Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks and Implications

by Scott Craver, Nasir Memon, Boon-Lock Yeo, and Minerva M. Yeung
presented by Hua Li

November 22, 1999

1 Introduction

Digital information can be duplicated and distributed easily, so some effective copyright protection tools are needed. Watermarking is one of the most important techniques.

- What is a watermark?

  Watermark is a piece of information embedded into the media(image, audio, video, software etc.) to prove the ownership.

- Types and purposes of watermarks

  In general, there are two types of digital watermarks: visible watermark and invisible watermark. They are designed for two purposes: copyright protection and data authentication. The paper focused on invisible watermarking techniques for copyright protection.

- Requirements for the watermarks

  Because watermarking is used to prove the ownerships, it must: be recoverable after intentional or unintentional modification; be robust against image processings and common image compression techniques; be invulnerable to those attempts of forging, removing or invalidation.

- Current invisible watermarking schemes

  The current invisible watermarking schemes can be broadly classified into two categories: *Spatial-domain-based* and *Transform-domain-base*. LSB(embed the least significant bits in an image's pixel data) is in spatial nature which has relatively low-bit capacity. Transform-domain-based techniques can embed a larger number of bits without incurring noticeable visual artifacts. The common transforms are discrete cosine transformation(DCT), wavelets, Fourier transforms.

- What watermarks can achieve and can not achieve?

  Most of watermark schemes are to deal with labeling an image invisibly and making the inserted label robust against malicious attacks. However, can a "robust" watermark solve all problems of ownerships dispute? The paper said that we can not count on a watermark scheme alone.

2 Several scenarios in contesting ownership

The paper described several scenarios about proving the rightful ownership. Two fictional characters are involved in these scenarios. Alice is the true originator of image, while Bob is a forger.

- Scenario 1

  Suppose Alice and Bob use the same digital watermarking technique. If the watermark extracted from the image matches Alice's particular watermark, the image is believed to belong Alice. Same for Bob. However, what if an image contains both Alice and Bob's watermark?

- Scenario 2

  Similar to scenario 1, except they use the entirely different watermarking techniques. Alice can take image and decode the label using her decoding scheme. If the watermark extracted from the image matches Alice's watermark, the image is Alice's. Same for Bob. However, if Alice's decoder indicates that image belongs to her while Bob's decoder indicates the image belongs to him, whose image is it?

- Scenario 3

  To solve the above problem, so-called *original image* can be used. Alice asks Bob for his original image and check if contains her watermark. Similarly, Bob can ask Alice for her original image to check if it contains his watermark. The fact is: if Bob obtained Alice's watermarked image and introduced his own watermark into it, then both of Bob's "original" and watermarked images will contain Alice's watermark. By this method, the problem seems to have been solved.

- Scenario 4

  How about if Alice's original and watermarked images contain Bob's watermark, who owns the image, Alice or Bob? Can this happen? The paper presented that this can happen in the real life. The attack known as *counterfeit watermarking scheme* allows multiple claims of ownership. In such case, rightful ownership can not be resolved by invisible watermark alone.

3 Multiple claims of ownership

By invalidating claims of ownership, Bob can be as much confidence as Alice. The paper explained how this can happen and how Bob can create such a counterfeit "original" image $\hat{I}^{'}$ from the watermarked image $\hat{I}$ without accessing the true original image I.

The key step in the counterfeiting technique is by "subtracting off" a watermark $S^{'}$ in $\hat{I}$ to create $\hat{I}^{'}$. Alice compares her original image to Bob's counterfeit original image and extract the label $T^{'}$ which is very similar to Bob's fake watermark $S^{'}$. The correlation between them is large and implies that they are most likely to be equal. By doing this, Bob can essentially cause a watermark to be present in I and claim that the true original I contains his watermark $S^{'}$ and that it is a modified version of $\hat{I}^{'}$.

4 Noninvertible watermarking of images

We had talked about Bob "remove" his fake watermark from an image rather than insert one, which is known as *invertible watermarking scheme*. Clearly, removing this invertibility is a crucial step in foiling the attack. The authors gave some math definitions and intensive proof of *noninvertible scheme*. Basically, if a mapping from $\hat{I}$ to $S^{'}$ and $\hat{I}^{'}$ does not exist, the watermarking scheme is called noninvertible scheme. However, noninvertible scheme is not a cure-all solution. It is a necessary but not sufficient condition for preventing more powerful attacks, such as *multiple watermarked images* which allows Bob to introduce more counterfeit watermarks into an image under noninvertible schemes.

5 A more powerful attack — attack with multiple watermarked image

In the previous sections, we assume there is one and only one watermarked version of an image I in an ownership dispute. If we can not find out which so-called "original" image is the "true original", how can we decide which watermarks in the image is the real one? Unlike the previous attack, we do not require that insertion of the fake watermark onto the counterfeit original image to produce same watermarked image $\hat{I}$. A counterfeit-attack involving two watermarked versions of an image is called *Twin Watermark Images Counterfeit Original attack* which is even more dangerous: Bob only needs to compute $\hat{I}'$ and S' to yield an image possibly different from but similar enough to $\hat{I}$, in terms of perceptual quality and the features preserved for watermark extraction, that S' can still be expected to lie in I. This leads us to define another type of watermarking scheme: *quasi-invertible.* The definition of quasi-invertibility is similar to the definition of invertibility except a single constraint is relaxed. Designers must therefore take extra care to ensure that theirs are non-quasi-invertible as well. Furthermore, attacks on public watermarking scheme is mentioned briefly in paper.

6 Discussion and conclusions

The current copyright law are quite inadequate to deal with digital data because it can be copied and manipulated very easily. Although, the research on how to make watermark more and more robust against attacks is successful, the watermark has its limitation. A "robust" watermark does not imply it is capable to prove the rightful ownership. We can not count on the watermarking techniques that are invertible or quasi-invertible, while prove that a technique is not quasi-invertible could be very difficult.

It is important to develop more robust watermarking schemes against general attacks while it is also critical to think about the application before we propose or adopt another watermarking scheme. In other words, ask questions like " in what ways can this watermarking technique be attacked?"