# Digital Image Watermarking: Issues in Resolving Rightful Ownership

Krishna Ratakonda, Rakesh Dugad and Narendra Ahuja *

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801

**Abstract**

In recent literature many strategies for attacking and subverting a watermark have been presented. Such attacks suggest that the ability to embed non-erasable watermarks does not necessarily imply that the watermarking scheme can be used to establish ownership. The main aim of this paper is to formulate *necessary and sufficient* requirements for a watermarking scheme to be able to resolve rightful ownership. It is shown that some of the popular schemes proposed in literature for watermarking images do not satisfy these requirements. Finally, we show that a modification of a watermarking scheme in literature [1] performs satisfactorily.

## 1 Background

Many schemes for casting invisible watermarks on images have been proposed in recent literature. The proponents of such schemes have mainly concentrated on showing that their scheme is robust to common image processing operations or malicious attempts to remove the watermark. However, recent literature has shown that the ability to embed robust watermarks does not necessarily imply that the watermarking scheme can be used to establish ownership [2, 3, 4, 5].

The mode of attack proposed in [3] is applicable to schemes which use the original image in the watermark detection process. To further illustrate this form of attack, let us assume the following scenario: Good(G) has watermarked his original(O) and made the watermarked copy(O') available to Bad(B) who wants to make false claims to ownership. Assuming an additive watermarking scheme which uses the original image (for example the scheme in [6][1]) B proceeds thus: he *subtracts* his watermark from O' (which is available to him) to obtain a new image (F). B claims F to be his original. The crucial point to note is that B *subtracts* his watermark from O' to obtain F, which implies that O'-F (the retrieved watermark) will have a strong correlation with B's watermark; this is also the case with O-F. From the way that this fake original F is constructed, it is evident that if F is used as the original in the watermark detection process, O (the original with G) and O' will be found to have the watermark of B. Therefore it is found that F, O' have the watermark of G while O, O' have the watermark of B. Hence, B has amassed equal evidence to prove his ownership as G has to prove his, thus subverting the watermarking scheme. [2] shows another variation of this attack, which can be used against a variation of the method in[6].

The above illustration of an attack on watermarking schemes relies on the usage of the original in the watermark detection process. This leads us to the following question: if the original image is not used in the detection process, can

* This research was supported by the Advanced Research Projects Agency under grant N00014-93-1-1167 administered by the Office of Naval Research and the NSF grant IRI 93-19038.

[1]The scheme in[6] functions thus: watermark casting is performed by adding the watermark, a pseudo random noise sequence, to the DCT of the image; watermark detection is performed by (i) retrieving the watermark by subtracting out the original from the test image and (ii) correlating the resultant with the actual watermark.



0-8186-8821-1/98 $10.00 

some conditions be imposed on the watermarking scheme to ensure that it can perform the duty of ensuring rightful ownership? We will show that this is indeed the case and further show that a modification of the scheme in [1] works reasonably well towards this end. We will also prove that there exist necessary and sufficient conditions, which are applicable to a general watermarking scheme, which ensure the ability of such schemes to prove rightful ownership.

## 2 Necessary and Sufficient Requirements

To start with, let us identify conditions which a watermarking scheme should *necessarily* satisfy inorder to avoid subversion. An obvious requirement is that the scheme should be resistant to attempts to erase the watermark from the image (we will call this REQ1 for future reference). In other words, successful attempts at removing the watermark from the image should result in significant perceptual degradation. If this requirement is not satisfied, a simple malicious attack can be made to claim ownership: erase the original watermark from the image and replace it with the attacker's watermark.

Another, somewhat more subtle requirement is that the correct owner should have in his possession a copy of the original image which should not have any other watermark except possibly his own (we will call this REQ2 for future reference). Note that the owner need not necessarily have the original. Even a copy of the image with his own watermark embedded in it would suffice, as long as no other watermark can be shown to be embedded within this image. We will call this uncorrupted copy with the owner to be the "pseudo-original" to distinguish it from the unmarked original. If this requirement is not satisfied, it follows that the watermark of a malicious attacker can be shown to be embedded within the pseudo-original. In such a situation the following false claim to ownership can be made by the attacker: both the owner and the attacker have the same amount of evidence to show that the image is his own, since each can only show that the watermarked image in the other's possession has his watermark; hence the actual owner does not have any special claim to ownership.

In the above discussion, it was shown that both REQ1 and REQ2 are *independently necessary* to claim rightful ownership. It will be now shown that REQ1 and REQ2 form a *sufficient set of requirements* for a watermarking scheme to be suitable to establish rightful ownership. REQ2 implies that the owner has within his possession a pseudo-original (i.e., an image which has no other person's watermark embedded within it). Therefore, under this watermarking scheme, a malicious attacker needs to show that he has a pseudo-original in his possession in order to prove his ownership (the correct owner can do this). Assuming that the owner only distributes the watermarked copy of his original image, REQ1 implies that any other person possessing a copy of this image, can only have a copy with the owner's watermark in it. Hence, the watermarking scheme is suitable to establish rightful ownership.

### 2.1 Practical Implications

The implications of satisfying REQ1 are clear: the watermark casting process should be robust enough that the watermark can be detected by the detection process even after degradations of the image. Hence in order to verify REQ1, we can subject the watermarked image to degradations which might remove the watermark and verify that the watermark is still intact.

REQ2 implies that, given the watermark detection process, it should not be possible for anyone else to establish the presence of his/her watermark in the pseudo-original with the owner. Therefore the robustness of a given watermarking (casting and detection) scheme depends also on how many *flexible parameters* the watermark detection process employs. For example, a scheme which uses only the watermark (i.e. the seed) during detection has only one flexible parameter – the seed and hence all that the attacker can specify is the seed (a standard generator for the watermark is already established) which has almost zero probability of matching with the seed of the owner. On the other hand a scheme using the original or any other image during detection

process gives the attacker that many more flexible parameters in setting up his attack. It is thus important, in the context of REQ2, to use as few flexible parameters as possible in the detection process and to carefully validate the fact that these might not be used to subvert the watermarking scheme. As pointed out in the introduction too much flexibility can be used by the attacker to design methods of attack which might render the watermarking process incapable of establishing rightful ownership.

### 2.2 Do popular watermarking schemes satisfy REQ1 and REQ2?

Since we have established that REQ1 and REQ2, as stated above, form *necessary and sufficient* requirements to establish rightful ownership, it is important to answer the question: Is it possible to construct a watermarking scheme which satisfies both REQ1 and REQ2 simultaneously?

Most schemes proposed in literature have been argued to satisfy REQ1 to some extent. In the following section we will see that REQ1 can be defeated and more stringent tests than conducted so far need to be used to test for REQ1. We will show a way to erase a watermark from an image without causing significant perceptual damage.

REQ2 is much harder to verify in practice. In section 1, we found that the method in [6] can be subverted by a sophisticated attack which was based on the fact that [6] uses the original image in the detection process. In view of the discussion in the previous sub-section, we can readily see that the attack proves that the scheme in [6] does not satisfy REQ2. Although [2] proposes enhancements to the scheme in [6], we find that no satisfactory arguments were made to prove that REQ2 cannot be subverted.

On the other hand, if the original image is not used in the detection process, one of the reasons that REQ2 can fail is that the attacker's watermark is an inherent part of the image (i.e., the attacker picks a watermark which is naturally present within the image). [3] shows that the scheme in [7], which does not use the original image, can still be subverted by the attacker, who picks a watermark which is an inherent part of the image and is therefore present in the owner's pseudo-original. This shows that we need to be careful in checking whether a watermarking scheme satisfies REQ2, even if the scheme does not use the original image.

The watermarking scheme reported in [1] functions thus: watermark casting is performed by adding a pseudo-random sequence to the top 25000 DCT coefficients in zig-zag scan order after leaving out the first 16000 coefficients[2]; watermark detection is performed by correlating these 25000 coefficients in the test image with the original copy of the watermark and looking for a peak in the detector response. It is easily seen that this additive watermarking scheme, which does not use the original image during detection, does satisfy REQ2, provided that the watermarks are truly pseudo-random sequences and therefore cannot be an inherent part of an image (also see section 2.2). The pseudo-random nature of the sequences can be easily assured by limiting the watermarks to those obtained from a standard generator. Hence, if the scheme of [1] can be proven to satisfy REQ1, it could work as a watermarking scheme. We will show, in the next section, that a modification of the method in [1] can indeed be used towards this purpose.

## 3 Towards a Robust Watermarking Scheme

In the previous section it was shown that the watermarking scheme proposed in [1] satisfies REQ2. From the discussion in the previous section, it is only necessary to show that this scheme casts a watermark that is not erasable without generating perceptually discernible damage to the image (REQ1). [1] shows that the scheme can tolerate 10 % quality JPEG compression and various other common image manipulations without erasing the watermark.

We note that the scheme embeds an additive pseudo-random sequence into the image; thus the most suitable way to remove the watermark might

[2] watermark casting also involves visual masking of the watermarked image to ensure perceptual invisibility of the watermark

be to use a de-noising algorithm for restoring images corrupted by AWGN sequences! Our first choice was to use an adaptive, localized Wiener filter to attack the watermark. Although the watermark was removed from the de-noised image, the image itself was blurred. Figure 3(a) shows the watermarked image and figure 3(b) shows the de-noised image. Recently, we had proposed a AWGN de-noising algorithm for images [8], which de-noises the image while producing perceptually better images than adaptive Wiener filtering. The algorithm combines estimates from hard thresholding in multiple signal compaction domains within an optimization framework. Further details are beyond the scope of this paper. Figure 3(c) shows the de-noised image using this method. It is seen that this de-noising scheme erases the watermark and yet maintains enough perceptual quality to subvert REQ1.

Is it possible to modify the scheme in [1] to make it more resistant to de-noising schemes? The crucial factor which undermines the watermarking scheme is the fact that the watermark is added to 25000 coefficients in the DCT domain, *after leaving out the first 16000 coefficients in zig-zag scan order.* Thus, we are adding the watermark to (typically) very small coefficients in the DCT domain. Since the scheme in [1] scales the added watermark according to the value of the coefficient, the amount of watermark added may be small. We found that adding the watermark to 25000 coefficients, after leaving out the first 1000 coefficients in zig-zag scan order vastly improves the resistance of the watermark to image manipulations. In order to ensure perceptual invisibility we reduced the factor $\alpha$, which determines the amount of watermark added, from 0.2 to 0.1.

Since we add the watermark to 25000 coefficients in zig-zag scan order, there is no guarantee that in every image these coefficients need necessarily have a large amount of image energy. Therefore, it may be still argued that the above watermarking scheme might fail in the case of some images, dependent on the distribution of their energy in the DCT domain. We recently proposed a DWT based scheme [9] which guar-

(a) (b) (c)

Figure 1: Results from Erasing the Water Mark: (a) Watermarked image (b) After adaptive, localized Wiener filtering and (c) After processing with the de-noising scheme in [8].

antees that the watermark is always added to the significant coefficients, thus ensuring resistance to de-noising schemes (apart from other image manipulations). This scheme has the further advantage that visual masking is implicit (taking advantage of the time-frequency localization properties of the DWT) as opposed to [1], which needs to perform explicit visual masking. Since explicit visual masking cannot be taken into account during the watermark detection process, the method in [9] has better detector response over the method in [1].

## 4 Conclusions

In this paper, we have suggested necessary and sufficient conditions which ensure the ability of a watermarking scheme to establish rightful ownership and how these may be employed in practical scenarios. Further more, we have shown that a modification of the algorithm proposed in [1] can be used to establish rightful ownership. It is also shown that popular de-noising schemes are possibly the most effective tools to remove an additive spread-spectrum watermark.

We would like to point out that the necessary and sufficient conditions that we have established for digital image watermarking schemes are not applicable in general to video sequences. Intuitively this can be attributed to the fact that video offers many more avenues for attack. These issues will be tackled in a forth coming paper.

## References


[1] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without restoring to the uncorrupted original image," in *International Conference on Image Processing*, vol. III, pp. 520–523, 1997.

[2] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "On the invertibility of invisible watermarking techniques," in *International Conference on Image Processing*, vol. III, pp. 540–543, 1997.

[3] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "Can invisible watermarks resolve rightful ownerships?," in *IBM Cyber Journal*, http://www.research.ibm.com:8080, July 1996.

[4] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications," *To be published in IEEE Journal on Selected Areas of Communications.*

[5] W. Zeng and B. Liu, "On resolving rightful ownerships of digital images by invisible watermarks," in *International Conference on Image Processing*, vol. III, pp. 552–555, 1997.

[6] I. J. Cox, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[7] I. Pitas, "A method for signature casting of digital images," in *International Conference on Image Processing*, pp. 215–218, 1996.

[8] P. Ishwar, K. Ratakonda, P. Moulin, and N. Ahuja, "Image denoising using multiple compaction domains," in *International Conference on Acoustics, Speech and Signal Processing (invited paper)*, 1998.

[9] R. Dugad, K. Ratakonda, and N. Ahuja, "A wavelet based scheme for watermarking image," in *International Conference on Image Processing*, Oct. 1998.