Transparent Robust Authentication and Distortion Measurement Technique for Images *

Bin Zhu, Mitchell D. Swanson, and Ahmed H. Tewfik
Department of Electrical Engineering, University of Minnesota
e-mail: binzhu, mswanson, tewfik@ee.umn.edu

ABSTRACT

We propose a novel scheme to embed an invisible signature into an image to check image integrity and measure its distortion. The technique is based on the pseudo noise sequences and visual masking effects. The values of an image are modified by a pseudo noise signature which is shaped by the perceptual thresholds from masking effects. The method is robust and can gauge errors accurately up to half of the perceptual thresholds. It also readily identifies large image distortion. Experimental results after applying JPEG and white noise to the image are also reported.

1. INTRODUCTION

We introduce a robust scheme to embed into an image an invisible pattern which is used to verify image integrity. Furthermore, the technique provides a distortion measurement indicating the amount of damage incurred upon an image. By exploiting limitations of the human visual system, the energy of the embedded pattern is maximized for robustness and yet is guaranteed to be perceptually invisible.

Using masking models, the tolerable error level at each pixel is obtained. We replace those values below the tolerable error level by a pseudo-noise sequence which is shaped by the tolerable errors. This shaping procedure hides high signature energy in perceptually important features such as edges. As a result, it increases to the robustness of the method. To check the integrity or measure distortion of an image, a receiver requires the key to regenerate the pseudo noise sequence. The pseudo noise is further shaped by the estimated perceptual thresholds from the received image. The receiver compares the received values against the regenerated signature, and thus can determine the integrity of the received image. If the image is manipulated by some lossy operations, such as coding, channel transmission errors, re-touching, etc, the distance between the received values and the regenerated signature can be used to measure the distortion. It can gauge the distortion at a pixel up to half of the perceptual threshold of the pixel. Distortion may be computed for the entire image or for local image regions. We assume here that the received image has limited perceptual distortions, since large perceptual modifications are readily apparent to the observer.

While current information systems provide efficient access to the data, they increase the problems associated with verifying data integrity. Image authentication may be used to address this problem. In a medical environment, for example, a doctor may want to verify whether a displayed image is authentic or re-touched, smoothed, etc. The scheme we propose here measures the amount of modification to the displayed image. An image, or areas within the image, exceeding a certain level of distortion may be discarded.

Our approach embeds the distortion measure directly into the image rather than an image header or separate file. As a result, we avoid problems associated with changing file formats and with storage and transmission of multiple files. The proposed technique is related to our image watermarking technique [1]. It differs from the watermarking techniques in that the intended audience must be able to extract the embedded signature without knowledge of the original image. Furthermore, it has to measure the distortion made to an image. The pattern we embed in the image must be perceptually invisible within the host media and robust to manipulation and signal processing operations on the image, e.g., filtering, compression, noise, etc.

Several techniques used for image verification have been proposed. Most approaches compute a discriminant (e.g., hash function [2]) and place the data in the image header or a separate file. In the approaches which embed verification into the image, most hide a pattern in the least significant bits (LSB) of an image based on the assumption that the LSB data are perceptually insignificant. However, any approach which only modifies the LSB data is highly sensitive to noise and is eas-

*This work was supported by AFOSR under grant AF/F49620-94-1-0461.

0-7803-3629-1/96/$5.00 @ 1996 IEEE

ily destroyed. Our technique uses visual masking to maximize the energy in the pattern, thereby increasing pattern robustness and the gauging range of distortion.

2. SPATIAL AND FREQUENCY DOMAIN MASKING

Spatial or frequency domain masking effects are used to shape the pseudo noise sequence to maximize the signature energy while maintaining the signature perceptually invisible. Visual masking refers to the psychophysical phenomena that a signal raises the perceptual thresholds of other signals around it. The masking values are obtained by the threshold visual masking models that were used in high quality, low bit rate image coding [3].

In the frequency domain, a grating signal raises the perceptual thresholds of other gratings whose frequencies are close to the masking frequency [4]. If the masking frequency is f_m, and the masking contrast is c_m, the contrast threshold at f due to the masker f_m is modeled as

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

where $c_0(f)$ is the detection threshold at frequency f. Since we use discrete cosine transform (DCT) to transform an image into frequency domain, the detection threshold $c_0(f)$ is corrected by multiplying the detection threshold obtained in psychophysics with a factor given by Nill [5]. The contrast threshold $c(f)$ at frequency f is obtained by a summation rule

$$c(f) = [\sum_{f_m \in \mathcal{S}(f)} c(f, f_m)^{\beta}]^{1/\beta}, \quad (2)$$

where the set $\mathcal{S}(f)$ is a range of frequencies at the neighborhood of f. If the contrast error at frequency f is less than $c(f)$, the model predicts that the error is perceptually invisible.

Similar masking effect exists around an edge in spatial domain. The model of spatial masking is a modified model from the threshold vision model proposed by Girod [6]. In the model, the processing channel is linearized under the assumption that the perceptual error at threshold vision is small. The perceptual threshold at each pixel is found reversely from the last stage to the first stage. For details, readers are referred to [3].

3. EMBEDDING SIGNATURE DESIGN

A signature embedded in an image can be generated either in the spatial domain or frequency domain. It is obtained by multiplying a pseudo noise sequence by the visual masking values obtained from either spatial masking model or frequency masking model. The operations are almost the same in both domains. We shall discuss the frequency domain design method in detail.

To design a signature in the frequency domain, we transform an image into the DCT domain first. Let $P(i,j)$ denote the value of frequency bin (i,j). Then we use the frequency masking values $M(i,j)$ at each frequency bin (i,j). Let $r(i,j)$ be the noise value generated by a pseudo noise generator which generates uniformly distributed white noise in the range of $(0,1)$. The frequency value $P(i,j)$ of the image is modified to $P_s(i,j)$ which is given by:

$$P_s(i,j) = M(i,j) \cdot \{\lfloor \frac{P(i,j)}{M(i,j)} \rfloor + sgn(P(i,j))r(i,j)\}, \quad (3)$$

where $\lfloor \cdot \rfloor$ rounds towards 0, and $sgn(x)$ is sign of x, defined as:

$$sgn(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ -1, & \text{if } x < 0. \end{cases} \quad (4)$$

It is easy to check that the error introduced by the above operation is smaller than the perceptual threshold, i.e., $|P_s(i,j) - P(i,j)| \leq M(i,j)$. Thus the signature is perceptually invisible.

4. EXTRACTING SIGNATURE AND CALCULATING DISTORTION

We assume that an intended receiver knows that a signature pattern is embedded in the received image. If not, we can use a similarity measure similar to that we use for image watermarking [1] to determine first if a given signature is embedded in the received image or not. To check the authentication and distortion of the received image, an intended receiver must be provided with the private key to regenerate the pseudo noise sequence embedded in the image. Since the pseudo noise is like white noise, a receiver without the key is unable to decode the embedded signatures.

The receiver uses the provided key to regenerate the pseudo noise sequence $r(i,j)$. We use the DCT to transform the received image into frequency domain with value $P'_s(i,j)$ at frequency bin (i,j). Assume the masking value estimated from $P'_s(i,j)$ is $M'(i,j)$. The error at (i,j) is estimated by the following equation:

$$\hat{e} = P'_s - M' \cdot \{sgn(P'_s)r + \lfloor \frac{P'_s}{M'} - (r - \frac{1}{2})sgn(P'_s) \rfloor\}, \quad (5)$$

where all the values are at the same frequency bin (i,j).

If $|P(i,j)| < M(i,j)$ and $r(i,j)$ is small, a small error can change the sign of $sgn(P'_s(i,j))$ which may result in large estimation error in Eq. 5. To avoid such a problem, when both P and r are small, r is multiplied by a factor to raise it to just above 0.5. The intended receiver can correct this situation if the error is not too large.

To estimate the performance of the error estimation, we assume that $P'_s(i,j) = P_s(i,j) + e(i,j)$ where $e(i,j)$ is the error produced by some lossy operations on the image. We have the following theorem:

Theorem 1 *Assume that the error $e(i,j)$ is small (i.e., $|e(i,j)| < \frac{M(i,j)}{2}$), and that for small errors the masking model produces the same masking values from the original and distorted images (i.e., $M'(i,j) = M(i,j)$), then*

$$\hat{e}(i,j) = e(i,j). \tag{6}$$

In other words, the error estimation given by the proposed method is accurate under the above conditions.

Proof. First we note that under the assumptions of the theorem, $sgn(P'_s) = sgn(P_s) = sgn(P)$. Substitution of $P'_s = P_s + e$ to Eq. 5 yields

$$\hat{e} = P_s + e - sgn(P)rM - M\lfloor \frac{P_s - sgn(P)rM}{M} + \frac{1/2sgn(P)M + e}{M} \rfloor.$$

From Eq. 3, $P_s - sgn(P)rM = \lfloor \frac{P}{M} \rfloor M$, the above equation can be simplified as

$$\begin{aligned} \hat{e} &= \lfloor \frac{P}{M} \rfloor M + e - M \lfloor \frac{P}{M} \rfloor \\ &= e, \end{aligned} \tag{7}$$

where we have used the fact that $|1/2sgn(P)M + e| < M$, and $\lfloor x \rfloor = x$ for an integer x. □

There are three factors which affect the accuracy of the estimated errors. The first is the robustness of the visual masking model. To accurately estimate errors, the masking model should give masking values based on the received image as close as possible to the actual masking values. The second factor is the errors incurred by some operations on the image. If the error at a frequency bin is larger than half of the masking value at that frequency bin, the estimated error is wrong. The third factor is that small errors result from inverse DCT and rounding to integers in the range from 0 to 255 for an 8 bit image.

The estimated error at each frequency bin can be used to find local or global distortions. A simple global distortion measurement is employed in this paper which is a weighted error at each frequency bin according to the masking values. More accurate distortion measurement can also be built from the estimated error at each frequency bin.

The same procedure can be applied to the spatial domain approach which uses the spatial domain masking model. The expressions are almost exactly the same as those for the frequency domain approach except that the results should be rounded to integers.

5. EXPERIMENTAL RESULTS

To illustrate the performance of the proposed approach, we have used both frequency and spatial domain design methods to design signatures to the 256 by 256 gray-scale (8-bit) image Lena shown in Fig. 1. The image with the embedded signature pattern using the frequency domain approach is shown in Fig. 2. When we tested both images on a Sun Sparc 5 monitor, we could not tell any difference between the two images.

Figure 1. The original 256X256 gray scale image Lena.

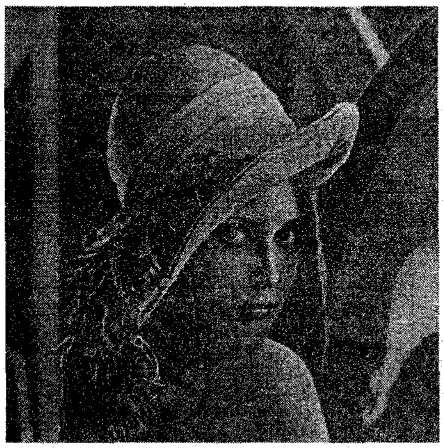

Figure 2. Image with embedded signature pattern.

We have applied JPEG and white noise to the images and used the proposed method to test its integrity and calculate the distortion of the resulting image. The estimated error for the JPEG processed image at different qualities is shown in

Fig. 3. The estimated error for the image with added white noise is shown in Fig. 4. The solid curves in both figures are the measured distortion given by the receiver while the dashed lines are for the ideal results if the original image is also known. Both figures were obtained from the frequency design approach. We note here that quality 100% for JPEG is in fact lossless coding.

As we can see from Figs. 3 and 4, the proposed method gives accurate results when the errors introduced to the image are small. The accuracy is reduced when the distortion grows larger. When there is no distortion, the measured distortion is small, but not 0. This is due to small errors introduced by the masking model as well as the inverse DCT and rounding off to integers in [0, 255]. From the results, we conclude that the frequency masking model is quite robust.

We have also used the spatial domain approach to design signatures. Due to the fact that the spatial masking model is sensitive to the noise introduced by the signature pattern, the results were not as good as those from the frequency domain approach. It results in errors relatively large even though the image has no distortion. With JPEG corrupted images, the result is shown by the solid curve in Fig 5, together with the ideal results (dashed curve). These results are normalized with respect to the distortion produced by JPEG compression at 0.95 quality. The results fits the ideal results reasonably well for small distortion.

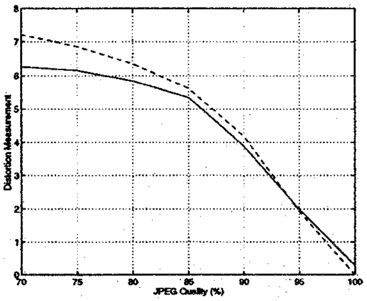

Figure 3. Distortion for JPEG processed image (freq. domain, see text)

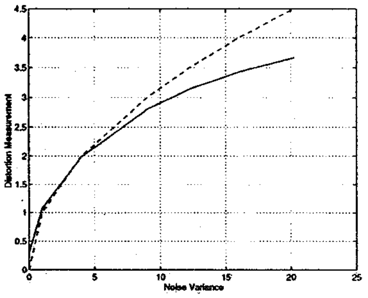

Figure 4. Distortion for image corrupted by white noise (freq. domain, see text).

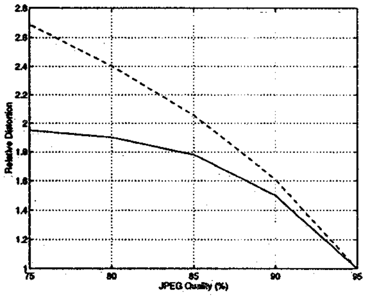

Figure 5. Normalized distortion for JPEG compressed images (spatial domain, see text).

6. CONCLUSION

In this paper, we have proposed a robust method to check the integrity of an image and to calculate the distortion if the image has incurred some lossy operations. It gauges image distortion accurately when the distortion is small. It also readily identifies large image distortion.

REFERENCES

[1] M. Swanson, B. Zhu, A. H. Tewfik, "Transparent Robust Image Watermarking," to appear *Proc. of the 1996 IEEE Int. Conf. on Image Processing*, September, 1996.

[2] J. P. Smith, "Authentication of Digital Medical Images with Digital Signature Technology," *Radiology*, Vol.194, No.3, p. 771-774, 1995.

[3] B. Zhu, A. H. Tewfik and Ö. N. Gerek, "Low Bitrate Near-Transparent Image Coding," *1995 SPIE Conf. on Wavelet Applications II*, Vol. 2491, PP. 173-184, 1995.

[4] G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *J. Opt. Soc. Am.*, Vol 70, No. 12, PP. 1458-1471, 1980.

[5] N. B. Nill, "A visual model weighted cosine transform for image compression and quality assessment," *IEEE Trans. Communications*, Vol. COM-33, No. 6, 1985.

[6] B. Girod, "The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals," *SPIE Vol. 1077 Human Vision, Visual Processing, and Digital Display*, pp. 178-187, 1989.