust for possible changes in the standard deviation of the suspected image, a simple one-dimensional search is performed to find the highest value of correlation. The weighted correlation automatically takes care of the fact that large DCT coefficients are more perceptually important than small ones. This increases the robustness of the method by a significant factor. On the other hand, it also decreases the information content of the watermark, because only those bits that correspond to the largest DCT coefficients actually carry information. Increasing the number of modified DCT coefficients can increase the information content of the watermark. The one-dimensional search for maximum correlation increases the overall value of the correlations. In order to avoid false detections, the threshold is set to the mean + five standard deviations of correlations produced by testing the presence of one watermark using 10000 randomly generated watermarks. The probability of a false detection is less than 1:1,700,000.

The second watermark was embedded into the middle frequencies of an image. We used a modification of the technique proposed by Ó Ruanaidh[9]. It is based on adding a noise-like signal that carries bits of watermark information to the middle frequencies of a DCT. The combined watermark exhibits remarkable robustness with respect to a number of different operations. The low-frequency watermark could be extracted from images that underwent extremely severe distortions, such as 11 consecutive blurring operations, 4% JPEG compression, random permutation of pixels in 10×10 squares, or applying

a mosaic filter with squares 10×10. The low-frequency watermark is also very robust to noise adding and median filtering. The second, spread spectrum watermark is extremely robust with respect to nonlinear transformations of the gray levels, such as gamma correction, histogram equalization, and dithering.

The main concern of this paper was robustness. The low-frequency watermark could be disrupted by modifying a selected fraction of the largest DCT coefficients by $100\alpha\%$. It seems that the only way to introduce security into this scheme without noticeably decreasing the robustness is to use secret, key-dependent basis functions instead of discrete cosines[2]. By making the basis key-dependent, one could intentionally encode clusters into the projections on the new basis functions (counterparts of DCT coefficients). The detector would then check for presence of regularly separated clusters in the set of projections. This offers an interesting idea for a detection function that is not based on a correlation. This could be very important for the design of a secure public watermark detector[3,4]. The major obstacle is a computationally efficient method for generating random, smooth orthogonal basis functions from a secret key. Smoothened random patterns orthogonalized using a Gram-Schmidt procedure were previously proposed[2] and shown to produce robust watermarking techniques. In our future effort, we intent to use warped weakly separable polynomial bases[27] for an efficient generation of such bases.

6. ACKNOWLEGEMENTS

The work on this paper was supported by Air Force Research Laboratory, Air Force Material Command, USAF, under a Phase II SBIR grant number F30602-98-C-0049. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation there on. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of Air Force Research Laboratory, or the U. S. Government.

7. REFERENCES

1. B. Schneier, "*Applied cryptography – protocols, algorithms and source code in C*", 2nd edition, John Wiley & Sons, New York, 1996.
2. J. Fridrich, "Robust digital watermarking based on key-dependent basis functions", *Proc. of the 2nd Information Hiding Workshop*, Portland, Oregon, April 15–17, 1998.
3. J. P. Linnartz and M. van Dijk, "Analysis of the sensitivity attack against electronic watermarks in images", *Proc. of the 2nd Information Hiding Workshop*, Portland, Oregon, April 15–17, 1998.
4. T. Kalker, J. P. Linnartz, and M. van Dijk, "Watermark estimation through detector analysis", *Proc. of the ICIP*, Chicago, October 1998. Submitted.
5. I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "A secure, robust watermark for multimedia", *Proc. of the 1st Information Hiding.Workshop*, edited by R. Anderson, LNCS, vol. 1174, pp. 183–206, Springer-Verlag, New York, 1996.
6. C. I. Podilchuk and W. Zeng, "Perceptual watermarking of still images", *Proc. of the First IEEE Signal Processing Society Workshop on Multimedia Signal Processing*, June 1997, Princeton, New Jersey.
7. M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking", *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, pp. 211–214, 1996.
8. M. G. Kuhn, "Stirmark", available at http://www.cl.cam.ac.uk/~mgk25/stirmark/, Security Group, Computer Lab, Cambridge University, UK (E-mail: mkuhn@acm.org), 1997.
9. J. J. K. Ó Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking", *Proc. of the ICIP*, vol. 1, pp. 536–539, Santa Barbara, California, 1997.
10. A. Herrigel, J. Ó Ruanaidh, H. Petersen, S. Pereira, T. Pun, "Secure copyright protection techniques for digital images," *Proc. of the 2nd Int. Information Hiding Workshop*, Portland, Oregon, April 15–17, 1998.
11. A. Piva, M. Barni, F. Bartolini, V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image", preprint, 1997.
12. M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust data hiding for images", *Proc. of the IEEE Digital Signal Processing Workshop*, pp. 37–40, Loen, Norway, September 1996.
13. J. J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland, "Phase watermarking of digital images", *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, pp. 239–242, Lausanne, Switzerland, September 1996.
14. J. J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland, "Watermarking digital images for copyright protection", *IEE Proc. Vision, Image and Signal Processing*, **143**(4), pp. 250–256, August 1996.
15. D. Kundur and D. Hatzinakos, "A robust digital image watermarking method using wavelet-based fusion", to appear in *Proc. Int. Conference in Image Processing*, 1997.

16. D. Kundur and D. Hatzinakos, "Digital watermarking based on multiresolution wavelet data fusion", *Proc. IEEE, Special Issue on Intelligent Signal Processing*, submitted, 1997.
17. I. Pitas, "A method for signature casting on digital images", *Proc. of the IEEE International Conference on Image Processing*, vol. 3, pp. 215–218, September 1996.
18. W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding", Technical report, MIT Media Lab, 1996.
19. H. S. Stone, "Analysis of attacks on image watermarks with randomized coefficients", NEC Research Institute, Technical Report, 1996.
20. E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling", *Proc. Nonlinear Signal Processing Workshop*, Thessaloniki, Greece, pp. 452–455, 1995.
21. G. C. Langelaar, J. C. A. van der Lubbe, and J. Biemond, "Copy protection for multimedia data based on labeling techniques", *Proc. 17th Symposium on Information Theory in the Benelux*, Enschede, The Netherlands, May 1996.
22. M. Schneider and S.-F. Chang, "A content-based approach to image signature generation and authentication", *Proc. ICIP '96*, vol. III, pp. 227–230, 1996.
23. M. Swanson, B. Zhu, and A. H. Tewfik, "Data hiding for video in video", *Proc. ICIP '97*, vol. II, pp. 676–679, 1997.
24. B. Girod, "The information theoretical significance of spatial and temporal masking in video signals", *Proc. of the SPIE Human Vision, Visual Processing, and Digital Display*, vol. 1077, pp. 178–187, 1989.
25. F. Hartung and B. Girod, "Digital watermarking of raw and compressed video", *Proc. European EOS/SPIE Symposium on Advanced Imaging and Network Technologies*, Berlin, Germany, October 1996.
26. B. O. Comiskey and J. R. Smith, "Modulation and information hiding in images", *Proc. of the 1st Information Hiding Workshop*, edited by R. Anderson, vol. 1174, LNCS, Cambridge, 1996.
27. W. Philips, "Fast coding of arbitrarily shaped image segments using weakly separable bases", *Optical Engineering*, **35**(1), pp. 177–187.