In IEEE Digital Signal Processing Workshop (DSP 96), Loen, Norway, pp. 37-40, September, 1996.

Robust Data Hiding for Images *

Mitchell D. Swanson *Bin Zhu* *Ahmed H. Tewfik*

Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455
e-mail: mswanson, binzhu, tewfik@ee.umn.edu

ABSTRACT

Data hiding is the process of encoding extra information in an image by making small modifications to its pixels. To be practical, the hidden data must be perceptually invisible yet robust to common signal processing operations. This paper introduces two schemes for hiding data in images. The techniques exploit perceptual masking properties to embed the data in an invisible manner. The first method employs spatial masking and data spreading to hide information by modifying image coefficients. The second method uses frequency masking to modify image spectral components. By using perceptual masking, we also increase robustness of the hidden information. Experimental results of data recovery after applying noise and JPEG coding to the hidden data are included.

1. INTRODUCTION

We introduce two robust schemes to hide information, e.g., labels, into an image by modifying perceptually irrelevant portions of image data. By exploiting the human visual system (HVS), our techniques embed a large amount of data into an image while guaranteeing that the hidden data are perceptually invisible. In particular, the data are hidden by modifying image coefficients according to masking levels based on the HVS. Information may be hidden throughout the image or confined to particular image objects and regions. The first scheme we introduce spreads the data to be hidden with a pseudo-noise sequence and then modifies them using spatial masking. The second data hiding scheme modifies the DCT coefficients of image blocks according to their frequency masking characteristics. In both schemes, masking characteristics are used to maintain high bit rates and robustness of the hidden data. We include experimental results which indicate the robustness of the data hiding techniques to common signal processing operations.

To be useful, the embedded data must be [1, 2]: *perceptually invisible* within the host media; *readily extracted* by its intended audience; *high bit-rate* for practical applications; and *robust* to manipulation and signal processing operations on the host image, e.g., filtering, re-sampling, compression, noise, cropping, etc. Hiding information in images may be used, e.g., to supplement an image with additional information, add copyright protection (i.e., digital watermarks), or verify image integrity. We consider here the generic problem of embedding supplementary information into image data with the goal of recovering the information bits *without* access to the original image. In such a scenario, the author of the hidden information supplies a public key which allows users to recover the hidden data. The hidden information may be text, audio, or image data. For example, text captions may be used to label faces and buildings in an image. A short audio clip may associate a train whistle with an image of a locomotive.

Note that embedding information *directly into image data* has several advantages over storing the data in an image header or a separate file. Specifically, header data are easily lost when the format of a file is changed or the image is cropped. Separate files also need to be transmitted when the image is transmitted and are difficult to maintain during cropping operations. Separate files and image headers also require additional storage. Hiding data directly into the image data resolves these problems.

2. PREVIOUS WORK

Several data hiding techniques have been proposed. The most common approaches modify the least significant bits (LSB) of an image based on the assumption that the LSB data are insignificant. In [3], two such techniques are described. 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. However, any approach which only modifies the LSB data is *highly sensitive to noise* and is easily destroyed. 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 [4]. A data hiding method similar to our frequency hiding method is proposed in [2], where the N largest frequency components of

---

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

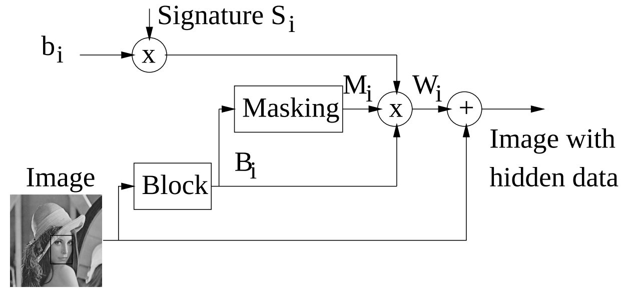

Figure 1. Diagram of spread spectrum data hiding technique.

an image are modified by a PN sequence. However, the scheme only modifies a subset of the frequency components and does not take into account the HVS. A method similar to our spatial hiding scheme is presented in [5], where the authors hide data in an image using spread spectrum techniques. Our approach employs spatial masking to *maximize* the amount of data hidden in the image.

3. SPATIAL AND FREQUENCY DOMAIN MASKING

We use masking models based on the HVS to ensure that the hidden data are perceptually invisible. Visual masking refers to a situation where a signal raises the visual threshold for other signals around it. Masking characteristics are used in high quality low bit rate coding algorithms to further reduce bit rates [6].

Our frequency masking model is based on the knowledge that a masking grating raises the visual threshold for signal gratings around the masking frequency [7]. The model we use [8] expresses the contrast threshold at frequency f as a function of f, the masking frequency f_m and the masking contrast c_m. 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. If the contrast error at f is less than $c(f)$, the model predicts that the error is invisible to human eyes.

Spatial masking refers to the situation that an edge raises the perceptual threshold around it. The model used here is similar to our image coding model [8] which is based on a model proposed by Girod [9]. In our approach, the upper channel of Girod's model is linearized under the assumption of small perceptual errors. The model gives the tolerable error level for each pixel in the image.

4. SPATIAL DATA HIDING

In Fig. 1, we show our first data hiding technique which uses spatial masking to shape hidden data. The first step consists of selecting arbitrarily sized image blocks B_i to embed the data. Note that the data may be embedded throughout the image or localized to specific regions and objects (e.g., the face outlined with a black box). This allows the author to associate image features with specific captions.

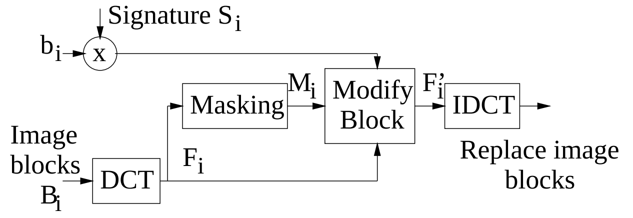

Figure 2. Diagram of frequency data hiding technique.

The message bits b_i are first spread using a pseudo-noise (PN) author signature S_i. As with spread spectrum communication systems, the PN sequence spreads the data spectrum, increases noise resistance, and hides the data. Spatial masking is then computed for the image blocks. The masking blocks M_i (i.e., tolerable error levels) are used to modify the hidden data $W_i = b_i(B_i * M_i * S_i)$, where $*$ is element-wise multiplication. The hidden data W_i are then added to the image. The masking blocks M_i shape the hidden data to guarantee invisibility and increase robustness.

As the hidden data are noise-like, they may only be extracted by a receiver that knows the PN signature S_i. A conventional receiver with access to S_i may be used to detect the hidden data. Specifically, the received image data blocks B_i' are projected onto the *author signature* S_i *weighted by the estimated image mask* M_i' and then thresholded to obtain an estimate of the data bit. In this way, image coefficients which are modified the most are given the most weight when the hidden data is recovered. Note that if the image has been cropped or translated, signature synchronization can be obtained using a two dimensional search. In some cases, e.g., spatial rescaling, a generalized likelihood ratio test must be applied.

5. FREQUENCY DATA HIDING

Our second data hiding approach is based on the frequency masking characteristics of the image data. The scheme is shown in Fig. 2. Again, blocks B_i are selected to hide the data b_i which are first spread by signature S_i. A discrete cosine transform (DCT) is applied to each block to form a DCT block F_i. A perceptual analysis stage is then applied to each DCT block to form frequency masking blocks M_i. A bit b_i is hidden in block F_i by modifying the DCT coefficients according to the equation

$$F_i'(j,k) = \left(\left[\frac{F_i(j,k)}{M_i(j,k)} \right] + \frac{1}{4} b_i S_i(j,k) \right) M_i(j,k),$$

where $[\cdot]$ denotes the rounding operation. The original image blocks B_i are replaced by the inverse DCT's of the modified blocks F_i'. Spatial masking is applied to the modified image to verify that the hidden data are invisible.

Given the image with (possibly modified) hidden data blocks F_i'', the data bit b_i may be recovered by forming the difference

$$\hat{b_i} = \sum_{j,k} M_i'(j,k) sgn\left(\frac{F_i''(j,k)}{M_i'(j,k)} - \left[\frac{F_i''(j,k)}{M_i'(j,k)}\right]\right)$$

where M_i' is the frequency mask *estimated* by the receiver times the signature S_i, i.e., $M_i' = M_i^{est} * S_i$, and $sgn(\cdot)$ is the sign value. Again, the bit decision for block B_i is weighted by the mask M_i'. Unlike spatial data hiding, the bit error rate (BER) of this scheme is *zero* when no distortion is present in the received image. We can derive a simple expression for the upper bound on the BER when zero mean Gaussian noise with variance σ^2 is added to the signal. Without loss of generality, assume that $b_i = 1$. A decision error occurs for coefficient $F''(j,k)$ whenever the magnitude of a noise sample $|w(j,k)|$ falls in one of the intervals

$$\left[\frac{(4n+1)M(j,k)}{4}, \frac{(4n+3)M(j,k)}{4}\right] = I_n$$

for $n = 0, 1, 2, \ldots$. Using the complementary error function $erfc(\cdot)$, we may write the probability of error for coefficient $F''(j,k)$ as

$$P_e(F''(j,k), \sigma) = 2 \sum_{n=0}^{\infty} erfc(\frac{I_n}{\sigma}).$$

For σ fixed, $P_e(F''(j,k), \sigma)$ decreases as $M(j,k)$ increases. Hence the receiver places more weight on coefficients with large masking values. The overall probability of error for bit b_i is a weighted combination of the $P_e(F''(j,k), \sigma)$ in block B_i.

6. PREPROCESSING HIDDEN DATA

To add further robustness to the hidden data, the data hiding techniques introduced here may be modified to take into account certain signal processing operations. If it is known that a JPEG coder will be applied to the image, for example, we can modify each data hiding procedure appropriately. In the spatial data hiding case, the signature-masking product is computed and then compressed/decompressed using a JPEG coder at an estimated quality factor Q. The decompressed product can then be modified to compensate for energy losses. In the frequency hiding scheme, the mask M_i may be preprocessed using the JPEG quantization table by substituting a new mask $\hat{M_i} = Q * M_i$ for M_i.

7. RESULTS

We illustrate our data hiding techniques on the 256 × 256 grayscale image shown in Fig. 3. Using each scheme, we embedded the text "We are investigating data hiding in multimedia systems" (432 bits) in the image by converting the text into bits and embedding each bit in an 8 × 8 image block. The text "Lena, 123 Main Street" (168 bits) is also hidden in blocks about the face object in the image (outlined by the block box in Fig. 1). The image with the hidden data is shown in Fig. 4.

In Fig. 5 we show a plot of BER for differing levels of white noises using spatial data hiding. Note that since the signature and noise are approximately uncorrelated, the BER remains constant for all noise levels. The BER of the scheme for JPEG coding at different quality settings is shown in Fig. 6. Preprocessing of the signature-mask product at a quality of 80% was used.

A plot of BER for differing levels of white noises using frequency data hiding is shown in Fig. 7. Note that this scheme performs better in low noise conditions and worse in high noise conditions than the spatial technique. Similarly for the plot of BER versus JPEG coding at different quality settings shown in Fig. 8. The frequency scheme works well under high quality coding conditions yet degrades more rapidly than spatial data hiding when the coding becomes too lossy. JPEG preprocessing at a quality of 70% was used.

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] E. Milbrandt, "Steganography Info and Archive." Available at URL *http://indyunix.iupui.edu/~emilbran/stego.html.*

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

[6] 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.

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

[8] 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.

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

Figure 3. Original 256x256 grayscale image.

Figure 4. Image with hidden data.

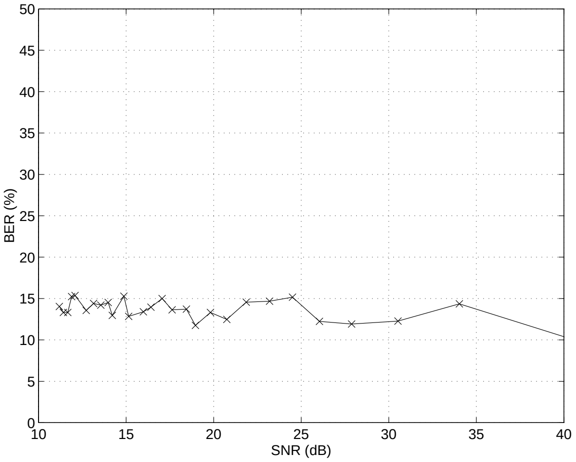

Figure 5. Bit error rate versus SNR using spatial data hiding.

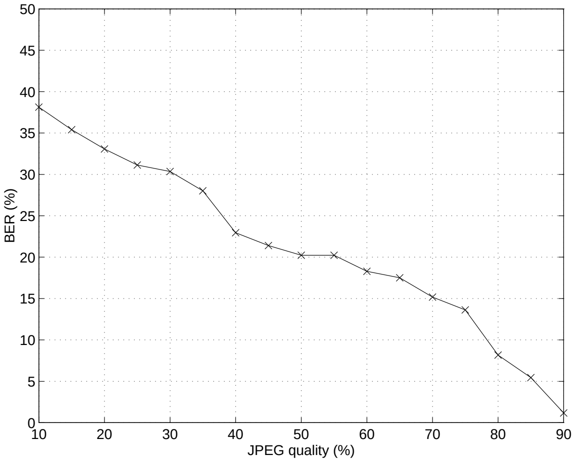

Figure 6. Bit error rate versus JPEG coding at different qualities using spatial data hiding.

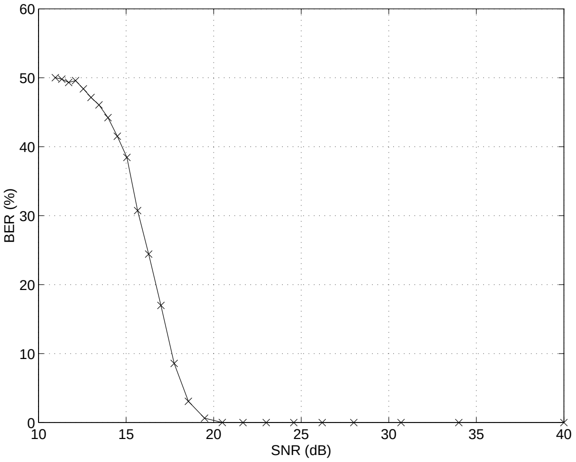

Figure 7. Bit error rate versus SNR using frequency data hiding.

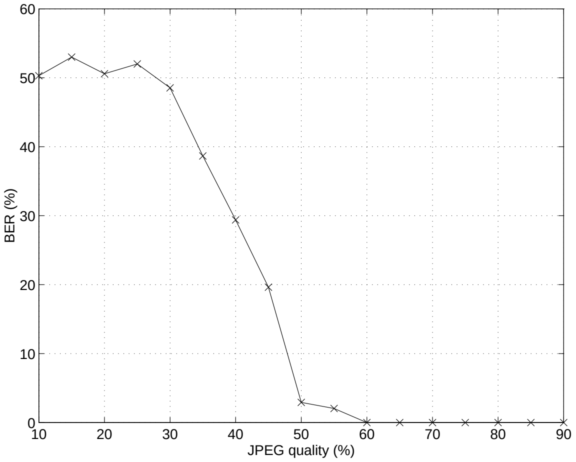

Figure 8. Bit error rate versus JPEG coding at different qualities using frequency data hiding.