4-7-2012

Performance measures for image watermarking schemes

Chaouki T. Abdallah

Gregory L. Heileman

Carlos E. Pizano

Recommended Citation

PERFORMANCE MEASURES FOR IMAGE WATERMARKING SCHEMES

Gregory L. Heileman[†‡], *Carlos E. Pizano*[‡], *and Chaouki T. Abdallah*[⋆]

Department of Electrical and Computer Engineering
University of New Mexico, Albuquerque, NM, USA 87131
{heileman,cpizano,chaouki}@eece.unm.edu

ABSTRACT

In this paper we propose performance measures and provide a description of the application of digital watermarking for use in copyright protection of digital images. The watermarking process we consider involves embedding visually imperceptible data in a digital image in such a way that it is difficult to detect or remove, unless one possesses specific secret information. The major processes involved in the watermarking processes are considered, including insertion, attack, and detection/extraction. Generic features of these processes are discussed, along with open issues, and areas where experimentation is likely to prove useful.

1. INTRODUCTION

The term *digital watermarking* is currently being used to describe the process of adding information, in particular digital watermarks, to digital multimedia content. Digital watermarks are being used for the purposes of documenting or ensuring (i.e., verifying, guaranteeing, or proving) the integrity of the multimedia content. Specifically, there are two general ways in which digital watermarks are being used with regards to multimedia content: (1) To add information to the content in such a way that it is clearly, and purposefully, available to those accessing the content. This is related to the traditional idea of watermarking (e.g., translucent marks that are routinely placed in paper currency or bond paper). A typical application is the automatic placement of a logo in an image taken by a digital camera. (2) To add information to the content in such a way that it is hidden from those accessing the content, unless they know to look for it, and possess any secret information needed to decode it. This is related to the traditional idea of *steganography*—a word derived from the Greek word meaning *covered writing*. The field of steganography is also referred to as *information hiding*; however, both of the uses described above are commonly referred to simply as *digital watermarking* in the current literature. The notion of embedding hidden or barely perceptible information within a message or picture is actually an old idea, dating back to antiquity. Interesting accounts of this history can be found in [7, 8]. In this paper we will focus on the use of digital watermarking techniques for the purpose of copyright protection and ownership verification of digital images. This requires information to be hidden in digital images (i.e., this is an application of usage (2) described above).

The need for developing watermarking techniques that protect electronic information has become increasingly important due to the widespread availability of methods for disseminating this information (e.g. via the Internet), and the ease with which this information can be reproduced [2, 3]. In fact, the inability to develop provably strong watermarking methods for copyright protection is often cited as a major stumbling block to the further commercial development of the Internet. For this reason, watermarking research has received considerable attention over the past few years, and an ever-increasing number of watermarking methods for copyright protection are being proposed both in the open and patent literatures [5, 4, 6, 10, 11, 12, 13]. These methods are invariably accompanied by the claim that they are "robust against malicious attacks," and these claims are often backed up via experimental tests devised by the developers themselves. Rarely do the authors of different papers consider the same suite of attacks, and worse yet, it is often the case that the same attack (e.g., sub-sampling) is implemented differently (e.g., some authors assume the image can be resized during a manual preprocessing step and others do not). The need for the establishment of a standard set of attacks by

which watermarking methods can be benchmarked is clearly evident. Similarly, since the performance of all watermarking methods is image dependent, a standard image database must be established. Thus, increased calls are being made for the establishment of watermarking standards (c.f., [9]). This paper presents a useful step in that direction by considering a mathematical model that has proved successful in fractal image processing [1]. We also make the case that experimentation is likely to be an important tool in the development, testing, and possible standardization of watermarking algorithms.

Throughout this paper we discuss a number of open issues related to the development of robust watermarking algorithms for copyright protection of digital images. We summarize the difficulties of this problem by considering the tradeoffs associated with watermark *robustness*, *imperceptibility*, and *information capacity*. A watermarking technique cannot simultaneously maximize these three quantities; rather, they tend to compete with each other, as depicted in Fig. 1. Specifically, in this figure we denote the space of all watermarking methods according to the three axes of robustness, imperceptibility, and information capacity. The methods that are realizable are shown as the shaded region in the figure, we envision them as forming a connected region about the origin, but we make no claims about the actual shape of this region. This figure supports the notion that it is possible to make a watermark more robust to various forms of attack, but this must come at the expense of conveying less information in the watermark, and more image degradation. For example, watermarks can generally be made more robust to attacks if the watermark bits are replicated throughout the image, but this obviously leads to a reduction in the information capacity of the watermark. Furthermore, a watermark can be made more robust to attack if it is placed in perceptually important regions of the image. In this case, attacks that attempt to remove the watermark are likely to result in noticeable changes to the image. However, only a limited amount of data can be hidden in these regions; if too much is hidden, the watermark itself leads to unacceptable perceptual changes. Finally, since an increase in the information capacity of an efficiently coded watermark will require an increase in the number of bits used by the watermark, we see that increasing the information capacity will eventually lead to perceptually significant image modifications. For this reason, a desirable property of watermarking methods used for copyright protection is the ability of the user to specify the level of robustness. The user can then examine the watermarked data to determine if it is acceptable, and repeat this process with
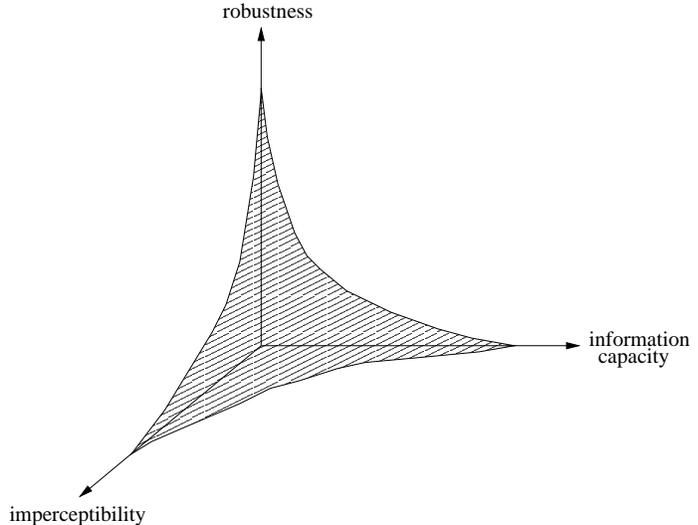

Figure 1: The space of all watermarking methods for copyright protection.

a different level of robustness if necessary. We propose such an approach accompanied with a test that measures the distortion between the attacked image and the original one using a norm weighed by a perceptual mask.

Increasingly, calls are being made for the establishment of watermarking standards (c.f., [9]). Prior to the establishment of any standards for attacks or image databases, it will be important to establish reasonable models for watermarking. This paper presents a useful step in that direction by considering a mathematical model that has proved successful in fractal image processing [1]. It will also be important to have unbiased means of judging watermarking methods during any standardization process. Based on the discussions contained in this paper, we believe experimental investigations can help lend direction to theoretical research on watermark robustness in this case.

We have also considered the issue of computational constraints when implementing an attack, since computational requirements may also play a limiting role in implementing one of the specific watermarking methods depicted in Fig. 1, and will factor prominently in the development of any watermarking protocols. These issues require further investigation.

2. CONSTRAINTS

The required invisibility of the watermark in the image leads to the following
    constraints: The watermark insertion cannot change the perceptual

content of the image I, and an attack will only be considered successful if it

defeats the detection/extraction processes without changing the perceptual content of

the image I_w. We will use the notation $I \sim J$ to denote that image I

is perceptually equivalent to J, and leave the exact definition of equivalence for alter. The constraint $I \sim I_w$ models the

typical situation in which a graphic artist, photographer, etc. has created an image in

which they would like to place a watermark, but they require that this be done

without altering the esthetics of the image. Now, if we let an attack be denoted by $\hat{I}_w = f(I_w)$, the constraint $I_w \sim \hat{I}_w$

captures the notion that a pirate will only find it useful to steal an image if

they can do so without altering its esthetics. Thus, if it is possible to construct a

a watermarking method in which perfect detection is obtained whenever $I_w \sim \hat{I}_w$, then a pirate will be forced to noticeably alter an image in order to

steal it. We believe that this is the best that one can hope for in this application.

In summary, the constraints we will assume are:

$$I \sim I_w \tag{1}$$
$$I_w \sim \hat{I}_w \tag{2}$$

We model an attack as a linear or nonlinear function $f(\cdot)$ applied to the

watermarked image I_w. That is, $f(I_w) = \hat{I}_w$ By transivity, constraints (1) and (2) imply that

$I \sim \hat{I}_w$ The main problem with the previous discussion is that the concept of a

perceptual invariant is ill-defined, and in fact involves open problems

in brain science, psychology, as well as subjective opinions. Nevertheless,

enough is known about the human visual system (HVS) to exploit the idea of

perceptual invariants in watermarking methods. A number of models for

the low-level processing portion of the HVS have been proposed, we will

denote these as a perceptual masking function $I^P = P(I)$. Specifically, we will

assume $P(I)$ operates on images I, returning an image the same size as I.

Entry $[ij]$ of the perceptual masking image matrix, i.e., $[P(I)]_{ij}$, is

directly porptional to the perceptual sensitivity of pixel $[I]_{ij}$. That

Figure 2: The distance between I_w and $\hat{I}_w$ in image and perceptual space.

is, a large value for $[P(I)]_{ij}$ indicates that pixel $[I]_{ij}$ in image

I can be changed signivicantly, without changing the perceptual characteristics

of the image. Likewise, a small value for $[P(I)]_{ij}$ indicates that

pixel $[I]_{ij}$ can not be altered much without altering the perceptual content

of the image I. For ease of presentation, we will assume the elements of $P(I)$

are all nonnegative. Now we can define equivalence between the 2 images I and J as follows,

$$I \sim J \iff I^P = J^P \tag{3}$$

Figure **??** demonsrates how images that are far apart in image space

may be close together in perceptual space

We are interested in measuring the distances between I, I_w, and $\hat{I}_w$ in perceptual space, and using this to develop a more formal notion of the

robustness of a watermarking method. In order to measure the distance between

two images I and J in perceptual space, we transform the matrix $P(I)$ in

the manner described below to create matrix $Q(I)$, and then we define the matrix $\Gamma(I, J)$ as follows

$$\Gamma(I, J) = Q(I) \circ |I - J|$$

where $A \circ B$ is the Schur product of two $m \times n$ matrices A and B:

$$[A \circ B]_{ij} = [A]_{ij} \cdot [B]_{ij},$$

$i = 1, \ldots, m$, $j = 1, \ldots, n$.

Matrix $Q(I)$ is obtained from $P(I)$ by first computing $[P]_{ij}^{-1}$ (i.e., inverting the individual elements of the matrix), and scaling the resulting

values to the interval $[0, 1]$. Thus, $[\Gamma(I, J)]_{ij} \in [0, 1]$,

$i = 1, \ldots, m$, $j = 1, \ldots, n$. We can think of $\Gamma(I, I)$ as being

the origin in this perceptual space, and $\Gamma(I, J)$ is the distance

in perceptual space of J from the origin. A number of norms can be used to

measure this distance. For example we may consider

$$\|\Gamma\|_1 = \max_j \sum_{i=1}^{m} |[\Gamma]_{ij}| \qquad (4)$$

$$\|\Gamma\|_2 = [\lambda_{\max}(\Gamma^T\Gamma)]^{\frac{1}{2}} \qquad (5)$$

$$\|\Gamma\|_\infty = \max_i \sum_{j=1}^{n} |[\Gamma]_{ij}| \qquad (6)$$

$$\|\Gamma\|_x = \sum_{i=1}^{m} \sum_{j=1}^{n} |[\Gamma]_{ij}| \qquad (7)$$

where $\lambda_{\max}(\Gamma^T\Gamma)$ is the maximum eigenvalue of $(\Gamma^T\Gamma)$. The norm $\|\Gamma\|_x$ is useful because it is sensitive to local "spikes" in Γ, which may correspond to point distortions in I_B, while the other norms consider a more global view of Γ, corresponding to distrotions of a more global nature.

To recap, we now have the space of images $\mathcal{S}$ which contains the original image I, its watermarked version $I_w = W(I)$ and its attacked version $\hat{I}_w = f(I_w)$. We also have the corresponsing perceptual mask imahes $I^P = P(I)$, $I_w^P = P(I_w)$, and $\hat{I}_w^P = P(\hat{I}_w)$. Because of the definition of the equivalence relation between any 2 members of $\mathcal{S}$, we have equivalence classes which completely partition $\mathcal{S}$. We also obtain the quotient space Q which is the set whose members are equivalent classes, i.e.

$$\begin{aligned} \pi &:& \mathcal{S} \longrightarrow Q \\ \pi I &=& \mathcal{I} \end{aligned}$$

such that $J \in \mathcal{I}$ if and only if $I \sim J$. We can also define attacks f which *respect* the equivalence relation $\sim$ as those functions $f : \mathcal{S} \longrightarrow]calS$ where

$$I \sim J \Longrightarrow f(I) \sim f(J)$$

We are now ready to mathematically define the concepts of robustness, imperceptibility, and infromation capacity of a watermark w. Let us first define $s_w = I - I_w$, and $\hat{s}_w = I_w - \hat{I}_w$ to be the watermark signature and its attacked version

Definition 1 *A watermark w on an image I is robust with respect to an attack f if*

$$I_w \sim \hat{I}_w \Longrightarrow s_w \sim \hat{s}_w$$

Definition 2 *A watermark w on an image I is imperceptible if $I \sim I_w$.*

Definition 3 *A watermark w on an image I has information content measure by $H(I; I_w)$.*

3. REFERENCES

[1] M. F. Barnsley and L. P. Hurd. *Fractal Image Compression*. AK Peters, Ltd., Wellesley, MA, 1993.

[2] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. Technical Report 95-10, NEC Research Institute, May 1995. http://www.neci.nj.nec.com/tr/index.html.

[3] S. Craver, N. Memon, B.-L. Yeo, and M. Yeung. Can invisible watermarks resolve rightful ownership. Technical Report RC20509, IBM Research Report, July 1996, and In *Proceedings SPIE Storage and Retrieval for Image and Video Databases V*, July 1997.

[4] Digimarc Homepage. http://www.digimarc.com.

[5] J. R. Hernández, F. Pérez-González, J. M. Rodríguez, and G. Nieto. Performance analysis of a 2d-multipulse amplitude modulation scheme for data hiding and watermarking of still images. to appear in *IEEE Journal on Selected Areas in Communication*.

[6] L. Irwin, G. L. Heileman, C. E. Pizano, C. T. Abdallah, and R. Jordán. The robustness of digital image watermarks. In *Proceedings of the International Conference on Imaging Science, Systems, and Technology*, pages 82–85, Las Vegas, NV, July 1998.

[7] D. Kahn. *The Codebreakers*. MacMillan, New York, 1967.

[8] D. Kahn. The history of steganography. In R. Anderson, editor, *Information Hiding, Springer Lecture Notes in Computer Science*, volume 1174, pages 183–206. Springer-Verlag, 1996.

[9] F. Mintzer, G. W. Braudaway, and A. E. Bell. Opportunities for watermarking standards. *CACM*, 41(7):57–64, 1998.

[10] N. Nikolaidis and I. Pitas. Copyright protection of images using robust digital signatures. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2168–2171, Atlanta, GA, May 1996. http://poseidon.csd.auth.gr/signatures/.

[11] F. A. Petitcolas. Image watermarking: Weaknesses of existing schemes. http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarki

[12] F. A. Petitcolas, R. J. Anderson, and M. G. Kuhn. Attacks on copyright marking systems. In *Proceedings Second International Workshop on Information Hiding, Lecture Notes in Computer Science*, Portland, Oregon, April 1998. Springer-Verlag. `http://www.cl.cam.ac.uk/~fapp2/papers/ih98-attacks/`.

[13] M. D. Swanson, M. Kobayashi, and A. H. Tewfik. Multimedia data-embedding and watermarking technologies. *Proceedings of the IEEE*, 86(6):1064–1087, 1998.