Digital Signature of Color Images using Amplitude Modulation

Martin Kutter

Signal Processing Laboratory, EPFL
1015 Lausanne Switzerland

Frédéric Jordan

Signal Processing Laboratory, EPFL
1015 Lausanne Switzerland

Frank Bossen

Signal Processing Laboratory, EPFL
1015 Lausanne Switzerland

ABSTRACT

Watermarking techniques, also referred to as digital signature, sign images by introducing changes that are imperceptible to the human eye but easily recoverable by a computer program. Generally, the signature is a number which identifies the owner of the image. The locations in the image where the signature is embedded are determined by a secret key. Doing so prevents possible pirates from easily removing the signature. Furthermore, it should be possible to retrieve the signature from an altered image. Possible alternations of signed images include blurring, compression and geometrical transformations such as rotation and translation. These alterations are referred to as attacks. A new method based on amplitude modulation is presented. Single signature bits are multiply embedded by modifying pixel values in the blue channel. These modifications are either additive or subtractive, depending on the value of the bit, and proportional to the luminance. This new method has shown to be resistant to both classical attacks, such as filtering, and geometrical attacks. Moreover, the signature can be extracted without the original image.

Keywords: Watermarking, digital signature, copyright, color image, geometrical attack, steganography

1. INTRODUCTION

The emergence of digital imaging and of digital networks has made duplication of original artwork easier. In order to protect these creations, new methods for signing and copyrighting visual data are needed. Watermarking techniques, also referred to as digital signature, sign images by introducing changes that are imperceptible to the human eye but easily recoverable by a computer program. Generally, the signature is a number which identifies the owner of the image. The locations in the image where the signature is embedded are determined by a secret key. Doing so prevents possible pirates from easily removing the signature. Furthermore, it should be possible to retrieve the signature from an altered image. Possible alternations of signed images include blurring, compression and geometrical transformations such as rotation and translation. These alterations are referred to as attacks.

Several watermarking algorithms have been developed in the past. Van Schyndel et al.[5] and Bender et al.[1] proposed a straightforward technique to sign gray scale images by adding a watermark image to the original image. A modification of the dithering rule was suggested by Matsui and Tanaka.[4] Another approach is based on the modification of DCT coefficients within a JPEG or an MPEG encoder.[2]

The main drawback of these early techniques is the lack of robustness to attacks. More recently, a spread spectrum technique has led to significant improvements.[3] Although it resists to filtering, it is vulnerable to geometrical attacks such as rotation, translation and image composition.

A new method based on amplitude modulation is presented. Single signature bits are multiply embedded by modifying pixel values in the blue channel. These modifications are either additive or subtractive, depending on the value of the bit, and proportional to the luminance. This new method has shown to be resistant to both classical and geometrical attacks. Moreover, the signature can be extracted without the original image.

This paper is structured as follows. Section 2 gives an overview of the new method. First the single embedding and retrieval of a single bit is described. This process is then generalized to multiple embedding of the same bit and to embedding of multiple bits. Section 3 describes how robustness to geometrical attacks is achieved. Section 4 shows some results and finally some conclusions are drawn in section 5.

2. ALGORITHM OVERVIEW

The main requirements for a digital signature are both invisibility to the human eye and robustness to alterations. To comply with the first requirement the signature is embedded in the blue channel, which is the one the human eye is least sensitive to. Also, changes in regions of high frequencies and high luminance are less perceptible, and thus favored. Robustness is achieved by embedding the signature several times at many different locations in the image.

First the single embedding and retrieval of a single bit is described. This process is then generalized to multiple embedding of the same bit and to embedding of multiple bits.

2.1. Single bit embedding

Let s be a single bit to be embedded in an image $I = \{R, G, B\}$, and $p = (i, j)$ a pseudo-random position within I. This position depends on a secret key K, which is used as a seed to the pseudo-random number generator. The bit s is embedded by modifying the blue channel B at position p by a fraction of the luminance $L = 0.299R + 0.587G + 0.114B$ as:

$$B_{ij} \leftarrow B_{ij} + (2s - 1)L_{ij}q \tag{1}$$

where q is a constant determining the signature strength. The value q is selected such as to offer best trade-off between robustness and invisibility.

2.2. Single bit retrieval

In order to recover the embedded bit, a prediction of the original value of the pixel containing the information is needed. This prediction is based on a linear combination of pixel values in a neighborhood around p. Empirical results have shown that taking a cross-shaped neighborhood gives best performance. The prediction $\hat{B}_{ij}$ is thus

computed as:

$$\hat{B}_{ij} = \frac{1}{4c}\left(\sum_{k=-c}^{c} B_{i+k,j} + \sum_{k=-c}^{c} B_{i,j+k} - 2B_{ij}\right) \tag{2}$$

where c is the size of the cross-shaped neighborhood.

To retrieve the embedded bit the difference δ between the prediction and the actual value of the pixel is taken:

$$\delta = B_{ij} - \hat{B}_{ij} \tag{3}$$

The sign of the difference δ determines the value of the embedded bit.

The embedding and the retrieval functions are not symmetric, that is the retrieval function is not the inverse of the embedding function. Although correct retrieval is very likely, it is not guaranteed. To further reduce the probability of incorrect retrieval, the bit is embedded several times, as described in the next section.

2.3. Multiple embedding

To improve retrieval performance, the bit can be embedded n times at different locations. These n positions $p_1, \ldots, p_n$ are determined by a pseudo-random sequence. As before the pseudo-random number generator is initialized with a seed equal to the secret key K.

By using a density parameter ρ, the redundancy control can be made image size independent. This density gives the probability of any single pixel being used for embedding. This value thus lies between 0 and 1, where 0 means that no information is embedded, and 1 that information is embedded in every pixel. The number of pixels used for embedding is equal to ρ times the total number of pixels in the image.

The locations for embedding are determined as follows: for each pixel of the image, a pseudo-random number x is generated. If x is smaller than ρ, then information is embedded into the pixel. Otherwise the pixel is left intact. In this process the scanning order is modified to make it image size independent. Instead of scanning the image line by line, column by column, a zig-zag like path is taken, as illustrated in figure 1.

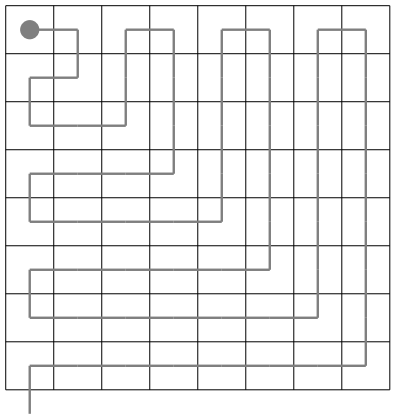


Figure 1: Modified image scanning order

To retrieve the bit, the difference between the prediction and the actual value of the pixel is computed for each p_k:

$$\delta_k = B_{p_k} - \hat{B}_{p_k} \tag{4}$$

These differences are then averaged:

$$\bar{\delta} = \frac{1}{\rho|I|} \sum_k \delta_k \tag{5}$$

where $|I|$ is the number of pixels contained in image I.

The sign of the average difference $\bar{\delta}$ determines the value of the multiple embedded bit.

2.4. Extension to an m-bit signature

The extension to an m-bit signature $S = \{s_0, \ldots, s_{m-1}\}$ is straightforward: let $p_1 \ldots p_n$ be the n positions selected for the multiple embedding of a single bit. For each of these positions a signature bit is randomly selected and embedded.

Given an $m-2$ bit string to be embedded, 2 bits are added to the string to form an m-bit signature. These two bits are always set to 0 and 1, respectively. There are two reasons to do so:

1. it allows to define a threshold τ which improves signature retrieval
2. it defines a geometrical reference which is used to counter geometrical attacks, such as rotation, cropping, translation

These items are further described in the next sections.

2.5. Adaptive threshold

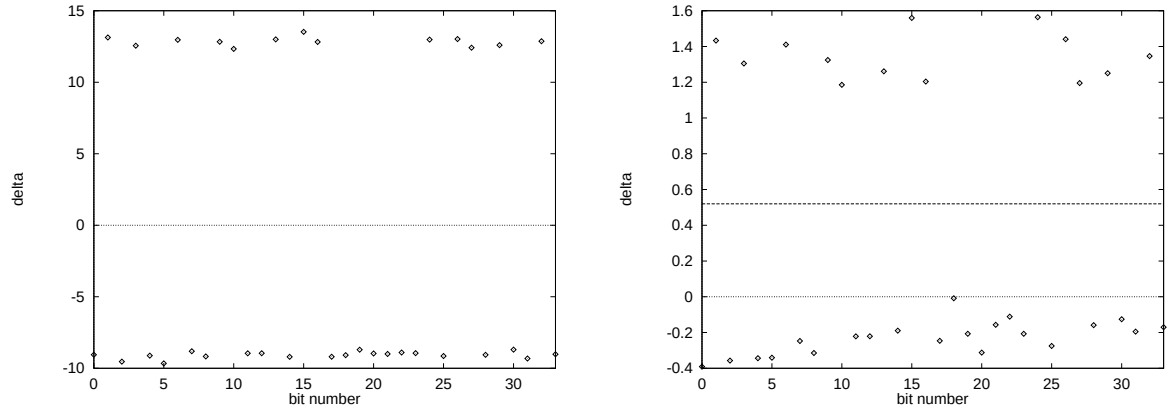
Considering each difference $\bar{\delta}^b$ that is used for information retrieval, the left graph in figure 2 clearly shows that the sign of $\bar{\delta}^b$ is a very good decision function. However, the right graph suggests that after an attack, this is not so anymore. Therefore the decision function needs to be adapted. Since it is known that the two first bits of the signature have values 0 and 1, respectively, this information can be used to compute an adaptive decision threshold. This threshold is defined as the average between $\bar{\delta}^0$ and $\bar{\delta}^1$:

$$\bar{s}^b = \begin{cases} 1 & \text{if } \bar{\delta}^b > \frac{\bar{\delta}^0 + \bar{\delta}^1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

3. ROBUSTNESS TO GEOMETRICAL ATTACKS

In order to resist to geometrical attacks, it is required for the recovering algorithm to be able to determine what operation (translation, rotation) has been applied to produce the tampered image. To estimate this transform, a reference is needed. The two first bits of the signature can fulfill this requirement. Since these bits always have the same value, a known pattern is hidden within the image. By looking for this pattern the transform can be found.

Let G be the transform applied to the signed image to obtain the tampered image J: $J = G(I)$. For now, it is assumed that the transform G is affine. Let (i, j) be the position of a pixel in I. The corresponding pixel in J



Signed image | Signed image after attack

Figure 2: Behavior of $\bar{\delta}^b$ (delta) before and after an image attack (low pass filtering)

is at position $(\tilde{i}, \tilde{j})$ and is related to (i, j) by:

$$\begin{pmatrix} \tilde{i} \\ \tilde{j} \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & d \\ c & d & e \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} i \\ j \\ 1 \end{pmatrix} \tag{7}$$

where $a, \ldots, e$ are the transform parameters.

In order to retrieve the signature, the inverse G^{-1} of G is needed. By applying G^{-1} to J the image I is recovered and the signature can then be extracted.

The transform G^{-1} can be found by looking for the pattern created by the two first bits of the signature. Let H be an estimation of G^{-1}, and I_H the image obtained by applying H to J. Let's first suppose that H is equal to G^{-1}. The two first bits of the signature can clearly be retrieved from I_H. Also, the confidence of the retrieval is very high, that is the difference between $\bar{\delta}^1$ and $\bar{\delta}^0$ is maximum. Suppose now that H is slightly different from G^{-1}. The signature can still be retrieved but the difference between $\bar{\delta}^1$ and $\bar{\delta}^0$ is smaller than before. This difference gets smaller as the divergence between H and G^{-1} grows.

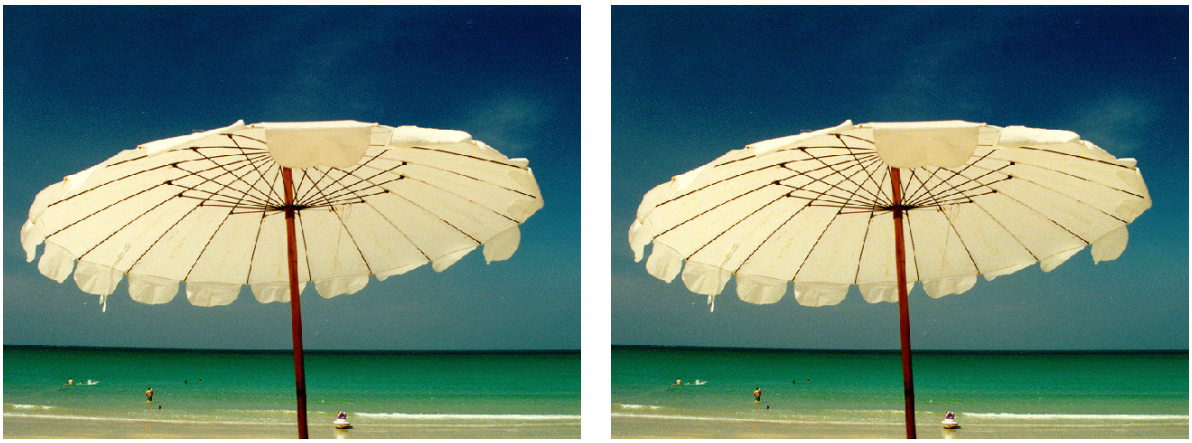
The difference can thus be used as an optimization criterion $q(H)$ defined as $q(H) = \bar{\delta}^1(I_H) - \bar{\delta}^0(I_H)$. As mentioned before $q(H)$ is maximal for $H = G^{-1}$. However the function $q(H)$ is not a smooth function. Optimization methods such as gradient descent would thus not be suitable. In this case full search methods have to be used.

The search can be sped up a lot if the nature of the transform G is given. For instance if it is known that the transform is a pure translation or rotation, the search space is greatly reduced.

4. RESULTS

To confirm the invisibility of the embedding process, an image (640 by 480 pixels, 24-bit color) with blue tones has been signed (see figure 3). For this particular image, the parameters have been set as follows: the signature S is 34 bit long and its value is the 32-bit number 1234567890 augmented by the two constant bits 0 and 1, the embedding density ρ is 0.55, the embedding strength is given by $q = 0.1$, and the size of the crossed-shape window

is $c = 3$.



Original image Signed image

Figure 3: Invisible embedding of information

To verify the robustness of the proposed method, the signed image has been attacked in several ways, namely:

- blurring
- JPEG encoding/decoding
- rotation
- composition with another image

The next sections describe more precisely each of these attacks. Although in this document comprehensive results are only provided for these four attacks, the method has also shown to be resistant to other attacks including pixel spreading, pixelizing, color quantization, translation, cropping, and despeckling (median filtering).

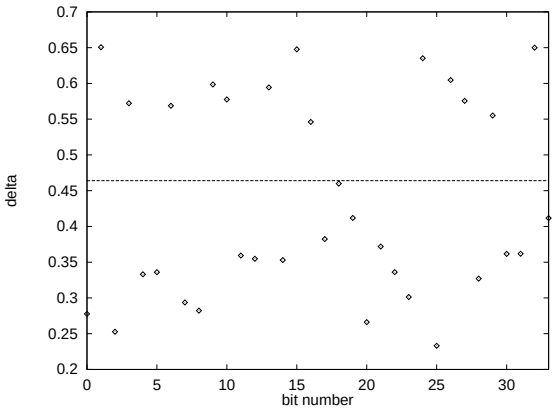
4.1. Blurring

Figure 4 shows the signed image after blurring. The blurring function is as follows. Each color value is replaced by the average value of pixels within a 5 by 5 neighborhood.

The graph on the right hand side of figure 4 clearly indicates that the strength of the signature is much lowered by the attack. The average absolute difference between the $\bar{\delta}^k$ and the threshold τ goes down from about 10 before the attack (see figure 2) to about 0.1 after the attack.

Although the signature is correctly retrieved after the blurring, the limits of the proposed method appear. Indeed the $\bar{\delta}^{16}$ lies very close to the threshold and blurring the image even more would probably result in an erroneous signature retrieval. However the image quality would then also be much lower.

Blurred image



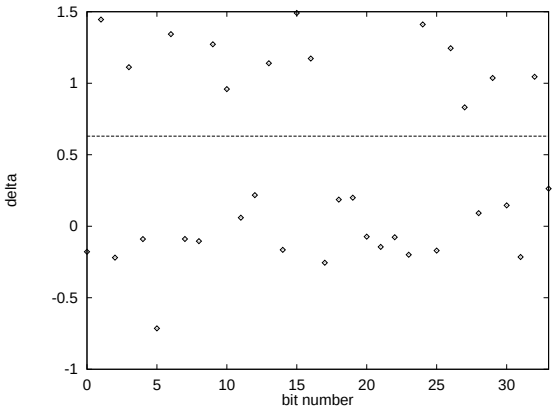
Retrieval quality

Figure 4: Simple image attack: blurring

4.2. JPEG encoding

Figure 5 shows the signed image after a JPEG encoding/decoding cycle. The quality factor for JPEG compression was set to 75 percent, which is the default value. Again the average absolute difference between the $\bar{\delta}^k$ and the threshold τ is much lower than before the attack. However each $\bar{\delta}^k$ clearly lies on one of the sides of the threshold, and the signature can thus be correctly retrieved with great confidence.

Image after JPEG compression/decompression



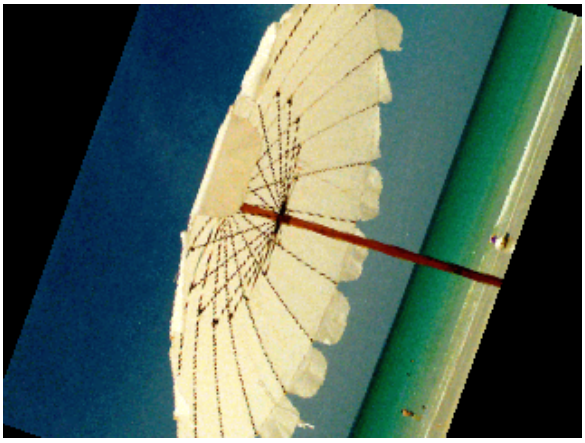
Retrieval quality

Figure 5: JPEG attack

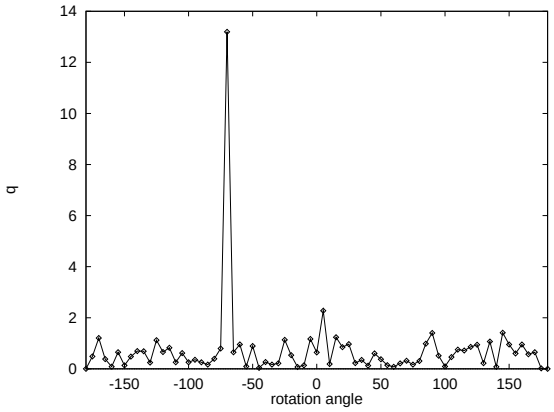
4.3. Rotation

To test the robustness of the proposed method against geometrical attacks, the signed image has been rotated to the left by an angle of 70 degrees (see figure 6). Considering that it is known that the attack is purely rotational, $q(H)$ is computed for every H defined as a rotation between 0 to 360 degrees, with increments of 5 degrees. The graph in figure 6 shows how the value $q(H)$ is affected by the angle. Clearly the optimum lies at -70 degrees, that

is, the amount by which the signed image was rotated. The signature can thus accurately be retrieved. Although a full search technique is used, the retrieval is still quite fast. Less than 20 seconds[1] were needed to do so.



The signed image rotated 70 degrees to the left



Searching for the transform: $q(H)$ as a function of the rotation angle

Figure 6: Geometrical attack: rotation

4.4. Composition with another image

Figure 7 shows an example of image composition. Given two signed images, each being signed with a different key, a third one is created by taking some pixels from the first one and some from the second one. This procedure can also be seen as a mixture of cropping and translation.

In this case, the algorithm is able to correctly retrieve both signatures given the appropriate keys.

5. CONCLUSION

A novel technique for image watermarking has been presented. The signing process has shown to be unnoticeable to the human eye. It has also been demonstrated that the signature is immune to a variety of attacks, including filtering, compression, and geometrical transforms. The resistance to the latter kind of attack without the need for the original image is the main improvement brought by this new method.

The proposed algorithm could be improved in several ways. First, all color channels could be exploited. The strength of the signature in each channel would be proportional to the sensitivity of the human eye to it. Also, robustness could be improved with the use of optimal error correcting codes. The current algorithm already features a primitive error correcting code based on the multiplicity of the embedding. However, it is well known that redundancy codes are far from optimality.

The authors would like to thank Vincent Vaerman for providing the images.

[1] On a Sun Ultra 1 workstation

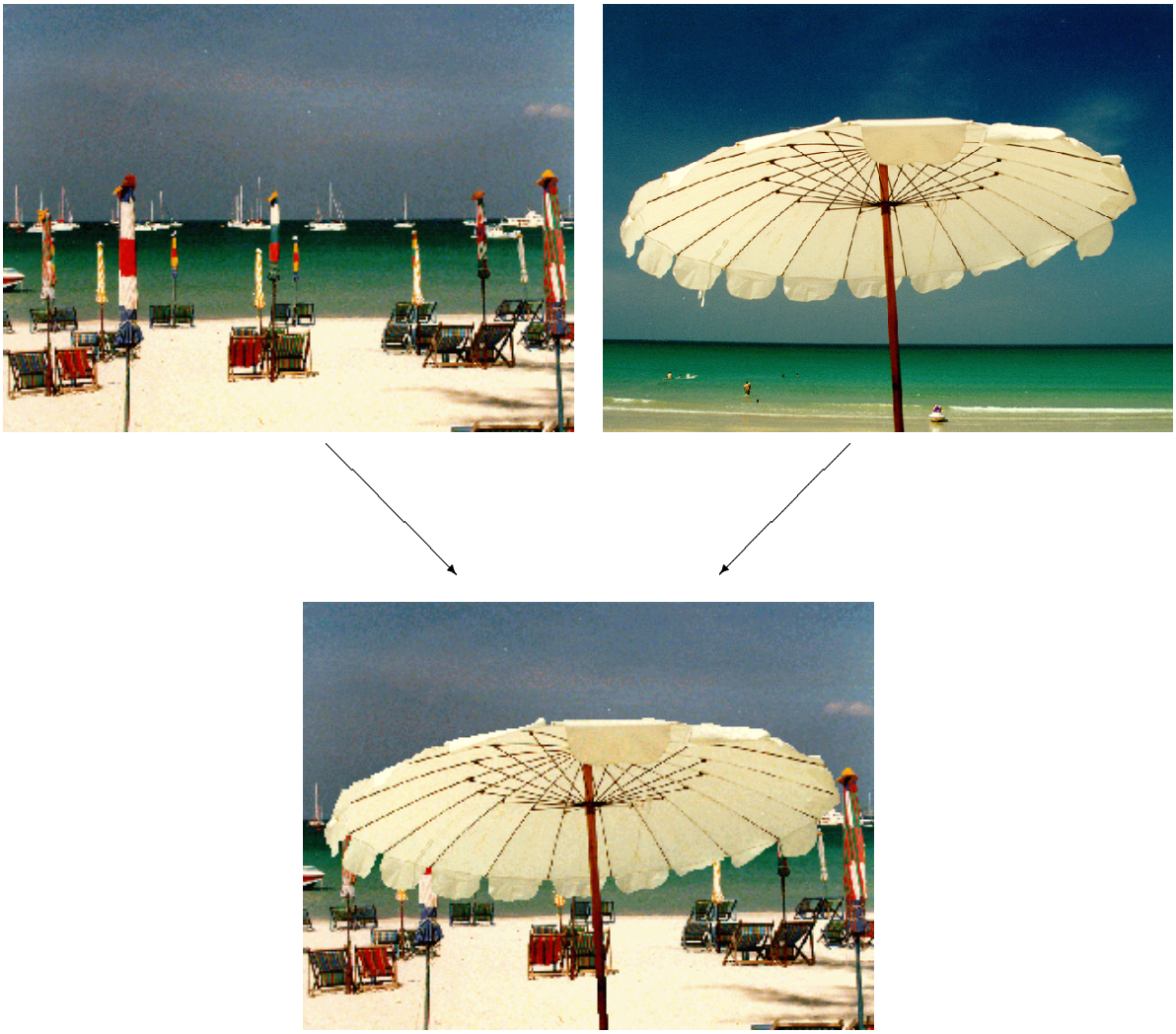


Figure 7: Composition example

6. REFERENCES

[1] W. Bender, D. Gruhl, and N. Mormoto. Techniques for data hiding. In *SPIE*, volume 2420, February 1995.

[2] S. Burgett, E. Koch, and J. Zhao. A novel method for copyright labelling digitized image data. *IEEE Transactions on Communications*, September 1994.

[3] I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. Technical Report 95-10, NEC Research Institute, 1995.

[4] K. Matsui and K. Tanaka. Video-steganography: How to secretly embed a signature in a picture. *Journal of the interactive Multimedia Association Intellectual Property Project*, 1(1):187–206, January 1994.

[5] R.G. van Schyndel, A.Z. Torkel, and C.F. Osborne. A digital watermark. In *1st IEEE International Conference on Image Processing*, volume 2, pages 86–90, 1994.