# A Role for Digital Watermarking in Electronic Commerce

Neil F. Johnson[1], Zoran Duric[2], Sushil Jajodia[1]

Center for Secure Information Systems
George Mason University
Fairfax, VA 22030-4444
http://isse.gmu.edu/~csis
{njohnson,zduric,jajodia}@gmu.edu

**Abstract**

Digital media are subject to illicit distribution and owners of such data are cautious about making their work available without some method of identifying ownership and copyright. Digital watermarks are employed in an attempt to provide proof of ownership and identify illicit copying and distribution of multimedia information. In this paper we briefly discuss the role of digital watermarking as a means of aiding in copyright and ownership issues. We present an overview of information hiding methods for digital media, propose a new way of categorizing watermark techniques through image modeling, and illustrate an alternative watermarking technique through the use of gradual masks. Attacks on watermarks are then discussed, and we propose a method for watermark recovery after attacks.

[1]Also with Information and Software Engineering Department, George Mason University.
[2]Also with Computer Science Department, George Mason University.

# 1 Introduction

The growth of digital media and the fact that unlimited numbers of perfect copies of such media can be illegally produced is a threat to the rights of content owners. A copy of digital media is an exact duplicate of the original. The authors of a work are hesitant to make such information available on the Internet as it may be copied and retransmitted without the permission of the author. An issue facing electronic commerce on the Internet for digital information is how to protect the copyright and intellectual property rights of those who legally own or posses digital works. Most electronic commerce systems use cryptography to secure the electronic transaction process. Encryption provides data confidentiality, authentication, data integrity, and in some cases authentication of the parties involved [Sch96, FB97]. However, the unencrypted data may still be copied and distributed (i.e., videotapes, DVD, and pay-per-view broadcasts). Authors also may want samples of their works to be available. In some cases, these samples may be the images used on a web site or the publication of information on the Internet. Copyright protection involves ownership authentication and can be used to identify illegal copies. One approach to copyrighting is to mark works by adding information about their relationship to the owner by a digital watermark. Digital watermarking provides a means of placing information within digital works. This information may be perceptible or imperceptible to the human senses. Early watermarking work investigated how documents can be marked so they can be traced in the photocopy process [BLMO94, BOML95].

Interest in digital watermarks is growing and seems to be motivated by the need to provide copyright protection to digital works. Watermarking can be used to identify owners, license information, or other information related to the cover carrying the watermark. Watermarks may also provide some control mechanisms such as determining if the work has been tampered with or copied illegally. Digimarc Corporation developed a search engine, *MarcSpider*, to search web sites for images that contain Digimarc watermarked images. When watermarked images are found, the information is reported back to the registered owners of the images [Dig]. In the realm of video and satellite broadcasts, watermarks are used to interfere with recording devices so copies of a broadcast are somewhat corrupt. A number of hardware and software technologies are being developed to deter illicit copying.

The remainder of the paper is organized as follows: Section 2 provides an introduction to digital watermarking, a brief survey and classification of watermarking techniques, and a simple, but

effective watermarking method. Some attacks on watermarking techniques are discussed in Section 3. In Section 4 we describe a method for recovering watermarks from images after some attacks and provide experimental results. This section contains background information on the recovery process which discusses some mathematical principles of the recovery processes. Conclusions and suggestions about future research are presented in Section 5.

## 2 Digital Watermarking

Digital watermarks have several desirable characteristics. The watermark should not degrade the image to a degree that interferes with its usefulness. The watermark should require no additional image formats or storage space. The watermark should be integrated with the image content so it cannot be removed easily without severely degrading the image. The watermark should be fairly tamper-resistant and robust to common signal distortions, compression, and malicious attempts to remove the watermark (though we will see that this is not achieved by many watermarks). The watermark can be made invisible to the human eye, but still readable by computer.

Digital watermarks may be perceptible or imperceptible to the human senses. (Throughout this paper the terms "invisible" or "imperceptible" watermarks relate to human visibility, not to computer detectability.) Many authors feel that image-based digital watermarks should be invisible to the human eye. If the watermark is to be imperceptible, then the existence of the watermark should not be advertised. Advertising the presence of watermarks invites "pirates" to attempt to alter or disable the watermarks [JJ98b]. Other authors prefer visible watermarks, or clearly advertise the existence of watermarks as a deterrent against illicit duplication or theft. Both viewpoints have merit; but the determination must be made by the owner of the images and depends on the intended use of the watermarked work. The visible watermarks in many television broadcasts are considered eyesores by some, while others simply ignore them. What constitutes interference with an image is subjective and dependent upon the end user.

Imperceptible watermarking is rooted in steganography which is the art and science of concealing the existence of a secret message [JJ98]. The difference between imperceptible digital watermarking and digital steganography is based primarily on intent. Steganography attempts to conceal the existence of a message, where the hidden message is the object of the communi-

cation, e.g., sending a satellite photograph hidden in another image. Digital watermarks contain information that may be considered attributes of the covering image such as copyright data, and the cover is the object of communication - not the watermark. Sometimes the methods used for steganography and watermarks are the same. However, unlike steganography, many imperceptible digital watermarks are made known to users, possibly as a deterrent to copying.

Various sorts of information can be stored in a watermark, including license, copyright, copy control, content authentication, and tracking information. This information can be used for copy protection, document identification, ownership designation, or as a means to track works to and from licensed users. Regardless of the type of information placed within the media content, a tradeoff exists between a watermark's payload (amount of information placed in the image) and its robustness to manipulation. Watermarks typically hide very little information and rely in part on redundancy of the mark to survive attacks such as cropping. A novel watermarking method is to use two images that appear to be the same but have slight differences so if they are viewed together as a 3D stereogram, the watermark becomes visible [HYQ98]. This approach has a low bandwidth for passing hidden information. Common steganography methods and bit-wise watermarking approaches hide larger amounts of information in a cover. However, these methods are vulnerable to attacks such as cropping, and if the embedding method relies on this "noise" level (least significant bits) of the cover, little processing is required to disable the readability of the embedded message.

## 2.1 Classifying Watermarking Techniques

Watermarks are embedded into images by changing some bits in image representation. Some methods operate on least significant bits, while others embed information into perceptually more significant image components. Current image-based digital watermarks may be grouped under two general classifications: those that fall into the *image domain* and those that fall into the *transform domain*. Tools used in the image domain include methods that use bit-wise techniques such as least significant bit (LSB) insertion and manipulation [CKSL96]. Patterns placed in the image [Car95] and spatial relationships between image components are another "additive" form of watermarking. Techniques that provide additive image information such as masking techniques [JJ98] without applying a function of the image to determine the watermark location are also categorized as being in the image domain, though they share the survivability properties of transform domain

watermarking techniques.

The transform domain classification of watermarks includes those that manipulate image transforms. Early work in this area considered the possibility that the dithering effect used for image quantization might be used to hide information [TNM90]. Transforms such as the fast Fourier transform (FFT), discrete cosine transform (DCT) [KRZ94, KZ95, ODB96], and wavelet transform [KH97, XBA97] hide information in the transform coefficients. Many variations on this approach exist, ranging from applying the transform to the entire image [CKLS95, HW96] to applying it to blocks of the image [Dig, Sig, SZT96], or using methods similar to those used in JPEG image compression [GW92, BS95]. These methods hide messages in relatively significant areas of the cover image. Transform domain watermarking and masking techniques are more robust against attacks such as lossy compression, cropping, and image processing techniques in which significant bits are changed.

Both image domain and transform domain methods may employ patchwork, pattern block encoding, or spread spectrum concepts which add redundancy to the hidden information [BGML96, CKLS95, SC96, MBR98]. These approaches help protect against some types of image processing such as cropping and rotating. The patchwork approach uses a pseudo-random technique to select multiple areas (or patches) of an image for marking. Each patch may contain the watermark, so if one is disabled or cropped, the others may survive. The message data becomes an integral part of the image by adjusting the luminance values, as in masking [JJ98]. Many of these techniques require use of the original, unwatermarked image to extract the watermark. In [ICIP97] transform domain watermarking techniques were introduced that do not require using the original to extract the watermark [PBBC97, FH97, OP97].

Watermarking techniques could also be classified based on whether an original (non-watermarked) image is needed for watermark recovery. Some watermarking techniques extract the watermark by comparing the original and the watermarked image [CKSL96, JJ98]. Other methods, particularly bit-wise techniques, do not need the original to recover the watermark, but use symmetric methods of placing data within images; the algorithm used to embed the watermark is also used to extract it. Softbots and "web crawlers" or "spiders" such as Digimarc's MarcSpider search for Digimarc watermarks in images and the related information can be retrieved from the registration service.

Another classification of watermarking techniques could be based on concepts of image mod-

eling [Ros92]. An image model has four basic components: image noise, texture, clutter (scene noise), and signal. Many of the techniques that are classified as being in the image domain are here categorized under the (image) noise domain. These include bit-wise watermarking techniques; they are sensitive to small amounts of image processing or lossy compression. Steganography tools frequently use this approach to hide data in images [JJ98].

Texture refers to the stochastic constraints on the appearance of objects/surfaces in an image or image component; examples of textures are leather, fur, plastic, etc.. The owner of an image could watermark the image by changing the texture of some background surfaces in the image; a sofa could have different appearances in several images — this could be accomplished by changing the texture of the sofa cover (leather, uniform colored cloth, etc.). Note that this method of marking images assumes that the owner can manipulate the appearances of the objects in the image. Clutter or scene noise refers to (small) objects which appear in the image. For example, in one image of an office there could be a cup in the corner of a desk and in the second image the cup could be different or replaced by another object — say a book. The owner of an image could use this method to mark differences between small numbers of images given to different users. Other examples of clutter used in watermarking are visible logos used by television broadcasters to mark their programs.

The image signal corresponds to the perceptually most significant components of an image. Watermarking techniques that change the lowest or medium frequency coefficients of a discrete cosine transform of an image could be classified in this category [CKSL96]. Another approach is rearranging the original work to create the watermark — e.g. in [LMP98] the authors embed the watermark into the layout of electronic circuits. Watermarks embedded by techniques in this category usually cannot be removed without significant change of the original work. This feature makes these techniques potentially much more useful for protecting images than the techniques that operate on the noise level.

## 2.2 A Watermarking Method

A widespread method of watermarking printed materials and television images relies on overlaying logos over images so that the logos become indelible parts of the originals. As an example a proposal draft could be marked by the large slanted word “draft” printed over each page. Similarly, network station logos appear in lower corners of television broadcasts. This idea can be applied to


(a) (b)

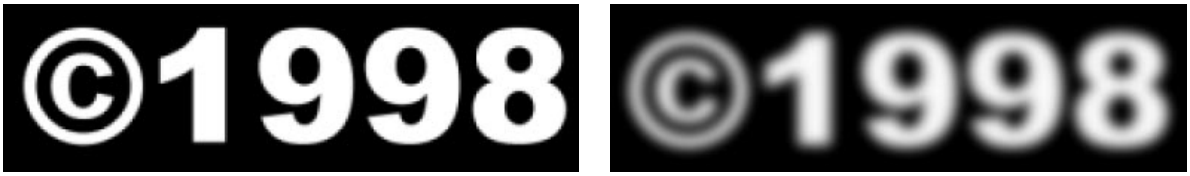
Figure 1: (a) "Flat" and (b) "gradual" image masks.

create imperceptible watermarks that can be embedded into digital images. The method employed here to embed a watermark in the image is known in graphics as "alpha channel compositing" (combining) [FvDFH96].

Figure 1 shows two masks used to create watermarks. The left image (Figure 1a) shows a "flat" mask while the right image (Figure 1b) shows a "gradual" mask. Applying the mask changes the gray values of pixels in the image. Any given image can be viewed as having 256 gray levels from black (0) to white (255). If the "flat" mask is applied, a brightness change of about 2% can be used before the watermark becomes visible in the low-frequency areas of the image (the sky in Figure 2a). In the higher-frequency areas of the image (busy areas such as the leaves of Figure 5a and lamp of Figure 2a) the possible brightness increase is between 3% and 4%. Applying the gradual mask (Figure 1b) produces a much "deeper" watermark. In low-frequency areas of the image (the sky and blurred background of Figure 2a) the brightness difference can be increased to 6% and in the higher-frequency areas, the brightness difference can be increased to 20%. The gradual shading and removal of sharp borders of the masks contribute to this approach. An example of applying a "gradual" mask is shown in Figure 2. The luminance of the sky was increased by 2%, the luminance of the blurred background was increased by 10-15%, and the luminance of the leaves and the lamp was increased by 25-30%. An interesting note is that this watermark is clearly visible after lossy

(a) (b) (c)

Figure 2: (a) The original image with no watermark. (b) The watermarked image. (c) The difference between the original and watermarked images.



(a) (b)

Figure 3: The (a) flat and (b) gradual masks.

JPEG compression to 96dpi and 15% quality.

The cross-section profiles of the two masks in Figure 3 can be compared by graphing the gray values across column coordinates. Figure 4 shows profiles of flat and gradual masks of the same size and varying luminance values. The thin lines represent the gradual masks at 30%, 20%, and 10% of luminance increase. The thick lines represent the flat masks at 5%, 2%, and 1% of luminance increase. The spatial image gradient (gray level change per pixel) is fairly similar for the 30% gradual mask and the 5% flat mask; its value is between 4 and 5 gray levels per pixel, because the human visual system responds to edges and sudden changes in image brightness. As a consequence, the gradual mask produces a better (stronger) watermark.

Figure 4: The graph represents the gray values of cross-section profiles of the flat and gradual masks of Figure 3 at different intensities.

## 3 Attacks on Watermarks

Objectives of attacks against watermarks and embedded information include rendering a watermark unreadable, revealing the existence of hidden information, or confusing the reader as to the authenticity of the watermark. Attacks on watermarks may be accidental or intentional. Accidental attacks may be the result of standard image processing or compression procedures. Illicit attacks may include dictionary attack, steganalysis, special image processing techniques, or other attempts to overwrite or remove existing watermarks [JJ98b, PAK98, JJ98c].

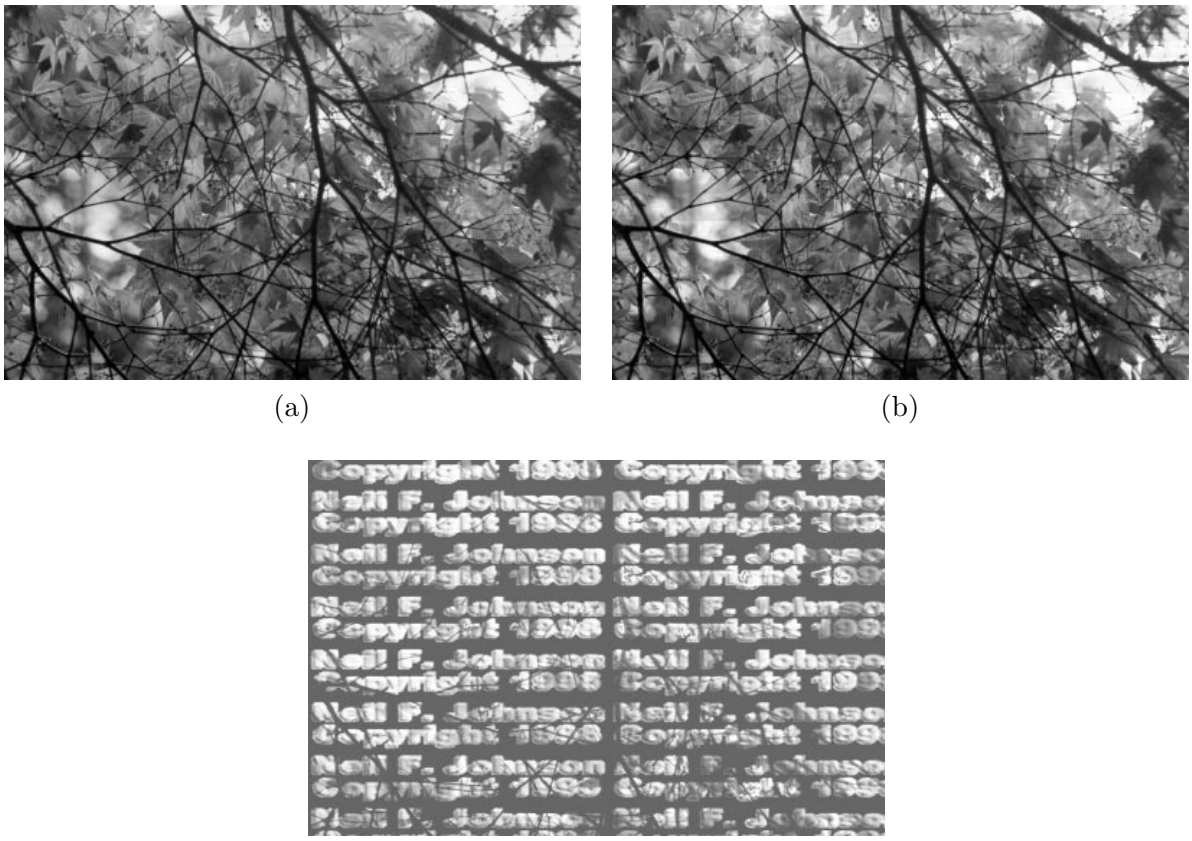
Compressing the carrier of the watermark may inadvertently render the watermark useless. This is especially likely for the bit-wise image domain methods used for watermarking. The original message can be reconstructed exactly if a lossless compression method is used. Lossy compression methods may yield better compression, but may not maintain the integrity of the original image [JJ98].

Some watermarking methods require a password to embed and detect watermarks and may be vulnerable to brute-force or dictionary attacks [Fli98, Mrf98]. Historically, the term "dictionary attack" refers to finding passwords by checking a list of terms. With improved processor speeds, a brute-force approach can find passwords by exhaustive search instead of using a dictionary list [EFF98]. Brute-force and dictionary attacks are general threats to passwords and are quite successful against poorly chosen passwords. Since the passwords used in watermarking are typically small by cryptographic standards, they can often be identified by guessing character combinations until the correct guess is made.

Other attacks may be achieved through the same means used to create watermarks. Image processing and transformations are commonly employed to develop and apply digital watermarks. These methods can also be used to attack and disable watermarks. Even with advances in watermarking technology, watermarks may be forged or overwritten; for example, multiple watermarks may be placed in an image and one cannot determine which of them is valid [CMYY97]. Current watermark registration services are "first come, first served", and someone other than the owner of a digital work may attempt to register a copyright first. Some watermarking tools are distributed with over-the-shelf software, such as Adobe Photoshop$^{TM}$ [Dig]. An exploitation of how to "crack" such a watermarking tool describes how to watermark any image with someone else's ID, or to overwrite valid watermarks with "forged" ones [Afp97]. If humanly imperceptible information is embedded in a cover, then humanly imperceptible alterations can be made to the cover which adversely affect this embedded information [JJ98b].

Many owners of watermarked works do not want the watermark to interfere with the use of the work by others. They therefore require that the watermark be imperceptible to the human sensory system. ("Human sensory system" is used here as an extension to human visual system - HVS, since media may pertain to senses other than vision.) This requirement works against the robustness of a watermark. Nevertheless, watermark users usually advertise the fact that a watermark exists. To a hacker, this constitutes a challenge to bypass the watermark.

Disabling a watermark or embedded message is fairly easy in cases where bit-wise methods are used since these methods employ the LSBs of images, which are changed by lossy compression or small amounts of image processing. Disabling watermarks created through masks and transforms requires more effort since the watermark is integrated more fully into the cover. A goal of many transform methods is to make the hidden information (the watermark) such an integral part of the



(a) (b)

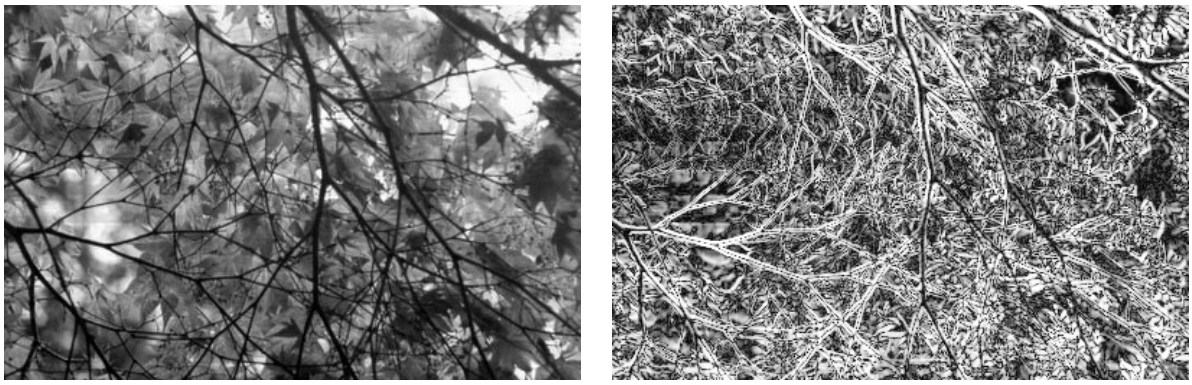
(c)

Figure 5: (a) An original image, (b) the watermarked image, and (c) the enhanced difference revealing the watermark.

image that the only way to remove or disable it is to destroy the marked image, rendering it useless to the attacker.

Attacks on transform watermarks are typically aimed against the watermark reader [CL97]. An attack may cause the reader to recognize a forged or counterfeited watermark [CMYY97]. Images may also be altered so the reader cannot see a watermark at all. The watermarked images are manipulated to a degree such that watermark cannot be recognized [PAK98, JJ98b]. Tools such as unZign and Stirmark can be use to test the robustness of watermarks [Unz, KP97]. These tools automate and apply image processing techniques and transforms to the watermarked image. The object is to show how vulnerable watermarks are to minor changes that have little visible effect on the images, but render the watermark unreadable. An illustration of such a process follows.

Figure 5 shows an example of a mask-based watermark. The image in Figure 5a is watermarked



(a) (b)

Figure 6: (a) The image distorted by Stirmark. (b) The enhanced difference between the distorted image and the original image.

using a mask to produce the image in Figure 5b. The watermark is not visible, but the enhanced image difference reveals it (Figure 5c). An attack on the watermark is conducted by applying Stirmark against the image in Figure 5b. Figure 6a shows the resulting "distorted" image after processing with Stirmark. Figure 6b shows the enhanced difference between the original image (Figure 5a) and the distorted image – the watermark is not visible. Stirmark applied image processing techniques of scaling, cropping, resampling, and rotating to interfere with the embedded mark.

## 4 Recovery from Attack

In the previous section we demonstrated how watermarks can be disabled using tools such as Stirmark. In this section we describe a method that can be used to recover watermarks in some cases. As with many watermark verification techniques, this method requires that the original be available for comparison. In general Stirmark applies unknown affine transforms to a watermarked image. The transformation is usually "small", but effectively hides most watermarks. Our recovery method applies an inverse transform to the distorted image $I'$ to reverse the distortion and recover the watermark. The method computes the displacement field between the original image $I$ and the distorted image $I'$; the field is then used to compute affine transformation parameters between the images. The inverse of the transformation is then applied to $I'$ resulting in an image in which the watermark can be detected. The details of the method and some experimental results are given

below.

### 4.1 Affine Transforms

Let $(x, y)$ be the pixel coordinates in an image $I(x, y)$ and let the image center be at $(0, 0)$. An affine transform of $I(x, y)$ is given by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \tag{1}$$

where $(x', y')$ are coordinates in the transformed image $I'(x', y')$ and $a - f$ are the transform parameters. Subtracting the vector $(x \;\; y)^T$ from both sides of equation (1), results in an expression for the displacement $(\delta x, \delta y)$ of the point $(x, y)$ due to the transform:

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \equiv \begin{pmatrix} x' - x \\ y' - y \end{pmatrix} = \begin{pmatrix} a - 1 & b \\ c & d - 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}. \tag{2}$$

### 4.2 Normal Displacement Fields

Let $\vec{\imath}$ and $\vec{\jmath}$ be the unit vectors in the $x$ and $y$ directions, respectively; $\delta\vec{r} = \vec{\imath}\delta x + \vec{\jmath}\delta y$ is the projected displacement field at the point $\vec{r} = x\vec{\imath} + y\vec{\jmath}$. If we choose a unit direction vector $\vec{n}_r = n_x\vec{\imath} + n_y\vec{\jmath}$ at the image point $\vec{r}$ and call it the normal direction, then the *normal displacement field* at $\vec{r}$ is $\delta\vec{r}_n = (\delta\vec{r} \cdot \vec{n}_r)\vec{n}_r = (n_x\delta x + n_y\delta y)\vec{n}_r$. $\vec{n}_r$ can be chosen in various ways; the usual choice (and the one that we use) is the direction of the image intensity gradient $\vec{n}_r = \nabla I / \|\nabla I\|$.

Note that the normal displacement field along an edge is orthogonal to the edge direction. If an edge element at position $\vec{r}$ is observed at time $t$, then the apparent position of that edge element at time $t + \Delta t$ will be $\vec{r} + \Delta t\delta\vec{r}_n$. This is a consequence of the well known *aperture problem* [J97]. Our method of estimating normal displacement field is based on this observation.

For an image frame (say collected at time $t$) edges are found using an implementation of the Canny edge detector [Can87]. For each edge element, say at $\vec{r}$, the image is resampled locally to obtain a small window with its rows parallel to the image gradient direction $\vec{n}_r = \nabla I / \|\nabla I\|$. A larger window is created for the next image frame (collected at time $t_0 + \Delta t$) we create a larger

window, typically twice as large as the maximum expected value of the magnitude of the normal displacement field. The first (smaller) window is compared with the second (larger) window by sliding the smaller window along the larger window and computing the difference between the image intensities. The zero of the resulting function is at distance $u_n$ from the origin of the second window; note that the image gradient in the second window at the positions close to $u_n$ must be positive. The estimate of the normal displacement field $(-u_n)$ is defined as the *normal flow*.

### 4.3 Computing Affine Transformation Parameters

Given an original image $I$ and a distorted image $I'$, such that $I'$ has been obtained by applying a unknown affine transform to $I$, the normal displacement field (the "normal flow") between $I$ and $I'$ is computed. Using Equation (2), the normal flow at image point $(x, y)$ is obtained as

$$\delta\vec{r}_n \cdot \vec{n}_r = n_x\delta x + n_y\delta y = a_1 n_x x + b n_x y + e n_x + c n_y x + d_1 n_x y + f n_y \equiv \mathbf{a}\cdot\mathbf{u} \tag{3}$$

where $\vec{n}_r = n_x\vec{\imath} + n_y\vec{\jmath}$ is the gradient direction at $(x, y)$, $\mathbf{a} = (n_x x \;\; n_x y \;\; n_x \;\; n_y x \;\; n_y y \;\; n_y)^T$, $a_1 = a - 1$, $d_1 = d - 1$, and $\mathbf{u} = (a_1 \;\; b \;\; c \;\; d_1 \;\; e \;\; f)^T$ is the vector of affine parameters. The method used to compute the normal flow, as described in Section 4.2, is applied. One normal flow value $u_{n,i}$ is used for each edge point $\vec{r}_i$ as the estimate of the normal displacement. This produces an approximate equation $\mathbf{a}_i\cdot\mathbf{u} \approx u_{n,i}$. Let the number of edge points be $N \geq 6$. We then have a system $\mathbf{A}\mathbf{u} - \mathbf{b} = E$, where $\mathbf{u}$ is an $N$-element array with elements $u_{n,i}$, $\mathbf{A}$ is an $N \times 6$ matrix with rows $\mathbf{a}_i$, and $E$ is an $N$-element error vector. We seek $\mathbf{u}$ that minimizes $\|E\| = \|\mathbf{b} - \mathbf{A}\mathbf{u}\|$; the solution satisfies the system $\mathbf{A}^T\mathbf{A}\mathbf{u} = \mathbf{A}^T\mathbf{b}$ and corresponds to the linear least squares solution [Ste73].

Given the estimate $\mathbf{u}$, based on the normal displacement field between images $I$ and $I'$, equation (1) is applied to obtain the inverse affine transform of $I'$ resulting in a "corrected frame" $I^{(1)}$. [The inversion of (1) is obtained implicitly. For each pixel position $(x, y)$ of $I^{(1)}$, the pixel position $(x', y')$ in $I'$ is computed (note that $x'$ and $y'$ may be non-integers). The gray level for $(x, y)$ is obtained by interpolating the gray levels of transformed image $I'$.] After obtaining $I^{(1)}$ the residual normal flow between $I$ and $I^{(1)}$ is computed. This residual is used to estimate the affine transform parameters $\mathbf{u}'$ between $I$ and $I^{(1)}$. If the estimated parameters are "small" we can stop and use $I^{(1)}$ as the recovered image; otherwise we can use $\mathbf{u}'$ to obtain $I^{(2)}$ from $I^{(1)}$. We say that an affine transform is small when applying Equation 2 to an image $I$ when we have $\max\{|\delta x|, |\delta y|\} < \varepsilon$; $\varepsilon$ is

typically 0.5.

### 4.4 Experimental Results

In this section, we provide two examples of experimental results. Figure 7 illustrates the recovery of a commercial watermark and Figure 8 illustrates a masked watermark recovery based on the attack shown in Figures 5 and 6. Figure 7a shows an original unwatermarked image ($I$). The image was watermarked using the version of Digimarc's PictureMark watermarking filter that is available with Adobe PhotoShop and distorted using Stirmark to produce the image in Figure 7b ($I'$). The watermark detection process was applied to $I'$ and the watermark could not be detected. The normal displacement field was computed between images $I$ and $I'$ (see Figure 7c) and the affine transformation parameters were estimated as $(0.0566 \;\; 0.0065 \;\; -0.0141 \;\; 0.0234 \;\; 3.0402 \;\; 3.6649)^T$. The inverse affine transform was applied to the image $I'$ to obtain the image $I^{(1)}$. The residual normal flow computed between images $I$ and $I^{(1)}$ (see Figure 7d) was used to estimate a new set of affine transformation parameters as $(0.0074 \;\; -0.0000 \;\; 0.0019 \;\; 0.0012 \;\; 0.6409 \;\; 0.3023)^T$. The inverse affine transform was applied to the image $I^{(1)}$ to obtain an image $I^{(2)}$. The residual normal flow computed between images $I$ and $I^{(2)}$ is shown in Figure 7e. Since the watermark was detected the process was stopped. Figure 7f shows the recovered image $I^{(2)}$.

Figure 8 illustrates the application of the method to the images in Figures 5 and 6. Figure 8a shows the normal flow computed between the original (Figures 5a) and the distorted watermarked image (Figures 6a). The affine transformation parameters were estimated as $(0.0255 \;\; 0.0012 \;\; -0.0049 \;\; 0.0045 \;\; 0.9685 \;\; 1.0939)^T$. The recovered image is shown in Figure 8b. Figure 8c shows the enhanced difference between the recovered image and the original image displaying the watermark.

## 5 Conclusion and Future Direction

Digital media are subject to illicit copying and distribution. The owners of such data are cautious about making their work available without some method of identifying ownership and copyright. Digital watermarks are currently being used in commercial applications to track the copyright and ownership of electronic works.

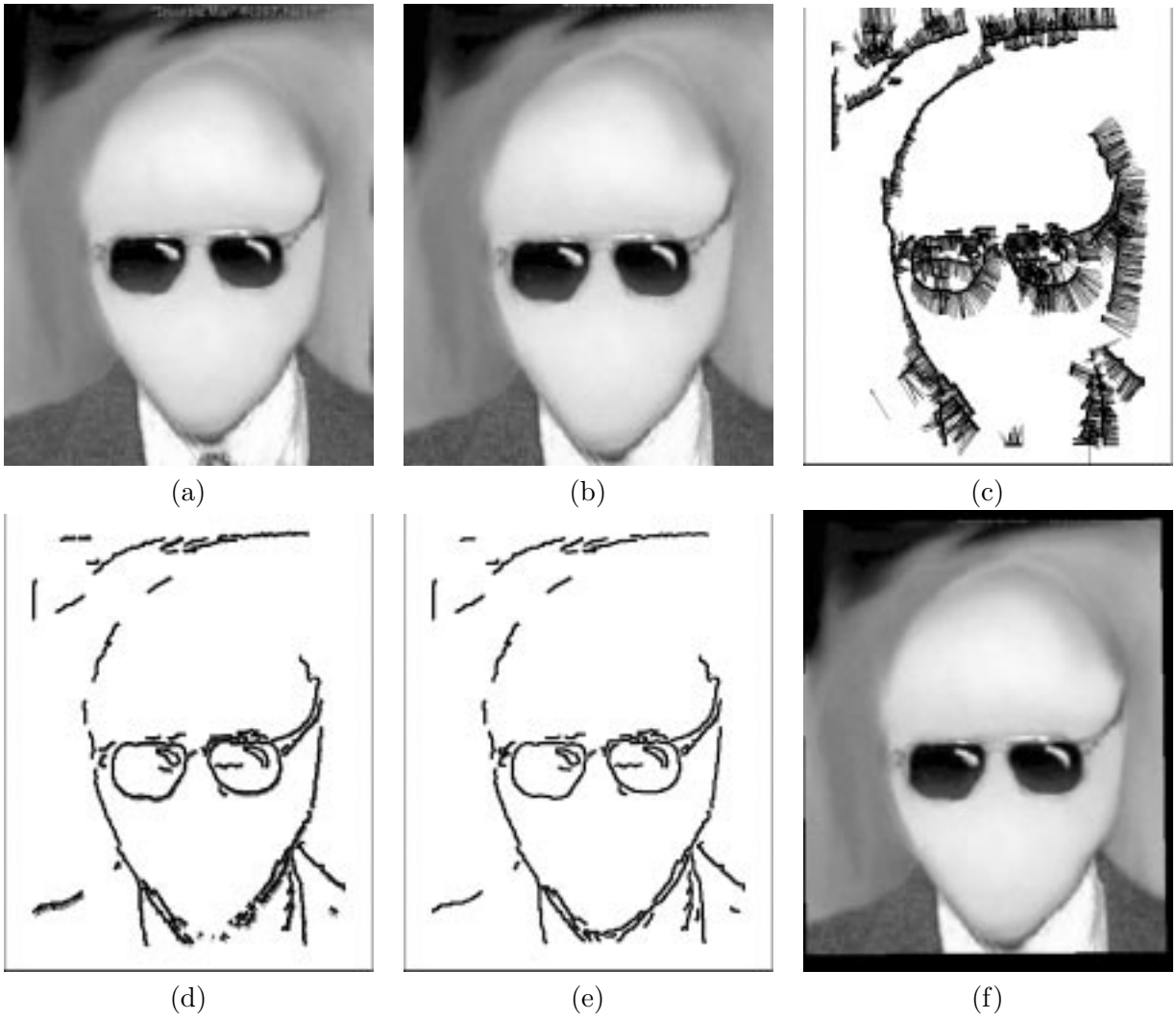
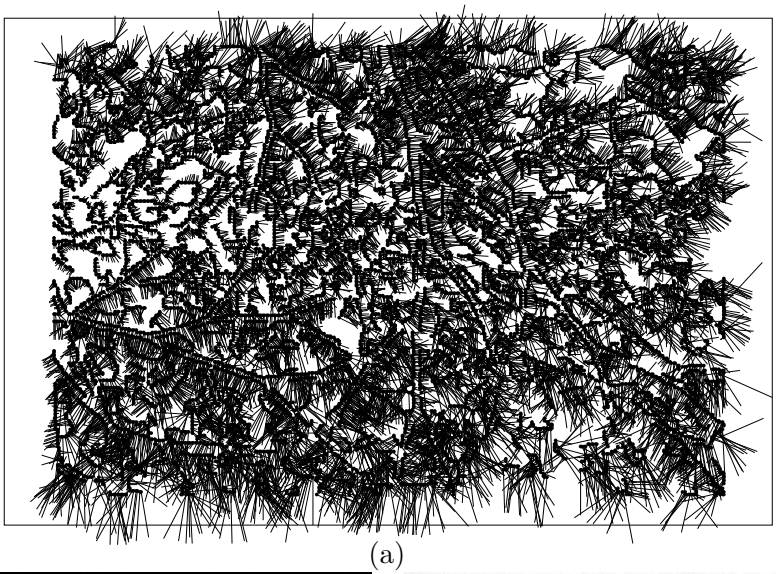
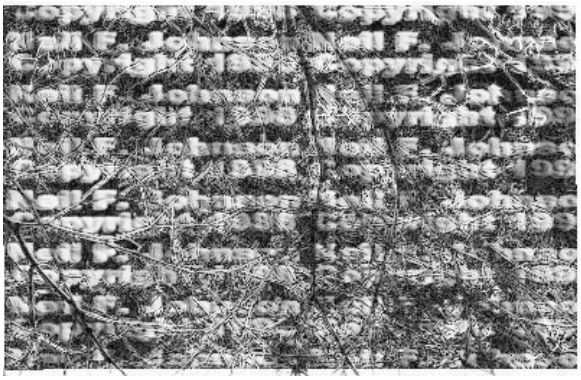


Figure 7: An illustration of the recovery method: (a) An original image. (b) A watermarked and distorted image. (c-e) Normal flows computed in various stages of the recovery process. (f) Recovered image in which the watermark can be detected.



(a)

(a)



(b)

Figure 8: (a) The normal flow field computed between the unwatermarked image in Figure 5a and the distorted watermarked image in Figure 6a. (b) The recovered image. (c) The enhanced difference between the recovered image and the original image showing the watermark.

We introduced desirable characteristics of watermarks, various methods of watermarking, and illustrated an alternative watermarking technique through the use of gradual masks. An important characteristic of an image-based watermark is it should be robust to common image processing and compression techniques, as well as to malicious attacks. To achieve this, the watermark should be placed in perceptually significant regions of the image.

Understanding and investigating the limitations of watermarking applications can help direct researchers to better, more robust solutions to ensure the survivability of embedded information such as copyright and licensing information. Methods that test the survivability of watermarks are essential for the development of stronger watermarking techniques [Unz, KP97]. Using such methods, as described in [JJ98b] and [PAK98], potential users of digital watermarking can see how much (or how little) effort is required to make a watermark unreadable by the watermarking tools. Given the ease of disabling watermarks, we provided an introduction to a method for watermark recovery after disabling attacks.

Digital watermarking does not prevent copying. Its effectiveness depends on providing evidence of illicit copying and dissemination by detecting watermarks in stolen images. Individual examination of suspicious images can be very costly, but it is unavoidable in cases where human intervention is needed to restore a seriously distorted image in order to detect the embedded watermark.

To date, the use of digital watermarks in images has not been recognized in legal cases for proving copyright and ownership. The cases that have gone to court have been resolved by physical evidence (negatives and photographic proofs). However, authors of purely digital media may not have such tangible evidence as claims of ownership.

An inherent weakness of many watermark approaches is the advertisement that an invisible watermark exists in a file. If the embedded message is not advertised, casual users will not know it exists and therefore will not attempt to remove it. Advertising the fact that hidden information exists raises the curiosity level of hackers who may see it as a challenge to overcome.

Further work is required to develop solutions to these problems and develop more robust watermarking techniques. An area for development is in watermark detection and third party authentication. Public-key steganography has been proposed in [Crav98]. Such an approach may be well suited for authentication of digital watermarks without relying on the original images or on proprietary watermark readers. Attacks against watermarking methods are being carefully considered in

current development of watermarking tools [Dig, Brau97]. Potential areas of research are in image recognition to locate images that have been damaged by attacks, and watermark recovery (and image reconstruction) from altered or damaged images [DJJ98]. We have provided some insight into this area and have illustrated that it is indeed possible.

## References


[And96] Anderson, R., (ed.), Information Hiding: First International Workshop, Cambridge, UK. Lecture Notes in Computer Science, Vol. 1174. Springer-Verlag, 1996.

[And96b] Anderson, R., Stretching the Limits of Steganography. In [And96], 39–48, 1996.

[AP98] Anderson, R., Petitcolas, F., On the Limits of Steganography, IEEE Journal on Selected Areas in Communications, 16(4): 474–481, 1998.

[Afp97] Anonymous (alias: Frog's Print, zguan.bbs@bbs.ntu.edu.tw), PhotoShop 4.0/Digimarc: Commercial stupidity–Digimarc downfall, http://www.fravia.org/frogdigi.htm and mirrored at http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/digimarc_crack.html. Original post in Learn Cracking IV on news:tw.bbs.comp.hacker, August 1997.

[Mrf98] Anonymous (alias: mrf), How to reverse engineer Steganos (First Step): Speed up brute force cracking, http://www.fravia.org/mrf_steg.htm, February 1998.

[Auc98] Aucsmith, D. (ed.), Information Hiding, Second International Workshop, Portland, Oregon, USA, April 1998, Lecture Notes in Computer Science, vol. 1525, Springer-Verlag, 1998.

[BGML96] Bender, W., Gruhl, D., Morimoto, N., Lu, A., Techniques for Data Hiding. IBM Systems Journal, 35(3-4): 313–336, 1996.

[BLMO94] Brassil, J., Low, S., Maxemchuk, N., O'Gorman, L., Electronic Marking and Identification Techniques to Discourage Document Copying. In Infocom94, 1994. ftp://ftp.research.att.com/dist/brassil/1994/infwocom94a.ps.Z.

[BOML95] Brassil, J., O'Gorman, L., Maxemchuk, N., Low, S., Document Marking and Identification using both Line and Word Shifting. In Infocom95, Boston, MA, April 1995, 853–860.

[Brau97] Braudaway, G.W., Protecting Publicly-Available Images with an Invisible Image Watermark. In [ICIP97].

[BS95] Brown, W., Shepherd, B.J., Graphics File Formats: Reference and Guide. Manning Publications, Greenwich, CT, 1995.

[Can87] Canny, J., A Computational Approach to Edge Detection, in M. A. Fischler and O. Firschein, editors, Readings in Computer Vision: Issues, Problems, Principles and Paradigms. Morgan Kaufmann, 1987.

[Car95] Caronni, G., Assuring Ownership Rights for Digital Images. Reliable IT Systems, Vieweg publications, Wiesbaden, 1995.

[CKLS95] Cox, I., Kilian, J., Leighton, T., Shamoon, T., Secure Spread Spectrum Watermarking for Multimedia. Technical Report 95-10, NEC Research Institute, 1995. An updated version of this paper with the same title is in IEEE Transactions on Image Processing, 6(12): 1673–1687, 1997.

[CKSL96] Cox, I., Kilian, J., Shamoon, T., Leighton, T., A Secure, Robust Watermark for Multimedia. In [And96], 185–206.

[CL97] Cox, I., Linnartz, J., Public Watermarks and their Resistance to Tampering. In [ICIP97].

[CMYY97] Craver, S., Memon, N., Yeo, B., Yeung, N.M., Resolving Rightful Ownerships with Invisible Watermarking Techniques. Research Report RC 20755 (91985), Computer Science/Mathematics, IBM Research Division, 1997. An updated version of this paper with the same title is in IEEE Journal on Selected Areas in Communications, 16(4): 573–586, 1998.

[Crav98] Craver, S., On Public-Key Steganography in the Presence of an Active Warden. In [Auc98], 39–48.

[Dig] Digimarc Corporation, PictureMarc$^{TM}$, MarcSpider$^{TM}$, http://www.digimarc.com

[DJJ98] Duric, Z., Johnson, N.F., Jajodia, S., Recovering Watermarks from Images. Internal technical report, Center for Secure Information Systems, George Mason University, 1998.

[EFF98] Electronic Frontier Foundation (EFF), Deep Crack, a hardware Data Encryption Standard (DES) cracker that cracked a 56-bit DES key in four hours. Prior to the construction of this $250,000 computer system, estimated cost of building such a machine was in the tens of millions. http://www.eff.org/descracker/, 1998.

[FB97] Ford, W., Baum, M., Secure Electronic Commerce. Prentice Hall, Upper Saddle River, NJ, 1997.

[FH97] Fleet, D., Heeger, D., Embedding Invisible Information in Color Images. In [ICIP97],

[Fli98] Flinn, J. (the_flynn@hotmail.com). A Journey within Steganos, on http://www.fravia.org/fly__01.htm, February 1998.

[FvDFH96] Foley, J.D., van Dam, A. Feiner, S.K., and Hughes, J.F., Computer Graphics: Principles and Practice, 2nd ed., Addison-Wesley, Reading, MA, 1996.

[GB98] Grhul, D., Bender, W., Information Hiding to Foil the Casual Counterfeiter. In [Auc98], 1–15.

[GW92] Gonzalez, R.C., Woods, R.E., Digital Image Processing. Addison-Wesley. Reading, MA, 1992.

[HW96] Hsu, C, Wu, J., Hidden Signatures in Images. In [ICIP96].

[HYQ98] Hou, S., Yvo, D., Quisquater, J., Cerebral Cryptography. In [Auc98].

[ICIP96] IEEE International Conference on Image Processing, Lausanne, Switzerland, September 16–19, 1996.

[ICIP97] IEEE International Conference on Image Processing, Santa Barbara, CA, October 26-29, 1997.

[J97] Jähne, B., Digital Image Processing: Concepts, Algorithms, and Scientific Applications, 4th ed. Springer-Verlag, Berlin, 1997.

[JJ98] Johnson, N.F., Jajodia, S., Exploring Steganography: Seeing the Unseen. IEEE Computer, 31(2): 26–34, 1998.

[JJ98b] Johnson, N.F., Jajodia, S., Steganalysis of Images Created using Current Steganography Software. In [Auc98], 273–289.

[JJ98c] Johnson, N.F. and S. Jajodia: Steganalysis: The Investigation of Hidden Information, Proceedings of the IEEE Information Technology Conference, Syracuse, New York, USA, September 1–3, 113–116, 1998.

[KH97] Kundur, D., Hatzinakos , D., A Robust Digital Image Watermarking Method Using Wavelet-Based Fusion. In [ICIP97].

[KRZ94] Koch, E., Rindfrey, J., Zhao, J., Copyright Protection for Multimedia Data. Proceedings of the International Conference on Digital Media and Electronic Publishing, Leeds, UK, 1994.

[KP97] Kuhn, M. and Petitcolas, F., StirMark, Tool for evaluating watermarks. http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/stirmark, 1997.

[KZ95] Koch, E., Zhao, J., Towards Robust and Hidden Image Copyright Labelling. Proceedings of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing, Neos Marmaras, Greece, 452–455, 1995.

[LMP98] Lach, J., Mangione-Smith, W., Potkonjak, M., Fingerprinting Digital Circuits on Programmable Hardware. In [Auc98].

[MBR98] Marvel, L.M., C.G. Boncelet, Jr., and C.T. Retter: Reliable Blind Information Hiding for Images, in [Auc98], 48–61.

[ODB96] O Ruanaidh, J., Dowling, W., Boland, F., Phase watermarking of digital images. In [ICIP96].

[OP97] O Ruanaidh, J., Pun, T., Rotation, Scale and Translation Invariant Digital Image Watermarking. In [ICIP97].

[PAK98] Petitcolas, F., Anderson, R., Kuhn, M., Attacks on Copyright Marking Systems. In [Auc98], 218–238.

[PBBC97] Piva, A., Barni, M., Bartolini, F., Cappellini, V., DCT-Based Watermark Recovering Without Resorting to the Uncorrupted Original Image. In [ICIP97], 520–523.

[Ros92] Rosenfeld, A., Models: The Graphics-Vision Interface. In T.L. Kunii, ed., Visual Computing, Springer, Tokyo, 1992, 21–23.

[SC96] Smith, J., Comiskey, B., Modulation and Information Hiding in Images. In [And96], 207–226.

[Sch96] Schneier, B., Applied Cryptography, Second ed. John Wiley & Sons, New York, 1996.

[Sig] Signum Technologies, SureSign, http://www.signumtech.com/

[Ste73] G. W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

[SZT96] Swanson, M., Zhu, B., Tewfik, A., Transparent Robust Image Watermarking. In [ICIP96], 211–214.

[TNM90] Tanaka, K., Nakamura, Y., and Matsui, K., Embedding Secret Information into a Dithered Multi-Level Image. IEEE Military Communications Conference, 216–220, 1990.

[Unz] Anonymous: unZign, Tool for Evaluating Watermarks. http://altern.org/watermark/, 1997.

[XBA97] Xia, X, Boncelet, C.G., Arce, G.R., A Multiresolution Watermark for Digital Images. In [ICIP97].