This paper appears in the Proceedings of the SANS Intrusion Detection and Response Conference (IDR'99) held in San Diego, CA, February 9-13, 1999.

An Introduction to Watermark Recovery from Images

Neil F. Johnson*

Center for Secure Information Systems
George Mason University
Fairfax, VA 220304444
http://isse.gmu.edu/~csis
njohnson@gmu.edu

Abstract: With the proliferation of the World Wide Web, authors of digital media now have an inexpensive means to distribute their works to a growing audience. Many authors are leery of distributing their works in fear that it may be copied illegally or represented as another's work. Digital watermarks provide means of placing additional information within digital media so if copies are made, the rightful ownership may be determined. Those wishing to make illicit copies of the digital can employ a number of methods against watermarks so the embedded information cannot be detected or read. In this paper we briefly discuss a method of recovering watermarks in digital images after such attacks and introduce related current and future work in the Center for Secure Information Systems (CSIS).

1 Introduction

With onset of the World Wide Web, authors of digital media can easily distribute their works by making them available on Web pages or other public forums. Anyone having access to those forums can copy the author's media. By the nature of digital media, a copy is an exact, perfect duplicate of the original. This brings to front a potential problem. How do authors claim ownership rights of such digital media if multiple persons have exact copies?[1] One method is to embed additional information and only distribute the media that contains this additional information. The embedded information is known as a watermark can provide, for example, information about the media, the author, copyright, or license information.

Interest in digital watermarks has grown out of an increasing interest in intellectual property and copyright protection. Digital watermarks may be perceptible (visible) or imperceptible (invisible) to human vision. Visible watermarks, by nature, are more intrusive to the media and act to deter theft of the media, such as a warning sign announces an alarm system even if one does not exist. Examples of such watermarks can be seen easily on most network television stations by the station's logo in the corner of the viewable screen. These watermarks are typically confined to an area of the image, which is less intrusive to the overall image. Attackers have a visible target and can remove the watermark by cropping the image.

---

* Also with the Department of Information and Software Engineering in the School of Information Technology Engineering

[1] Digital media has unique characteristics not found in other media. Though the proof of ownership of digital copies of photographs can be resolved by presenting negatives, authors of purely digital media may not have such tangible evidence. In this paper, we will use digital copies of photographs in examples to represent digital media.

Invisible watermarks have an advantage over visible watermarks, in that their location may be unknown. A common practice is to distribute the watermark (or watermarks) across the entire image. This provides some protection against cropping attacks. However, the less perceptible a watermark is, it may be more vulnerable to manipulation. Assume an image (I) is composed two types of data based on the human visible threshold. These types are visible data (v) and invisible data (w). Thus, an image can be defined as I = v + w. To further define these types, any manipulation to (v) will result is noticeable distortion in the image. Modifying (w) will not be noticeable. The size of (w) is available to both the owner and attacker. Since (w) remains imperceptible, there exist some (w') such that I' = v + w' and there is not perceptible difference between I and I'. An attack may be to replace, remove, or distort (w). One such attack described in [1] discusses adding illicit watermarks as means to counterfeit valid watermarks. If information is added to some media such that the added information cannot be detected, then there exists some amount of additional information that may be added or removed within the same threshold, which will overwrite or disable the embedded information. If the attacker is intent on disabling the watermark, this can be easily done [2,3]. One way around this is to produce a more perceptible watermark thus impacting some part of the visible portion of the image (v).

Note: Though the remainder of the paper emphasizes images in discussion and examples, please bear in mind that watermarking and recovery processes apply to a number of other mediums and signals such as text, audio, graphics, multimedia, signal processing, and telecommunications.

2 Attacks

Attacks on watermark may not necessarily remove the watermark, but disable its readability. Image processing and transforms are commonly employed to create and apply watermarks. These same techniques can also be used to disable or overwrite watermarks. Multiple watermarks can be placed in an image and one cannot determine which one is valid [1]. Currently watermark registration service is "first come, first served." Someone other than the rightful owner may attempt to register a copyright first. Figure 1 illustrates applying image processing techniques (skew, warp, blur, and rotate) to attack a mask-based watermark. These processing techniques have been automated in tools available on the Internet [4,5].

(a) Original Image

(b) Watermarked Image

(c) Watermark - difference between (a) and (b)

(d) StirMark is used to process (c) (e) Difference between (a) and (d)

Figure 1: illustration of an attack on a watermark using StirMark.

3 Countermeasures

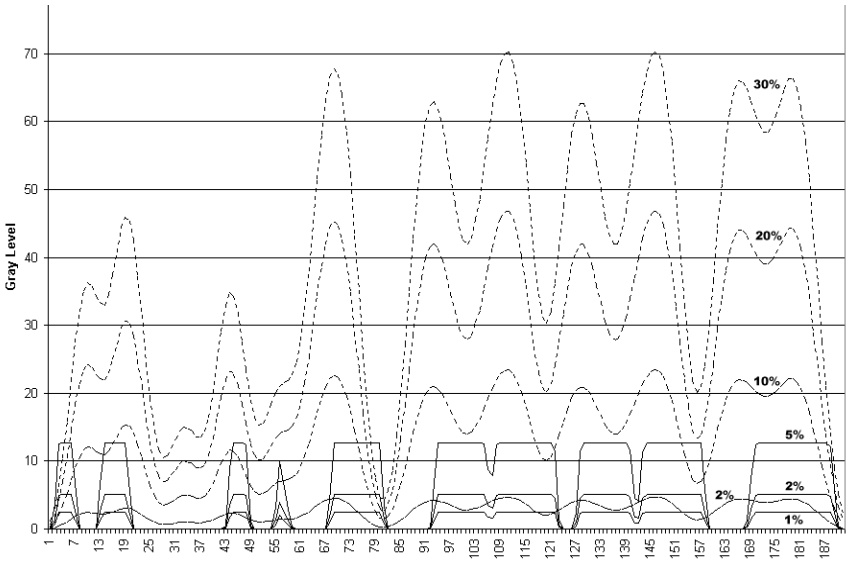
Granted, this image is very busy and 'you cannot see the watermark for the trees.' The watermark appears to be lost. What can be done to counter attacks on watermarks? We have several options available. Depending upon the image, a stronger watermark may be a viable solution and can survive some image processing. If an image is processed to the degree that the watermark cannot be recognized (see Figure 1), then reconstruction of the image properties may be possible through the use of an original image. This reconstruction recovers features of the image that may have been lost including the watermark. More information about image recognition watermark recovery is covered in [6].

Stronger Watermarks

The human eye is drawn to patterns such as lines and edges in images. A watermark that is composed of sharp edges is more likely to be visible than one with smoothed edges [7,8]. Let us look at an example (see Figure 2).

(a)

(b)



(c)

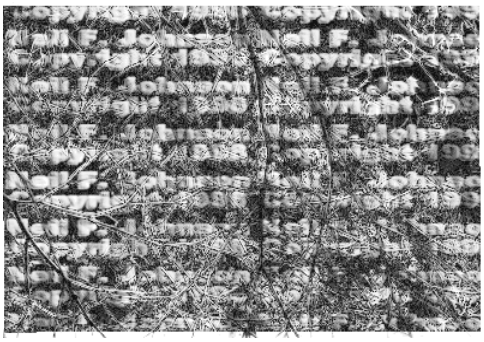
Figure 2: The top two images are masks to create a "©1998" watermark in an image; (a) a “sharp” mask and a (b) “gradual” mask. (c) The graph represents the gray values of a latitudinal cross-section of each mask at different intensities. The intensity percentage associated with each line plot in the graph is roughly the amount of luminance applied to the mask as a watermark.

A watermark is created from the "sharp" mask (Figure 2a) by increasing the luminance. At an increase of about 5%, the watermark starts to become visible in busy areas of an image on a high-resolution computer monitor. In relatively flat areas of an image (i.e.; a clear sky) only a 1% increase is possible before becoming visible. If the "gradual" mask (Figure 2b) is applied to the same image by increasing luminance, the watermark is not visible until nearly a 30% increase in luminance. This produces a watermark that is more resistant to the changes of lower bits. However, given enough image processing, these are also vulnerable, but the resulting image may not be usable to the attacker.

Watermark Recovery

In instances where the embedded watermark cannot be read, another approach is to attempt to recover the watermarks from damaged images. The image size and aspect can be recovered by applying the displacement between the original and damaged images. The features of the damaged image are “refined” toward those of the original image. Details of the creation of the parameters for recovery and refinement from the corresponding points between the original image and the damaged image is described in [6] and beyond the scope of this paper. An example of the recovery is shown in Figure 3.

(a)   (b)



(c) Recovered watermark

Figure 3: Watermark recovery: (a) The difference between the "original" (Figure 1a) and "attacked" image (Figure 1d). (b) The “recovered” image is created through computing the differences between the original and the “attacked” image. (c) The recovered watermark seen by the enhanced difference between the "original" image and (b)

4 Conclusion and Future Work

The use of tools to test the survivability of watermarks is necessary to understand the limitations of existing techniques and to nudge us to develop stronger watermarking methods [4,5]. Using these tools and methods described in [2] and [3], potential customers of digital watermarking can see how much (or little) effort is required to disable a watermark.

Further work is necessary to improve the reliability of watermark systems to protect intellectual property and copyrights. Attacks on watermarks are being considered in current development of watermarking tools [9,10]. Areas for development include watermark detection, recovery, and authentication. One possible approach for authentication is to apply public-key steganography as introduced in [11] and further explored in [12].

The intent of this paper is to provide a high-level, introduction to the watermark recovery we are pursuing and document preliminary results. This work will be further detailed in future papers [8,6]. We are currently expanding the ideas introduced here to include automatic image recognition, image refinement in the recovery phase, and the investigation of the invariant properties between point clusters between images (some of which is touched on in [6])

5 References:

[1] S. Craver, N. Memon, B. Yeo, N.M. Yeung, "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, attacks, and implications." IEEE Journal on Selected Areas in Communications, vol. 16 no. 4, pp. 573-586 (1998)

[2] N.F. Johnson, S. Jajodia, "Steganalysis of Images Created using Current Steganography Software," in [13], pp. 273-289 (1998)

[3] F. Petitcolas, R. Anderson, M. Kuhn, "Attacks on Copyright Marking Systems," in [13], pp. 218-238 (1998)

[5] Anonymous, unZign, Tool for evaluating a variety of watermarks, http://altern.org/watermark/, (1997)

[4] F. Petitcolas, StirMark v. 2.x, Tool for evaluating a variety of watermarks based on the initial version (StirMark 1.0) by Markus Kuhn, http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/stirmark/ (1998)

[6] Z. Duric, N.F. Johnson, S. Jajodia, "Recovering Watermarks from Images," work in progress, Center for Secure Information Systems, George Mason University. Publication TBA.

[7] D. Grhul, W. Bender, “Informaiton Hiding to Foil the Casual Counterfeiter,” in [13], pp. 1-15 (1998)

[8] N.F. Johnson, Z. Duric, S. Jajodia, "A Role of Digital Watermarking in Electronic Commerce," accepted for publication in special issue of the ACM on Electronic Commerce in 1999.

[9] Digimarc Corporation, http://www.digimarc.com
[10] G.W. Braudaway, “Protecting Publicly-Available Images with an Invisible Watermark,” Proceedings of the (ICIP97) IEEE International Conference on Image Processing, Santa Barbara, CA, USA (1997)
[11] R. Anderson, “Stretching the Limits of Steganography,” Lecture Notes in Computer Science, vol. 1525, Springer-Verlag, pp. 39-48 (1997)
[12] S.Craver, “On Public-Key Steganography in the Presence of an Active Warden,” in [13], pp. 355-368 (1998)
[13] D. Aucsmith (Ed.), "Information Hiding," Second International Workshop, IH'98, Portland, Oregon, USA, April 1998, Proceedings. Lecture Notes in Computer Science, vol. 1525, Springer-Verlag (1998)