Effective and Ineffective Digital Watermarks

Fred Mintzer, Gordon W. Braudaway and Minerva M. Yeung
IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598
Emails: {mintzer,braud,yeung}@watson.ibm.com

Abstract

We have entered an era where inexpensive and readily-available equipment can produce perfect copies of digital multimedia materials, such as CD-quality audio, publication-quality images, or digital video. In this environment, it has become easier for malicious parties to make salable copies of copyrighted content without compensation to the content owner. Many media content owners are concerned about the potential loss of revenue from multimedia piracy, especially when the content will be exposed to the Internet. Digital watermarking is seen by many as a potential solution to this problem.

Many different watermarking schemes have been proposed. Often, however, there is little discussion of how effective a proposed watermarking technique may be at solving a particular problem. In this paper, we describe a number of proposed image-watermarking application scenarios and form a small number of watermark-application categories. Then, with these applications in mind, we discuss the desired technical properties of watermarks for each category. Finally, we discuss some watermarking techniques developed by the authors, in light of the desired properties.

1 Introduction

The field of digital watermarking is rather new; indeed, at this point many of its terms are not well defined. We define *watermarking* to be a process that embeds data, called *a watermark*. into a multimedia object to help protect the owner's rights to that object. The watermark may be either visible or invisible.

We define *visible watermarking* to be a process that embeds data that is intentionally perceptible to a human observer; *invisible watermarking* to be a process that embeds data that is not perceptible, but may be extracted by a computer program. Although "invisible" and "watermark" are visual terms, invisible watermarking is not limited to marking images. There is interest (in both the commercial and technical communities) in applying invisible watermarks to digital audio, still-frame images of various types, and digital video.

2 Some Proposed Applications of Image Watermarks

In this section, we describe a number of possible scenarios for applications of image watermarking. In composing the scenarios, we have taken into consideration the concerns and comments expressed by content owners, and by others involved in the field, about the problems that content owners encounter.

- **Scenario A: visible watermarking for enhanced copyright protection**.

  In this scenario, images are made available through the Internet, and the content owner is concerned that the images will be used commercially (e.g., imprinting coffee mugs) without payment of royalties. Here, the content owner desires an ownership mark that is visually-apparent, but which does not prevent the image from being used for other purposes (e.g., scholarly research). The underlying rationale is that the visibility of mark will make apparent any commercial use of the images, and hence make it easier to enforce the copyright (and collect licensing revenue).

- **Scenario B: visible watermarking used to indicate ownership of originals**.

  In this scenario, images are made available through the Internet, and the content owner desires to indicate the ownership of the underlying materials (e.g., the library that owns the manuscript), so an observer might be encouraged to patronize the institution that owns the materials. Here, the content owner desires a visible mark that makes clear the source of the materials. Loss of revenue is a lesser concern than for A.

- **Scenario C: invisible watermarking for a trustworthy camera**.

  In this scenario, images are captured with a digital camera for later inclusion in news articles. Here, it is the desire of a news agency to verify that an image is true to the original capture, and has not been edited to falsify a scene. In this case, an invisible watermark is embedded at capture time; its presence at the time of publication is intended to indicate that the image has not been altered since it was captured. This scenario has also been described in [1].

- **Scenario D: invisible watermarking to detect alteration of images stored in a digital library**.

  In this scenario, images (e.g., human fingerprints) have been scanned and stored in a digital library; the content owner desires the ability to detect any alteration

of the images, without the need to compare the images to the scanned materials. Here, the underlying rationale is that the content owner will extract an invisible watermark from the image that will indicate whether that image has been altered or replaced since it was entered into the digital library. This is an especially keen desire when the digital library is exposed to an external network, such as the Internet.

- **Scenario E: invisible watermarking to detect misappropriated images**.

  In this scenario, the seller of digital images, a.k.a. photo clip art, is concerned that his fee-generating images may be purchased by an individual who will make them available for free; this would deprive the owner of licensing revenue. In this case, a "web crawler" is desired that would search images on web sites to look for the seller's watermark and determine whether the seller's images are being made available there.

- **Scenario F: invisible watermarking as evidence of ownership**.

  In this scenario, the seller of digital images suspects that one of his images has been edited and published without payment of royalties. Here, the detection of the seller's watermark in the image is intended to serve as evidence that the published image is property of the seller. This scenario, and some of its subtleties, has also been described in [2].

- **Scenario G: invisible watermarking to determine the identity of a misappropriator**.

  In this scenario, the seller of digital images suspects that one of his images has been edited and published without payment of royalties. Here, the seller adds an invisible watermark to his images, at distribution time, to indicate to whom they were sold. The extracted watermark is intended to reveal the identity of the buyer of the image that was published. This would permit the seller to discontinue business with the buyer because of the risk (of loss of assets) it entails.

- **Scenario H: invisible watermarking for a digital VCR**.

  In this scenario, an invisible watermark is embedded in MPEG-compressed video. The digital VCR looks for a "special watermark" to determine whether the video may be copied, or only played.

3 Clustering the Applications

Scenarios A and B can both be satisfied by visible image watermarks. Their underlying desired technical properties will be discussed in Section 4.

Scenarios C and D can both be satisfied by invisible image watermarks that will change, or disappear, if a watermarked image is altered. Indeed, for these applications, it is generally desired that the watermarks are very sensitive to many sorts of image processing. We call these watermarks *fragile invisible watermarks* because it is desired that they be altered or destroyed by most common image processing techniques; their desired technical properties will be discussed in Section 5.

Scenarios E, F and G can be be satisfied by invisible image watermarks that persist even if someone tries to remove them. Since they are desired to survive intentional attacks, we call them *robust image watermarks*. Their desired properties will be discussed in Section 6.

We note that Scenario H seems to bear some resemblance to E and F. However, since the authors' focus has been on high-quality still-frame imaging, and the requirements for H are still evolving, we will not discuss it further.

4 Desired Properties for Visible Watermarks

Scenarios A and B, above, share some common desired technical properties:

V 1. that the watermark is readily visible,
V 2. that the watermark is unobtrusive,
V 3. that the watermark is hard to remove, and
V 4. that the watermark may be applied automatically, with consistent visual prominence, to batches of diverse images.

Requirements V1, V2, and V3 have been discussed in our earlier papers [3, 4] and will not be more fully described.

Requirement V4 is not easily satisfied. It has been our experience that applying the same watermark to different images can result in varying results; the watermark may be quite prominent in one image, yet nearly invisible in others.

There is some difference in the properties desired of the watermarks needed to satisfy Scenarios A and B. In Scenario A, there is an economic incentive to remove the watermark; hence, its robustness to removal is more strongly desired. For the techniques of [5, 3, 4] this difference can generally be accommodated by applying a more prominent watermark for Scenario A applications than for Scenario B applications.

5 Desired Properties for Fragile Invisible Watermarks

Scenarios C and D, above, also share some common desired technical properties.

F 1. the watermark is invisible to a human observer,
F 2. the watermark is altered by the application of most common image processing techniques,
F 3. it is difficult for an unauthorized person to insert a false watermark,
F 4. the watermark can be quickly extracted by an authorized person,
F 5. the watermark survives image cropping, and
F 6. the extracted watermark indicates where alterations have taken place.

It is difficult, if not impossible, to determine whether a watermark is strictly invisible, as its invisibility depends on the viewer, the image, and how the image is presented (e.g., displayed/printed, under what illumination, with what

gamma). Still, there are some tests that we commonly use. They involve preparing images in which the invisible watermark is most likely to appear, and displaying them on a high-resolution monitor. These prepared images are:

1. an image composed by applying the watermark to a uniform gray image, (which reveals added texture in the absence of scene texture),
2. an image composed by applying the watermark to a source image with large dark areas (which reveals added texture in the dark regions where texture is often most visible),
3. an image composed by compositing the source image and the watermarked image, wherein one fills the lower left corner, one fills the upper right corner, and the dividing line runs from corner to corner (which permits the viewer to judge whether the image texture is different in the marked and unmarked images).

Property F3 addresses the concern that the watermark might be extracted from a marked image and inserted in a substitute image. For this condition to be met, it is desired that it be difficult for a malicious party to determine if and/or how an image has been marked. We favor marks that can only be extracted with a hidden data sequence (key) that can unlock them, where that key is stored separately in a secure database. This makes the watermarking hard to reverse engineer.

Property F5, we note, is desirable in some applications, where the image is intended to be cropped after marking. In other applications where this is not intended, it may not be desirable. But, whether this property is present or not, it is important that the watermark extraction also be able to detect whether cropping has occurred. Similarly F6 may be desired in some applications, but not others.

In [6] we report a technique for watermarking color and gray-scale images that possesses these properties. We note that the watermark is quite hard to uncover in the absence of its key.

6 Desired Properties for Robust Invisible Watermarks

Scenarios E, F, and G, above, share some common desired technical properties.

R 1. the watermark is invisible to a human observer,

R 2. the watermark remains in a watermarked image, even after it has been processed by common image processing techniques,

R 3. the watermark is hard to detect by an unauthorized person,

R 4. the watermark may be quickly extracted by an authorized person, and

R 5. after a watermarked image is printed and re-scanned, the watermark can still be extracted.

Designing watermarks that possess R3 and R5 is a daunting task, but a robust watermark is not of much use if it is easily removed. While we cannot specify all of the image processing attacks that can be envisioned, we can enumerate the image processing methods that we normally use to prepare high-quality images for display, printing, or/or transmission. They include

1. cropping,
2. brightness and contrast modifications,
3. sharpening, blurring, and other filtering operations,
4. enlargement, reduction, and rotation, and
5. lossy JPEG compression.

   In addition, we believe the watermark should survive intentional attacks that

6. add correlated or uncorrelated noise to the image.

Designing watermarking techniques that survive only these operations is a daunting task. Perhaps even more daunting is the developing the software that detects a watermark after the watermarked image has been subjected to cropping, reduction, and rotation. Many thousands of lines of software may be required to do the detection, and it may require considerable processing power to execute. Thus, there seems a natural conflict between R2 and R4. Indeed, there may be many useful robust watermarking methods that achieve different balances between these two desired properties.

We believe property R3, like F3, is often best achieved by using hidden marks that require a key to extract them. In [7], we report on a technique for watermarking color and gray-scale images that possesses these properties, including some ability to survive printing, R5.

7 Remarks

Misappropriation of assets is a great concern of multimedia content owners, especially when the content is to be made available through the Internet, which is not inherently secure. Watermarking is seen by many as a potential solution to this problem.

In this paper, we have listed a number of application scenarios for image watermarking. Some of these scenarios may turn out to not be commercially attractive or technically feasible. There are undoubtedly other scenarios that we have overlooked. Still, the set of scenarios given represent a set of potential applications for which desired properties can be described, and against which various proposed watermarking techniques may be judged.

Furthermore, we have clustered the scenarios into three watermark-application categories that we described as visible watermarking, fragile invisible watermarking, and robust invisible watermarking. Desired technical properties of each application group were also given. Some of the desired technical properties are quite subjective. In some cases, we presented some specific conditions that should be considered in claiming these properties. We have evaluated

three techniques of our creation [5,6,7] with respect to the desired properties reported.

We note that attaining the desired properties may involve different techniques for different types of images. The best fragile invisible watermark for a G4-compressed binary image may not be the best fragile watermark for an uncompressed high-resolution 36 bit-per-pixel color image. Still, we believe all proposed techniques would benefit from an evaluation of the extent to which the techniques possess the properties desired of the underlying application; only in this way can we judge whether proposed techniques are effective or ineffective.

Moreover, we believe watermarking is not a single technique that fits all situations, but a set of techniques that can each reduce the risk of misappropriation within a specific application domain. There are also other tools that can help protect digital assets; they include cryptography, time-stamping, and registration. Multiple watermarking techniques, used in conjunction with the other asset protection tools, will undoubted be required to best address the diverse problems of content protection.

References

[1] G.L. Friedman, "The trustworthy digital camera: Restoring credibility to the photographic image", *IEEE Transactions on Consumer Electronics*, vol. 39, pp. 905–910, Nov. 1993.

[2] S. Craver, N. Memon, B.L. Yeo, and M.M. Yeung, "Resolving rightful ownership with invisible watermarking techniques: Limitations, attacks and implications", accepted for publication *IEEE Journal of Selected Areas of Communications* (also IBM Research Report RC 20755, March 1997).

[3] F.C. Mintzer, L.E. Boyle, A.N. Cazes, B.S. Christian, S.C. Cox, F.P. Giordano, H.M. Gladney, J.C. Lee, M.L. Kelmanson, A.C. Lirani, K.A. Magerlein, A.M.B. Pavani, and F. Schiattarella, "Toward on-line, worldwide access to Vatican Library materials", *IBM Journal of Research and Development*, vol. 40, 1996.

[4] K.A. Magerlein G.W. Braudaway and F.C. Mintzer, "Protecting publicly-available images with a visible image watermark", *in Proceedings, SPIE Conference on Optical Security and Counterfeit Deterrence Techniques*, vol. SPIE 2659, pp. 126–132, Feb. 1996.

[5] J. Pickerell and A. Child, *Marketing Photography in the Digital Environment*, DiSC Co., 1994.

[6] M.M. Yeung and F.C. Mintzer, "An invisible watermarking techniques for image verification", *in International Conference on Image Processing*, 1997.

[7] G.W. Braudaway, "Protecting publicly-available images with an invisible image watermark", *in International Conference on Image Processing*, 1997.