International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 12 Issue: 12 | Dec 2025 www.irjet.net p-ISSN: 2395-0072

A Study on the Adequacy of Digital Watermarking in Establishing Ownership Rights

Pavan A C[1], M T Somashekara[2]

[1]*Assistant Professor, Department of Computer Science and Engineering, PES University, Bengaluru, India*
[2]*Associate Professor, Department of Computer Science and Applications, Bangalore University, Bengaluru, India*

Abstract - *Over the years, researchers have developed numerous methods aimed at undermining or defeating digital watermarking systems. These developments indicate that the presence of a robust, non-removable watermark alone is not enough to establish ownership with certainty. This paper aims to identify and formalize the critical conditions a watermarking scheme must meet to reliably determine rightful ownership. Our evaluation shows that several prominent image watermarking approaches fall short of these requirements. However, we demonstrate that a modified version of a known watermarking technique [1] performs effectively under the proposed criteria.*

Key Words: Digital image watermarking, Image Normalization, Robustness, Ownership Rights

1. INTRODUCTION

A wide range of techniques for embedding invisible watermarks into images have been introduced in recent literature. Most of these approaches focus on demonstrating robustness against common image processing operations or intentional attempts to remove the watermark. However, studies [2, 3, 4, 5] have revealed that the ability to embed a robust watermark does not necessarily guarantee that the watermarking scheme can reliably establish ownership.

The attack strategy described in [3] targets watermarking schemes that rely on the original image during the detection phase. To illustrate this attack, consider the following scenario: a user, Good (G), embeds a watermark into his original image (O) and distributes the watermarked version (O′). An adversary, Bad (B), who has access to O′, intends to falsely claim ownership. Assuming an additive watermarking scheme that requires the original image during detection (such as the scheme in [6]), B subtracts his own watermark from O′ to generate a new image F, which he then presents as his original. Because F is derived in this manner, the difference O′ – F will correlate strongly with B's watermark. Similarly, O – F will also exhibit this correlation. As a result, if F is treated as the original during detection, both O and O′ appear to contain B's watermark. This means B can produce the same level of evidence to support his ownership claim as G can, effectively undermining the integrity of the watermarking scheme. Another variant of this attack, targeting a modification of the method in [6], is discussed in [2].

This class of attacks highlights a significant vulnerability that arises when the original image is used in the detection process. It prompts an important question: can we ensure ownership verification without referencing the original image during detection? In this paper, we argue that it is indeed possible. We present a set of necessary and sufficient conditions that a general watermarking scheme must satisfy to establish rightful ownership. Furthermore, we show that a modified version of the scheme proposed in [1] adheres to these conditions and performs effectively in this context.

2. NECESSARY AND SUFFICIENT REQUIRE-

To begin with, we identify the essential conditions that a watermarking scheme must fulfil to prevent subversion. The first and most straightforward requirement is that the scheme should be resistant to watermark removal—referred to hereafter as REQ1. Specifically, any attempt to remove the watermark should cause noticeable degradation in image quality. If this condition is not met, an attacker could simply erase the watermark and embed their own, thereby making a false claim to ownership.

A second, more nuanced condition is that the legitimate owner must possess a version of the image that contains no watermark other than their own. We refer to this as REQ2. Importantly, the owner does not necessarily need to retain the original, unmarked image; a version bearing only their watermark is sufficient, provided that no other watermark can be detected within it. We refer to such a version as the pseudo-original, to distinguish it from the actual, unwatermarked original. If REQ2 is not met, it becomes possible for an attacker to claim that their watermark is also present in the pseudo-original, thereby creating a situation in which both the attacker and the rightful owner appear to have equal proof of ownership.

The above discussion establishes that both REQ1 and REQ2 are individually necessary to assert rightful ownership. We now argue that these two conditions, taken together, are also sufficient. REQ2 ensures that the legitimate owner has access to a pseudo-original that is uncontaminated by other watermarks. Consequently, any adversary attempting to claim ownership must also produce such a pseudo-original, which is infeasible. Assuming the owner only distributes watermarked copies, REQ1 guarantees that any unauthorized attempts to remove or replace the watermark

would significantly distort the image, thereby making the tampering detectable.

Therefore, a watermarking scheme that satisfies both REQ1 and REQ2 can be considered sufficient to reliably establish rightful ownership.

a. Practical Implications

The implications of satisfying **REQ1** are straightforward: the watermark embedding process must be sufficiently robust so that the watermark remains detectable even after the image undergoes degradation. To verify REQ1, one can subject the watermarked image to various forms of distortion or attack and then check whether the watermark can still be reliably detected.

REQ2, on the other hand, dictates that no other party should be able to demonstrate the presence of their watermark in the pseudo-original possessed by the rightful owner. This highlights an important aspect of watermarking robustness—namely, the number of adjustable parameters available in the detection process. A watermarking scheme that relies solely on the watermark seed for detection offers minimal flexibility to an attacker, since the only tunable parameter is the seed itself, and the probability of an attacker's seed coinciding with the owner's is extremely low (assuming a standard, publicly known generator is used).

Conversely, schemes that involve the original image or other reference inputs during detection provide attackers with multiple avenues to manipulate the detection outcome. Each additional input parameter effectively increases the attacker's chances of mounting a successful subversion. Therefore, in the context of REQ2, it is essential to minimize the number of flexible parameters used in the detection process and to rigorously assess any potential vulnerability these parameters might introduce.

As noted in the introduction, excessive flexibility in watermark detection can be exploited by adversaries to design sophisticated attacks that undermine the scheme's ability to verify rightful ownership.

b. Do popular watermarking schemes satisfy REQ1 and REQ2?

Having established that **REQ1** and **REQ2** are both necessary and sufficient conditions for a watermarking scheme to ensure rightful ownership, we now address a key question: Can a watermarking scheme be constructed that satisfies both REQ1 and REQ2 simultaneously?

Most watermarking techniques proposed in literature claim to meet REQ1 to some degree. However, as we will demonstrate in the following section, these claims are not always reliable—REQ1 can indeed be circumvented. Specifically, we will illustrate how a watermark can be removed from an image without causing perceptible degradation, thereby highlighting the need for more rigorous testing methods to validate REQ1.

In contrast, REQ2 proves far more difficult to verify in practical settings. Earlier, in Section 1, we discussed how the scheme in [6] can be undermined through a sophisticated attack that exploits its use of the original image during detection. Based on the analysis in the previous subsection, this clearly shows that the scheme fails to satisfy REQ2. While [2] offers certain enhancements to [6], it does not provide convincing evidence that REQ2 is preserved.

Interestingly, even schemes that do not rely on the original image for detection are not immune. One potential vulnerability arises when an attacker selects a watermark that is inherently present in the image itself. As shown in [3], this form of attack can compromise the scheme in [7], which does not use the original image in the detection process. In such cases, the attacker's watermark appears embedded in the owner's pseudo-original, thereby violating REQ2. This underlines the importance of rigorous evaluation of REQ2, even in schemes that avoid the original image during detection.

The scheme proposed in [1] operates differently: watermark embedding is achieved by adding a pseudo-random sequence to the top 25,000 DCT coefficients (ordered in a zig-zag scan) while excluding the first 16,000 coefficients. Detection involves correlating these 25,000 coefficients in the test image with the original watermark sequence and checking for a peak in the detector's output. This additive watermarking scheme does not require the original image during detection, and—assuming the watermark is truly pseudo-random—can be considered compliant with REQ2. The randomness can be ensured by using only sequences generated by a standardized pseudo-random generator, minimizing the likelihood that such a sequence exists naturally in any image.

Therefore, if the scheme in [1] can also be shown to satisfy REQ1, it becomes a viable candidate for a robust watermarking scheme. In the next section, we will demonstrate how a modified version of this approach can fulfil both requirements effectively.

3. TOWARDS A ROBUST WATERMARKING SCHEME

In the previous section, we established that the watermarking scheme proposed in [1] satisfies **REQ2**, i.e., it prevents unauthorized parties from claiming ownership of the image. To complete the analysis, it is now necessary to verify **REQ1**—ensuring that the watermark remains detectable even after typical image degradations, and cannot be removed without causing perceptible damage to the image.

The authors of [1] demonstrate that their scheme withstands JPEG compression (up to 10% quality) and several common

image-processing operations, indicating a level of robustness. The scheme operates by embedding an additive pseudo-random sequence into selected Discrete Cosine Transform (DCT) coefficients. A natural approach to attempt watermark removal, therefore, is to treat the watermark as noise and apply a de-noising algorithm—specifically, one designed for images corrupted by Additive White Gaussian Noise (AWGN).

An initial attempt to erase the watermark used an adaptive, localized Wiener filter. While this filter did remove the watermark, the de-noised image exhibited noticeable blurring, compromising perceptual quality (see Figures 3(a) and 3(b)). A more effective attack was achieved using a custom AWGN de-noising algorithm previously proposed in [8]. This method leverages hard-thresholding across multiple signal compaction domains, integrated within an optimization framework. Though details are beyond the scope of this paper, the results (Figure 3(c)) show that the watermark can be erased with minimal perceptual distortion—effectively defeating REQ1 for the original scheme in [1].

This leads to the question: Can the scheme in [1] be modified to improve resistance to such de-noising attacks? The vulnerability arises from embedding the watermark into the last 25,000 DCT coefficients (excluding the first 16,000 in zig-zag order), which typically correspond to low-energy (high-frequency) components. Additionally, since the watermark is scaled based on the magnitude of the coefficients, the actual signal added may be too weak to survive de-noising.

A simple but effective modification was found by reducing the number of skipped coefficients—from 16,000 to 1,000—before embedding the watermark into the next 25,000. This change ensures that the watermark is added to higher-energy coefficients, which are more robust to de-noising. To maintain perceptual invisibility, the watermark strength factor (α) was reduced from 0.2 to 0.1.

Despite this improvement, there is still no guarantee that the selected 25,000 coefficients will consistently carry sufficient image energy across all images. The scheme's performance thus remains somewhat image-dependent.

(a)

(b)

(c)

To overcome this limitation, a Discrete Wavelet Transform (DWT)-based watermarking scheme was recently proposed in [9]. Unlike the DCT-based method, the DWT approach ensures the watermark is embedded in significant coefficients across multiple resolutions, making it inherently more robust to de-noising and other distortions. Furthermore, the DWT's time-frequency localization enables **implicit visual masking**, allowing the watermark to be embedded where the human visual system is less sensitive. In contrast, [1] relies on **explicit visual masking**, which cannot be factored into the detection process. As a result, the scheme in [9] not only enhances robustness but also yields a stronger detector response compared to [1].

4. CONCLUSIONS

In this paper, we have proposed a set of necessary and sufficient conditions that a digital image watermarking scheme must satisfy to establish rightful ownership. We have also demonstrated how these conditions can be practically applied to evaluate the robustness and reliability of existing watermarking techniques. Specifically, we have shown that a modified version of the algorithm introduced in [1] can indeed satisfy both conditions, thereby serving as a viable candidate for ownership verification. Additionally, our investigation reveals that popular image de-noising algorithms are among the most effective tools for attacking additive spread-spectrum watermarks, potentially undermining their robustness.

It is important to note, however, that the conditions outlined in this work are specific to digital images and may not directly extend to video watermarking. The increased temporal redundancy and complexity in video content

provide significantly more opportunities for sophisticated attacks, necessitating a distinct and more comprehensive set of requirements. These challenges will be addressed in future work.

5. REFERENCES

[1] Piva et al., 1997: "DCT-based watermark recovering without restoring to the uncorrupted original image," Int. Conf. on Image Processing, vol. 111, pp. 520-523.

[2] Craver et al., 1997: "On the invertibility of invisible watermarking techniques," Int. Conf. on Image Processing, vol. 111, pp. 540-543.

[3] Craver et al., 1996: "Can invisible watermarks resolve rightful ownerships?," IBM Cyber Journal, http://www.research.ibm.com:8080, July.

[4] Craver et al., 1997: "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications," IEEE Journal on Selected Areas of Communications, to be published.

[5] Zeng and Liu, 1997: "On resolving rightful ownerships of digital images by invisible watermarks," Int. Conf. on Image Processing, vol. 111, pp. 552-555.

[6] Cox et al., 1997: "Secure spread spectrum watermarking for multimedia," IEEE Trans. on Image Processing, vol. 6, pp. 1673-1687, Dec

[7] Pitas, 1996: "A method for signature casting of digital images," Int. Conf. on Image Processing, pp. 215-218.

[8] Ishwar et al., 1998: "Image denoising using multiple compaction domains," Int. Conf. on Acoustics, Speech and Signal Processing (invited paper).

[9] Dugad et al., 1998: "A wavelet based scheme for watermarking image," Int. Conf. on Image Processing, Oct.

[10] Yen and Huang, 2016: "Frequency domain digital watermark recognition using image code sequences with a back-propagation neural network," Multimedia. Tools Appl., 75, 9745-9755.

[11] Thahab and Shahadi, 2018: "Robust Image Watermarking Against Crop and Rotation Attacks for Public Communication Networks," SCEE 2018, IEEE, 105-110.

[12] Mishra et al., 2018: "Bi-directional extreme learning machine for semi-blind watermarking of compressed images," J. Inf. Secur. Appl., 38, 71-84.

[13] Fatahbeygi and Akhlaghian Tab, 2019: "A highly robust and secure image watermarking based on classification and visual cryptography," J. Inf. Secur. Appl., 45, 71-78.

[14] Moosazadeh and Ekbatanifard, 2017: "An improved robust image watermarking method using DCT and YCoCg-R color space," Optik (Stuttg)., 140, 975-988.

[15] Chutani and Goyal, 2018: "Improved universal quantitative steganalysis in spatial domain using ELM ensemble," Multimed. Tools Appl., 77, 7447-7468.

[16] Ismael et al., 2014: "Lossless audio steganography based on lifting wavelet transform and dynamic stego key," Indian J. Sci. Technol., 7, 323-334.