SECURE SPREAD SPECTRUM WATERMARKING FOR IMAGES, AUDIO AND VIDEO

Ingemar J. Cox†, Joe Kilian†, Tom Leighton‡, and Talal Shamoon†

†NEC Research Institute, 4 Independence Way, Princeton, NJ 08540
Email: ingemar|joe|talal@research.nj.nec.com
‡ Mathematics Department and Laboratory for Computer Science
MIT, Cambridge, MA 02139. Email: ftl@math.mit.edu

ABSTRACT

We describe a digital watermarking method for use in audio, image, video and multimedia data. We argue that a watermark must be placed in perceptually significant components of a signal if it is to be robust to common signal distortions and malicious attack. However, it is well known that modification of these components can lead to perceptual degradation of the signal. To avoid this, we propose to insert a watermark into the spectral components of the data using techniques analogous to spread sprectrum communications, hiding a narrow band signal in a wideband channel that is the data. The watermark is difficult for an attacker to remove, even when several individuals conspire together with independently watermarked copies of the data. It is also robust to common signal and geometric distortions such as digital-to-analog and analog-to-digital conversion, resampling, quantization, dithering, compression, rotation, translation, cropping and scaling. The same digital watermarking algorithm can be applied to all three media under consideration with only minor modifications, making it especially appropriate for multimedia products. Retrieval of the watermark unambiguously identifies the owner, and the watermark can be constructed to make counterfeiting almost impossible. We present experimental results to support these claims. A longer version of this document is available at: http://www.neci.nj.nec.com/tr/neci-tr-95-10.ps.

1. INTRODUCTION

The proliferation of digitized media (audio, image and video) is creating a pressing need for copyright enforcement schemes that protect copyright ownership. A digital watermark is a visible, or preferably invisible, identification code that is permanently embedded in the data, that is, it remains present within the data after any decryption process. In the context of this work, data refers to audio (speech and music), images (photographs and graphics), and video (movies). It does not include ASCII representations of text, but does include text represented as an image.

In order to be effective, a watermark should be unobtrusive (perceptually invisible), robust (difficult to remove), resilient to common signal processing and geometric distortions and intentional attacks. Intentional attacks include forgery and attacks using one or more watermarked copies of a document. Finally, the watermark should be universal, i.e. applicable to all three media under consideration, and the watermark should unambiguously identify its owner.

We believe that in order for the watermark to be robust, it should be placed in perceptually *significant* regions of the data despite the risk of potential fidelity distortions. Conversely, if the watermark is placed in perceptually insignificant regions, it is easily removable, either intentionally or unintentionally by, for example, signal compression techniques that implicitly recognize that perceptually weak components of a signal need not be represented. We have also found that resistance to multiple-copy (collusion) attacks is highly dependent on how the watermark is generated. To combat these attacks, we propose a normally distributed watermark that is shown to be robust in this sense.

Previous digital watermarking techniques [1–8, 10–14] are often not robust, particularly to collusion attacks. One reason for these weaknesses is that previous methods have not explicitly identified the perceptually most significant components of a signal as the destination for the watermark. In fact, it is often the case that the perceptually significant regions are explicitly avoided. The reason for this is obvious – modification of perceptually significant components of a signal results in perceptual distortions much earlier than if the modifications are applied to perceptually insignificant regions. The perceptually significant regions of a signal may vary depending on the particular media (audio, image or video) at hand, and even within a given media. For example, it is well known that the human visual system is tuned to certain spatial frequencies and to particular spatial characteristics such as line and corner features. Consequently, many watermarking schemes that focus on different perceptually significant phenomena are potentially possible. In this paper, we focus on perceptually significant *spectral* components of a signal.

Section 2 outlines our watermarking procedure. The experiments described in Section 3 strongly suggest that our system is robust to a variety of common signal distortions and displays strong immunity to a variety of attacks. Finally, Section 4 discusses possible weaknesses and enhancements to the system.

Authors appear in alphabetical order.

[0]Based on "A Secure, Imperceptable yet Perceptually Salient, Spread Spectrum Watermark for Multimedia" that appeared in Proc. IEEE Southcon96, June 1996. Copyright IEEE.

0-7803-3258-X/96/$5.00 © 1996 IEEE

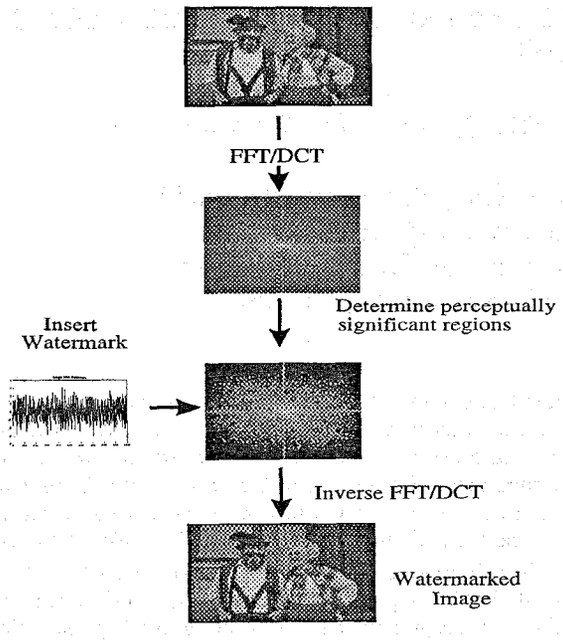

Figure 1: Stages of watermark insertion procedure.

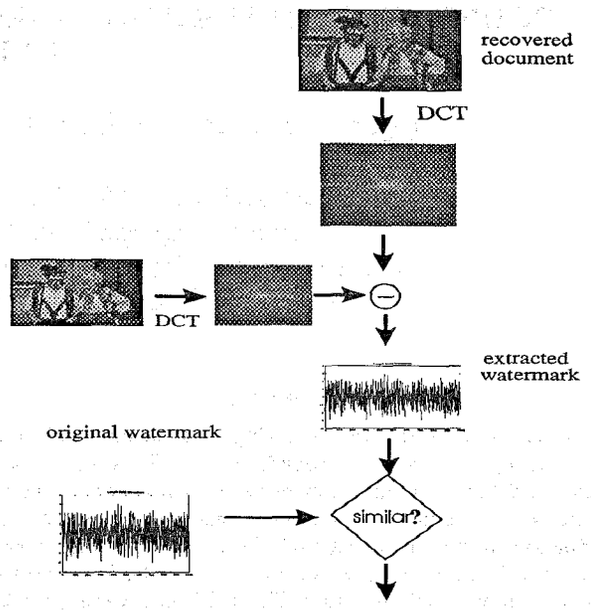

Figure 2: Stages of watermark extraction procedure.

2. WATERMARKING

Figures 1 and 2 outline the steps to insert and extract a watermark. Rather than encode the watermark into the least significant components of the data, we originally conceived our approach by analogy to spread spectrum communications [9]. In spread spectrum communications, one transmits a narrowband signal over a much larger bandwidth such that the signal energy present in any single frequency is undetectable. We place a watermark into a set of (possibly all) frequency components that are perceptually significant. The duality between the frequency and time/pixel domains causes the watermark to be spread over all pixels. Furthermore, since the watermark consists of a relatively weak noise signal in the frequency domain, it is virtually undectable to an attacker in that domain too. However, the watermark verification process knows the location and content of the watermark making it possible to concentrate these many weak signals into a single signal with high signal-to-noise ratio. The straightforward way to destroy such a watermark is to add high amplitude noise to *all* frequency bins that may contain watermark signal. In short, spreading the watermark throughout the spectrum of an image ensures a large measure of security against unintentional or intentional attack: First, the location of the watermark is not obvious. Furthermore, frequency regions are selected in a fashion that ensures severe degradation of the original data following any attack on the watermark.

In its most basic implementation, our watermark consists of a sequence of real numbers $X = x_1, \ldots, x_n$. In practice, we create a watermark where each value x_i is chosen independently according to $N(0,1)$ (where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2). The choice of n dictates the degree to which the watermark is spread out among the relevant components of the image. In general, as the number of altered components are increased the extent to which they must be altered decreases.

When we insert the watermark, X, into the image, V, to obtain V' we specify a scaling parameter α which determines the extent to which X alters V, so that $v_i' = v_i(1+\alpha x_i)$, where v_i denote perceptually significant spectral components of the image.[1]

Given V and a possibly distorted watermarked image V^*, we can extract a possibly distorted watermark X^* by essentially reversing the steps used to insert X into V. If V^* differs from V' (through unintentional distortion or active attack) it is highly unlikely that the extracted mark X^* will be identical to the original watermark X. We measure the similarity of X and X^* by $\mathsf{sim}(X, X^*) = \frac{X^* \cdot X}{\sqrt{X^* \cdot X^*}}$.

Our choice of a normally distributed watermark is motivated by the resistance that this distribution displays in the face of multiple-document attacks that use t multiple watermarked copies $D_1', \ldots, D_t'$ of document D to produce an unwatermarked document D^*. We note that most watermarking schemes proposed seem quite vulnerable to such attacks.

Our use of continuous valued watermarks appears to give greater resilience to such attacks. Interestingly, we have experimentally determined that if one chooses the x_i uniformly over some range, then one can remove the watermark using only 5 documents. We believe a Gaussian distribution is somewhat stronger.

Using a probabilistic analysis, it can be shown that any attack on Gaussian watermarks must make use of $\Omega(\sqrt{n/\ln n})$ watermarks in order to have any chance of destroying the watermark. (This is provided only that the original image

[1]For the experiments of Section 3, $\alpha = 0.1$.

Figure 3: "Bavarian Couple" courtesy of Corel Stock Photo Library.

Figure 4: Watermarked version of "Bavarian Couple".

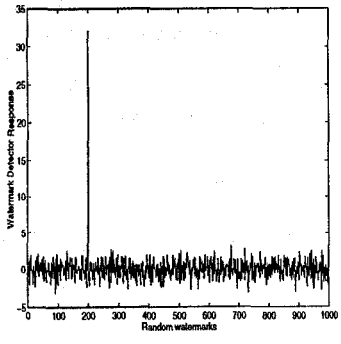

Figure 5: Watermark detector response to 1000 randomly generated watermarks. Only the watermark to which the detector was set to respond matches the one in Figure (4).

Figure 6: JPEG encoded version of "Bavarian Couple" with 5% quality and 0% smoothing.

Figure 7: Dithered version of "Bavarian Couple".

Figure 8: Clipped version of JPEG encoded (10% quality, 0% smoothing) "Bavarian Couple".

Figure 9: Printed, xeroxed, scanned and rescaled image of "Bavarian Couple".

to be protected comes from a Gaussian distribution and that the second moment of the deviation of the new image from the original is small compared to n.) Hence, Gaussian watermarks are better than uniform watermarks, particularly when n is large.

3. EXPERIMENTAL RESULTS

In order to evaluate the proposed digital watermark, we first took the "Bavarian Couple"[2] of Figure (3) and produced the watermarked version of Figure (4).

Figure (5) shows the response of the watermark detector to 1000 randomly generated watermarks of which only one matches the watermark that is present. The positive response due to the correct watermark is very much stronger that the response to incorrect watermarks, suggesting that the algorithm has very low false-positive (and false negative) response rates.

Next, we evaluated the performance of the watermark to common signal distortions such as scaling in which 75% of the data is removed,[3] JPEG encoding (Figure (6)) with a quality factor or 5%, dithering (Figure (7)), clipping (Figure (8)) in which only the central quarter of a JPEG image (quality factor 10%) is retained,[4] and printing, photocopying and subsequent re-scanning and scaling (Figure 9). Despite very severe distortions and loss of data, the detector responses were 13.4, 13.9, 10.5,[5] 10.6 and 7.0[6] respectively. These responses are well above random chance levels and indicate with a very high degree of certainty that the watermark is present.

Next, we considered resilience to malicious attacks on a watermarked image. First, we watermarked the image five successive times (that is, five distinct watermarks are were inserted into an image). This may be considered another form of attack in which it is clear that significant image degradation eventually occurs as the process is repeated. This attack is equivalent to adding noise to the frequency bins containing the watermark. Interestingly, though the response of the detector is reduced to approximately 14, all five watermarks are clearly identifiable, demonstrating that the method is robust to successive watermarking, i.e. additive noise attacks.

In a similar experiment, we took five separately watermarked images and averaged them together in order to

[2]The common test image "Lenna" was originally used in our experiments and similar results were obtained. However, Playboy Inc. refused to grant copyright permission for electronic distribution.

[3]In order to recover the watermark, the quarter-sized image was re-scaled to its original dimensions.

[4]In order to extract the watermark from this image, the missing portions of the image were replaced with portions from the original **unwatermarked** image.

[5]We achieved this response with a more robust version of the extraction process that removed any non-zero mean from the extracted watermark.

[6]We achieved this response with a more robust version of the extraction process in which only the sign of the elements of the watermark are used.

simulate a simple collusion attack. Once again, the five watermarks are clearly identifiable, demonstrating that the method is robust to simple collusion based on averaging of a few images. Of course, given a sufficient number of independently watermarked images, averaging can destroy the watermark. However, the number of images needed may be large.

4. CONCLUSION

A need for electronic watermarking is developing as electronic distribution of copyright material becomes more prevalent. Above, we outlined the necessary characteristics of such a watermark. These are: fidelity preservation, robustness to common signal and geometric processing operations, robustness to attack, and applicability to audio, image and video data. To meet these requirements, we proposed a watermark whose structure consists of 1000 randomly generated numbers with a normal distribution having zero mean and unity variance. We rejected a binary or a uniformly distributed watermark on the grounds that they are much less robust to attacks based on collusion of several independently watermarked copies of an image. The length of the watermark is variable and can be adjusted to suit the characteristics of the data. For example, longer watermarks might be used for an image that is especially sensitive to large modifications of its spectral coefficients, thus requiring weaker scaling factors for individual components.

The watermark is then placed in the perceptually *most* significant components of the image spectrum. This ensures that the watermark remains within the image even after common signal and geometric distortions. Modification of these spectral components results in severe image degradation long before the watermark itself is destroyed. Of course, to insert the watermark, it is necessary to alter these very same coefficients. However, each modification can be extremely small and, in a manner similar to spread spectrum communication, a strong narrowband watermark may be distributed over a much broader image (channel) spectrum. Conceptually, detection of the watermark then proceeds by adding all of these very small signals, and concentrating them once more into a signal with high signal-to-noise ratio. Because the magnitude of the watermark at each location is only known to the copyright holder, an attacker would have to add much more noise energy to each spectral coefficient in order to be sufficiently confident of removing the watermark. However, this process would destroy the image.

In our experiments, we added the watermark to the image by modifying 1000 of the more perceptually significant components of the image spectrum. More specifically, the 1000 largest coefficients of the DCT (excluding the DC term) were used. Further refinement of the method would identify perceptually significant components based on an analysis of the image and the human perceptual system and might also include additional considerations regarding the relative predictability of a frequency based on its neighbors. The latter property is important to consider in order to minimize any attack based on a statistical analysis of frequency spectra that attempts to replace components with their maximul likelihood estimate, for example.

It was shown that the algorithm can extract a reliable copy of the watermark from imagery that has been significantly degraded through several common geometric and signal processing procedures. These include, zooming (low pass filtering), cropping, lossy JPEG encoding, dithering, printing, photocopying and subsequent rescanning.

More experimental work needs to be performed to validate these results over a wide class of data. Application of the method to color images should be straightforward though robustness to certain color image processing procedures should be investigated. Similarly, our ideas may be applicable to text images and line-art. However, more work is needed to address the binary nature of these images and their highly structured spectral distributions. Furthermore, application of the watermarking method to audio and video data straightforwardly, although one must pay attention to the time varying nature of these data. A more sophisticated watermark verification process may also be possible using methods developed for spread spectrum communications.

Acknowledgements

This work was in part influenced by the work of Larry O'Gorman and colleagues at AT&T Bell Laboratories on the watermarking of text. We also thank Harold Stone for advice on image transforms and Sebastien Roy for testing the robustness of the watermark.

5. REFERENCES

[1] E. H. Adelson. Digital signal encoding and decoding apparatus. Technical Report 4,939,515, United States Patent, 1990.

[2] W. Bender, D. Gruhl, and N. Morimoto. Techniques for data hiding. In *Proc. of SPIE*, volume 2420, page 40, February 1995.

[3] J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman. Electronic marking and identification techniques to discourage document copying. In *Proc. of Infocom'94*, pages 1278–1287, 1994.

[4] G. Caronni. Assuring ownership rights for digital images. In *Proc. Reliable IT Systems, VIS'95*. Vieweg Publishing Company, 1995.

[5] E. Koch, J. Rindfrey, and J. Zhao. Copyright protection for multimedia data. In *Proc. of the Int. Conf. on Digital Media and Electronic Publishing*, 1994.

[6] E. Koch and Z. Zhao. Towards robust and hidden image copyright labeling. In *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, June 1995.

[7] B. M. Macq and J-J Quisquater. Cryptology for digital tv broadcasting. *Proc. of the IEEE*, 83(6):944–957, 1995.

[8] K. Matsui and K. Tanaka. Video-steganography. In *IMA Intellectual Property Project Proceedings*, volume 1, pages 187–206, 1994.

[9] R. L. Pickholtz, D. L. Schilling, and L. B. Millstein. Theory of spread spectrum communications - a tutorial. *IEEE Trans. on Communications*, pages 855–884, 1982.

[10] G. B. Rhoads. Indentification/authentication coding method and apparatus. *World Intellectual Property Organization*, IPO WO 95/14289, 1995.

[11] W. F. Schreiber, A. E. Lippman, E. H. Adelson, and A. N. Netravali. Receiver-compatible enhanced definition television system. Technical Report 5,010,405, United States Patent, 1991.

[12] K. Tanaka, Y. Nakamura, and K. Matsui. Embedding secret information into a dithered multi-level image. In *Proc, 1990 IEEE Military Communications Conference*, pages 216–220, 1990.

[13] L. F. Turner. Digital data security system. Patent IPN WO 89/08915, 1989.

[14] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne. A digital watermark. In *Int. Conf. on Image Processing*, volume 2, pages 86–90. IEEE, 1994.