IN *PROC. IEEE INT. CONF. ON IMAGE PROC. '96*, VOL. III, PP. 211-214

TRANSPARENT ROBUST IMAGE WATERMARKING

Mitchell D. Swanson, Bin Zhu, and Ahmed H. Tewfik
Department of Electrical Engineering
University of Minnesota, Minneapolis, MN 55455 USA
email: mswanson, binzhu, tewfik@ee.umn.edu

ABSTRACT

We propose a watermarking scheme to hide copyright information in an image. The scheme employs visual masking to guarantee that the embedded watermark is invisible and to maximize the robustness of the hidden data. The watermark is constructed for arbitrary image blocks by filtering a pseudo-noise sequence (author id) with a filter that approximates the frequency masking characteristics of the visual system. The noise-like watermark is statistically invisible to deter unauthorized removal. Experimental results show that the watermark is robust to several distortions including white and colored noises, JPEG coding at different qualities, and cropping.

1. INTRODUCTION

Digital images facilitate efficient distribution, reproduction, and manipulation over networked information systems. However, these efficiencies also increase the problems associated with copyright enforcement. To address this issue, digital watermarks (i.e., author signatures) are under investigation. Watermarking is the process of encoding hidden copyright information in an image by making small modifications to its pixels. Unlike encryption, watermarking does not restrict access to an image. Watermarking is employed to provide solid proof of ownership. To be effective, the watermark must be [1, 2]: **perceptually invisible** within the host media; **statistically invisible** to thwart unauthorized removal; **readily extracted** by the image owner; and **robust** to incidental and intended signal distortions incurred by the host image, e.g., filtering, compression, re-sampling, re-touching, cropping, etc.

In this paper, we introduce a novel watermarking scheme for images which exploits the human visual system (HVS) to *guarantee that the embedded watermark is imperceptible*. Our watermark is generated by filtering a pseudo-noise sequence (author id) with a filter that approximates the frequency masking characteristics of the HVS. The image watermark is constructed by computing watermarks for individual image blocks. The blocks may be $n \times m$ or may be defined in terms of image objects/regions. This helps deter pirating of image objects. Furthermore, the noise-like watermark is statistically invisible. We include experimental results which indicate that the watermark is readily extracted and *robust* to common signal processing operations.

This work was supported by AFOSR under grant AF/F49620-94-1-0461. Patent pending, Media Science, Inc., 1996.

2. PREVIOUS WORK

The most common watermarking approaches modify the least significant bits (LSB) of an image based on the assumption that the LSB data are insignificant. Two LSB techniques are described in [3]. The first replaces the LSB of the image with a pseudo-noise (PN) sequence, while the second adds a PN sequence to the LSB of the data. Another LSB data hiding method called "Patchwork" [1] chooses n pairs (a_i, b_i) of points in an image and increases the brightness of a_i by one unit while simultaneously decreasing the brightness of b_i. Several executable software packages (e.g., Stego, S-Tools) based on LSB approaches are also available. However, any approach which only modifies the LSB data is *highly sensitive to noise* and is easily destroyed. Furthermore, image quality may be degraded by the watermark. Other watermarking approaches include [4, 5, 6].

A method similar to ours is presented in [7], where the authors hide data by adding fixed amplitude pseudo-noise to the image. The approach presented here employs masking to vary the amplitude of the hidden data. Specifically, the tolerable error levels obtained using masking provide us with the maximum amount the image data may change. Pseudo-noise techniques are also used in [2], where the N largest frequency components of an image are modified by Gaussian noise. However, the scheme only modifies a subset of the frequency components and does not take into account the HVS. The watermark we propose here embeds the *max-*

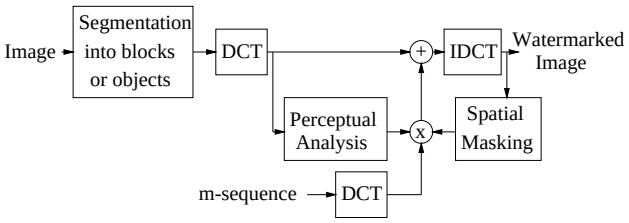

Figure 1: Diagram of new watermarking technique.

imum amount of information throughout the spectrum. Since more data is embedded, this scheme is *guaranteed* to be more robust to modifications than a technique which only modifies a subset of the image data.

3. WATERMARK GENERATION

In Fig. 1, we show our watermarking technique. The initial step consists of segmenting the image into blocks. Using a traditional approach, the blocks may be $n \times n$ (e.g., 8×8 like JPEG). An option at this stage is to segment the image into blocks of *objects and texture regions*. In either case, blocking the image adds detection robustness to cropping and localized signal processing operations. Upon applying a discrete cosine transform (DCT) to each block, a frequency mask is computed for each block in a manner similar to low bit rate coding algorithms [8]. The resulting perceptual mask is scaled and multiplied by the DCT of a maximal length pseudo-noise sequence (author id). Note that a different pseudo-noise sequence is used for each image block. This *watermark* is then added to the corresponding DCT block. The watermarked image is obtained by assembling the inverse DCT's of each block. Spatial masking is used to verify that the watermark is invisible and to control the scaling factor.

Pseudo-noise (PN) sequences form the signatures in our watermarking scheme because of their noise-like characteristics, resistance to interference, and their good auto-correlation properties. PN sequences are periodic noise-like binary sequences generated by length m linear shift registers [9]. Furthermore, the period N autocorrelation function has peaks equal to 1 at 0, N, 2N, etc., and is approximately equal to $1/N$, elsewhere. These periodic peaks allow the author to *synchronize* with the embedded watermark during the detection process.

Visual masking models are used to modify the author signature. Visual masking refers to a situation where a signal raises the visual threshold for other signals around it. Both frequency and spatial masking are employed by our watermarking scheme. Our frequency masking model is based on the observation that a masking grating raises the visual threshold for signal gratings around the masking frequency [10]. The model we use [11] expresses the contrast threshold at frequency f as a function of f, the masking frequency f_m and the masking contrast c_m:

$$c(f, f_m) = c_0(f) \cdot Max\{1, [k(f/f_m)c_m]^{\alpha}\},$$

where $c_0(f)$ is the detection threshold at frequency f. To find the contrast threshold $c(f)$ at a frequency f in an image, we first use the DCT to transform the image into the frequency domain and find the contrast at each frequency. Then we use a summation rule of the form $c(f) = [\sum_{f_m} c(f, f_m)^{\beta}]^{1/\beta}$. If the contrast error at f is less than $c(f)$, the model predicts that the error is invisible to human eyes.

After adding the watermark in the frequency domain, spatial masking is checked. The spatial model is used to verify that the watermark designed with the frequency masking model is invisible for local spatial regions. The model used here is similar to our image coding model [11] which gives the tolerable error level for each coefficient. Each watermark coefficient is compared with the tolerable error level obtained to assure that it is invisible. A visible watermark is rescaled via a weighting factor.

4. WATERMARK DETECTION

The watermark should be extractable even if common signal processing operations are applied to the host image. This is particularly true in the case of deliberate unauthorized attempts to remove it. For example, a pirate may attempt to add noise, filter, code, re-scale, etc., an image in an attempt to destroy the watermark. As the embedded watermark is noise-like and its location (based on multiple blocks) is unknown, a pirate has insufficient knowledge to directly remove the watermark. Furthermore, a different m-sequence is used for each block to further reduce unauthorized watermark removal by cross-correlation. Therefore, any destruction attempts are done blindly. Unlike other users, the author has copies of the original signal S and the signature. Detection of the watermark is accomplished via hypotheses testing:

$$\begin{array}{lll} \mathbf{H}_0: & X = R - S = N & \text{(No watermark)} \\ \mathbf{H}_1: & X = R - S = W' + N & \text{(Watermark)} \end{array}$$

where R is the potentially pirated signal, W' is the potentially modified watermark, and N is noise. The correct hypothesis is obtained by applying a correlating detector on X with W and comparing with a threshold. In some cases, e.g., spatial rescaling, a generalized likelihood ratio test must be applied.

5. EXPERIMENTAL RESULTS

To illustrate our watermarking technique, the 256×256 grayscale (8-bit) image shown in Fig. 2 was segmented into 8×8 blocks and watermarked. The watermarked image is shown in Fig. 3. The images appear identical.

We tested the robustness of the watermark to several degradations. To model perceptual coding techniques, we corrupted the watermark with *worst case colored noise which follows the image mask.* We generated colored noise with SNR of 10dB and added it to the image with (hypothesis $\mathbf{H}_1$) and without ($\mathbf{H}_0$) the watermark. The watermarked image with colored noise is shown in Fig. 4. The hypothesis test was applied to each block in the image. This testing process was repeated 250 times. The normalized correlation coefficients indicate easy discrimination between the hypotheses as shown in Fig. 6. In particular, the correlation coefficient for the image with and without the watermark was approximately 0 and 1 respectively.

To further degrade the watermark, we applied JPEG coding at 0.38 bpp (10% quality, c.f. Fig. 5) and 1.32 bpp (50% quality) to each of the images *already corrupted* with colored noise. Note that the image is significantly degraded at 0.38 bpp, yet the watermark is still easily detected as shown in Fig. 6. It is unlikely a pirate would do so much irreparable damage to the image. Setting a decision threshold of 0.15 results in no decision errors. In Fig. 7, we show the result of applying JPEG coding at different quality factors to the noisy image with and without the watermark. It is clear that the correlation coefficient values for the two hypotheses are well separated for all JPEG coding qualities. We also investigated cropping robustness by determining the minimum number of image blocks required to make a confident decision on whether the watermark is present in an image ($P_D = 1$ and $P_F < 10^{-4}$). Each noisy image in the above tests was *randomly cropped* and tested. The results indicate that only 0.4%, 2%, and 15% of the image is needed for a confident decision when coded at 8 bpp, 1.32 bpp, and 0.38 bpp.

For comparison, we implemented the system described in [2]. For JPEG coding at 10%, we obtained (unnormalized) correlation coefficients using their system of 5.09 and 2.34 for the same test image with and without the watermark, respectively. The ratio is significantly smaller than ours. While testing their system, we were unable to reproduce the detection results as claimed in [2]. This may be the result of special post-processing operations they implement. The robustness of our watermarking scheme to re-sampling, multiple watermarking, vector quantization, and other distortions is described in [12].

6. REFERENCES

[1] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for Data Hiding." Tech. Rep., MIT Media Lab, 1994.

[2] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia." Tech. Rep. 95-10, NEC Research Institute, 1995.

[3] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A Digital Watermark," in *Proc. 1994 IEEE Int. Conf. on Image Proc.*, vol. II, (Austin, TX), pp. 86–90, 1994.

[4] I. Pitas and T. Kaskalis, "Applying signatures on digital images," in *Nonlinear Signal Processing Workshop, Thessaloniki, Greece*, pp. 460–463, 1995.

[5] E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling," in *Nonlinear Signal Processing Workshop, Thessaloniki, Greece*, pp. 452–455, 1995.

[6] O. Bruyndonckx, J.-J. Quisquater, and B. Macq, "Spatial method for copyright labeling of digital images," in *Nonlinear Signal Processing Workshop, Thessaloniki, Greece*, pp. 456–459, 1995.

[7] J. R. Smith and B. O. Comiskey, "Modulation and Information Hiding in Images." to appear *1996 Workshop on Information Hiding*, University of Cambridge, UK.

[8] N. Jayant, J. Johnston, and R. Safranek, "Signal Compression Based on Models of Human Perception," *Proc. of the IEEE*, vol. 81, pp. 1385–1422, oct 1993.

[9] S. Haykin, *Communication Systems, 3rd Edition.* New York, NY: John Wiley and Sons, 1994.

[10] G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *J. Opt. Soc. Am.*, vol. 70, no. 12, pp. 1458–1471, 1980.

[11] B. Zhu, A. Tewfik, and O. Gerek, "Low Bit Rate Near-Transparent Image Coding," in *Proc. of the SPIE Int. Conf. on Wavelet Apps. for Dual Use*, vol. 2491, (Orlando, FL), pp. 173–184, 1995.

[12] A. H. Tewfik, M. D. Swanson, B. Zhu, K. Hamdy, and L. Boney, "Transparent Robust Watermarking for Images and Audio." To be submitted *IEEE Trans. on Signal Proc.*, 1996.

Figure 2: Original 256x256 grayscale image.

Figure 3: Watermarked image using 8×8 blocks.

Figure 4: Watermarked image with colored noise (SNR 10dB).

Figure 5: JPEG coded version of watermarked image at 0.38bpp (10% quality).

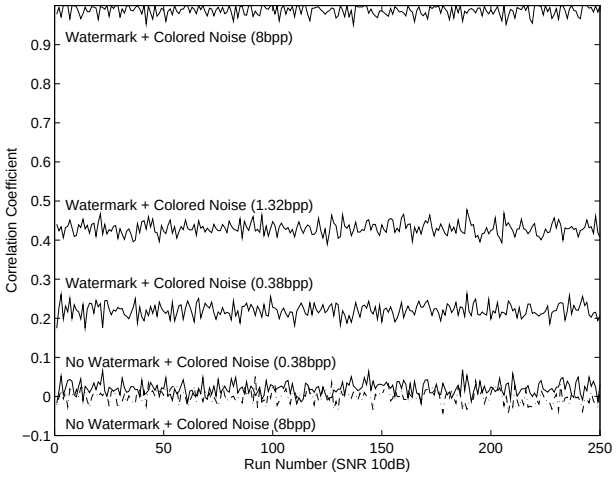

Figure 6: Watermark detection after adding colored noise with and without JPEG coding.

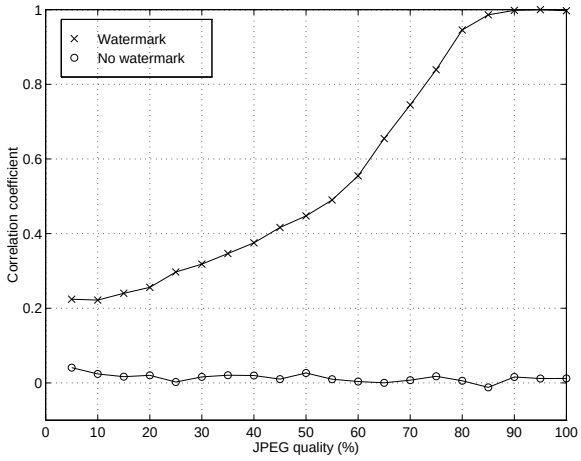

Figure 7: Watermark detection after JPEG coding at different qualities.