Robust Protocols for Proving Ownership of Images

Mahalingam Ramkumar and Ali N. Akansu
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
New Jersey Center for Multimedia Research
University Heights, Newark, NJ 07102
ali@megahertz.njit.edu

Abstract

We explore the problem of proving ownership or origin of multimedia content like image / video or audio signals through watermarking. The need for watermarking *arises out of the unsuitability of present copyright laws for claiming ownership of digital content. Watermarking schemes however, are threatened by* counterfeit attacks*, which primarily use the* freedom *available in the choice of* signature *or choice of the* watermarking method*. A restrictive protocol for watermarking could go a long way in rendering counterfeit attacks extremely difficult. We suggest a comprehensive protocol that makes it possible for the true owner to claim ownership unambiguously, while making it extremely difficult for a pirate to do so.*

1 Introduction

Establishing ownership of books or blueprints, have traditionally been done by obtaining copyright on that content, perhaps from the copyright office. However, the nature of digital content makes traditional copyright mechanisms unsuitable for establishing ownership. Figure 1 depicts two typical scenarios, where existing copyright mechanisms may be unsuitable for securing copyright of say, digital images. In scenario I, A and B represent two distinct but identical photographs created by Alice and Bob respectively (both photographs may have been shot from the same place at different instances of time). Alice is responsible for circulating copies of her art as A_1. Meanwhile, Bob circulates his creation as B_2. Both Alice and Bob register their contents A and B with the Copyright Office [1]. If both A and B (and hence A_1s and B_2s) look identical, Bob can claim that A and all A_1s are violations of his copyright while Alice can claim that B and B_2s are violations of her copyright. Obviously, this is not a desirable situation. In a second scenario, the photograph was created by Alice who is not *interested* in obtaining a copyright. Bob may have received a copy of A (which Alice may have made freely available on her website), for which he promptly obtains a copyright, and then circulates it as A_2. While it may still be acceptable for Bob to claim ownership of all A_2s (circulated by Bob) it is definitely not ethical to let the copyright law enable Bob to claim ownership of the original A created by Alice. The key issue here (which cannot be determined by traditional copyright mechanisms) is to determine copies which *originate* from a *particular source*. *Watermarking the source* can effectively address this problem.

Digital watermarking [1] is a means of protecting multimedia content from intellectual piracy. It is achieved by imperceptibly modifying the original content to insert a "signature". The signature is extracted when necessary to show proof of ownership. In this paper, for the purpose of illustration, we assume that the original content is a digital image. However, the proposed protocol is equally applicable for video and audio signals as well.

Let I be the original (cover) image. A watermark embedding function E inserts a watermark S in the image I to generate the watermarked image $\hat{I} = E(I, S)$. The existence of the watermark S in an image $\tilde{I}$ is checked by a detector D. Watermark detectors can be broadly classified into two categories. *Non-oblivious* detectors need the original image I to check for the presence of the signature S in $\tilde{I}$. On the other hand, *oblivious* detectors [2, 3] *do not* require the original image. We shall term the output of the detector,

$$s_d = \begin{cases} D(\tilde{I}, S, I) & \text{non-oblivious detector} \\ D(\tilde{I}, S) & \text{oblivious detector} \end{cases} \tag{1}$$

as the *detection statistic*. The detection statistic is an indication of the *degree of certainty* with which the signature S is detected in the image $\tilde{I}$.

[1] to register a work of *visual art* a completed application form, a non-refundable filing fee of \$30 and a non-returnable deposit of the material to be registered are to be mailed to the Copyright Office. See http://www.loc.gov/copyright/reg.html for more details

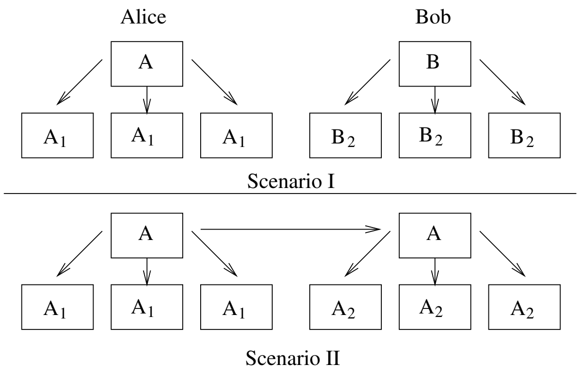

Figure 1. Top - Scenario I: A and B are two similar photographs created by Alice and Bob. Scenario II : Alice (creator of A) does not want to obtain copyright.

1.1 Detection Statistic

Typically, the signature S takes the form of a Gaussian or binary pseudo random sequence $\mathbf{s}$ (say of length N) generated from a *key* $\mathcal{K}$. The watermark embedding and detection operations can therefore be written as

$$\hat{I} = \mathcal{E}(I, \mathbf{s}) \quad \tilde{\mathbf{s}} = \mathcal{D}(\tilde{I}, < I >) \quad s_d = \frac{\mathbf{s}^T \tilde{\mathbf{s}}}{|\mathbf{s}||\tilde{\mathbf{s}}|} \tag{2}$$

The detection statistic ($-1 \leq s_d \leq 1$) is a measure of (normalized) *inner product*, or normalized correlation of the embedded and the detected signature sequences. The normalized inner product of randomly generated signature sequences will also be random. More specifically, for large N, the distribution of the inner product will be Gaussian. Let x_i and y_i, $i = 1, \ldots, N$ be i.i.d. of variances $\sigma_x^2 = \sigma_y^2 = \frac{1}{N}$, and let $q_i = x_i y_i$. The inner-product is calculated as $p = \sum_{i=1}^N q_i = \sum_{i=1}^N x_i y_i$. Since x_is and y_is are independent, the variance of q_i is $\sigma_q^2 = \sigma_x^2 \times \sigma_y^2 = \frac{1}{N} \times \frac{1}{N} = \frac{1}{N^2}$. Therefore, for large N, from central limit theorem (CLT) [7] the $p \sim \mathcal{N}[0, \frac{1}{N}]$.

If the creator (or pirate) has *absolutely no freedom* in choosing the signature, and if the detection statistic s_d obtained is say 6 times the standard deviation (if $s_d = 6\frac{1}{\sqrt{N}}$), then we could say that the signature is detected with a probability of error (or probability of false alarm) of less than $Q(6) \approx 1 \times 10^{-9}$, where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{\frac{t^2}{2}} \mathrm{dt}$. In other words, on an average, only 1 out of 1×10^9 signatures chosen randomly can yield such a high correlation. Note that N is also the *degree of freedom* of the signature. Even if the detection statistic is unity, the signature is still detected with a finite false alarm probability, given by $Q(\sqrt{N})$. *It is therefore very important to have a sufficiently large N to be able to yield "acceptable" levels of P_e.*

2 Counterfeit Attacks on Watermarks

One of the primary problems to be addressed by watermarking methods is their ability to make a counter-claim *practically impossible*. A counter-claim arises from situations where a pirate can use the *inadequacies of watermarking protocols* [4, 5] to "demonstrate" the presence of a his / her "watermark" (fake watermark or signature) in the actual original content.

Time stamping [6] has been proposed as an *enhancement* to the security provided by watermarking to overcome the problems associated with counter-claims in watermarks. In addition to watermarking, the creator can obtain a time stamp from a time stamping service (TSS). If the time stamp is obtained before the content is released to the public, (before the *pirate* can obtain a time stamp on the content) nobody else can claim legitimate ownership of the content. However, time-stamping has the disadvantage of requiring *ongoing involvement* of a third party. Moreover, there are some situations for which it does not provide acceptable solutions. Time-stamping does not protect people who do not want to obtain time stamp and/or watermark their content, like Alice in Scenario II. If Bob is able to show a counterfeit signature in image A created by Alice, and if Alice has not obtained a time-stamp, then Bob will be able to claim ownership of content created by Alice. Clearly, time stamping does not help in situations like this. Obviously, time stamping is also not a good solution for time sensitive applications (like live broadcasts). However, with a suitable protocol, which would lay some (very reasonable) restrictions on watermarking algorithms, the above mentioned problems can be effectively addressed.

2.1 Freedom in Choice

Let Alice be the creator of the original image I. She embeds her signature S_A in I to obtain the watermarked image $\hat{I}_A = E_A(I, S_A)$. The presence of her signature S_A in $\hat{I}_A$ or any image $\tilde{I}_A$ *derived* from $\hat{I}_A$ (or $\tilde{I}_A = \hat{I}_A + P$) can be demonstrated with a reasonably good degree of certainty, by obtaining a sufficiently high detection statistic

$$s_{d_A} = D_A(\tilde{I}_A, S_A, < I >). \tag{3}$$

In the above equation $< I >$ denotes that the original image I may or may not be used by the detector. The job of Bob, an aspiring pirate, is to demonstrate the presence of his (arbitrary) signature S_B in Alice's original image I. In other words, Bob has to obtain a "large"

$$s_{d_B} = D_B(I, S_B, < I_f >), \tag{4}$$

where I_f may be Bob's *fake original* image. Note that Bob is at liberty to *choose his own watermarking scheme* (E_B, D_B). If Bob has freedom in choosing his signature

S_B, he could fix some (E_B, D_B), and "construct" a signature S_B that yields a high detection statistic s_{d_B}. Note that even though Bob does not possess a copy of I (which is never released to the public by Alice), he does have $\hat{I}_A$, which is "very close" to I. If Bob does not have freedom in choosing his signature (say, if the signature is assigned to him by a Watermarking Authority), he could still try to *tweak* the watermarking scheme (E_B, D_B) to obtain a high detection statistic. Therefore, it is clear that a good watermarking protocol should lay some restrictions *on the choice of signature* and *choice of the embedding / detector functions* (or the watermarking algorithm).

2.2 Fake Originals

Even if the watermarking scheme *and* the choice of signature is fixed, it may still be possible for a pirate to engineer a counterfeit attack, *if the detection scheme is non-oblivious*. This would permit the pirate to create a *fake original* (cover) image, for which there are no restrictions! This problem can be solved if the detection method is *strictly* oblivious.

For strictly oblivious watermarking, the watermark detector may not even know the size of the original. The received image may have been resized, rotated, cropped, undergone some histogram modification, and probably some geometric distortion such as those introduced by StirMark [8]. If by some way, a watermarking scheme is able to detect the watermark in a strictly oblivious fashion (in the face of all possible attacks), the watermarking scheme would have drastically reduced degrees of freedom for the choice of the signature. Reduced degrees of freedom implies lower separation between possible watermarks, and also, in general, lower robustness to attacks. As mentioned in Section 1.1 the maximum false alarm probability obtainable is $Q(\sqrt{N})$. Therefore, such schemes may not be able to *unambiguously establish ownership* in a court of law. In other words, a *strictly* oblivious watermarking method, capable of resolving ownership may never be practical. Therefore, the content creator may be forced to use the original image to *undo* the distortions introduced by the pirate. Under this condition, the pirate may gain some freedom in *choosing* an algorithm for undoing the geometric distortions. Therefore, a good watermarking protocol should also fix (or *regulate*) the algorithms used for undoing the distortions introduced. However, the pirate still has *freedom in choosing the fake original which will be used by the fixed algorithm for undoing geometric distortions*. In other words, the pirate (Bob) can *engineer* a (fake) original which when used in conjunction with the fixed algorithm, can "undo the distortion" in Alice's original image I (Bob would claim that the fake original he has created to be the original, and Alice's image I, to be an image derived from his original) to show the presence of his watermark. Again, to engineer the attack, he has the image $\hat{I}$ which is "very close" to I.

2.3 Multiple Watermarks

Another possible counterfeit attack on watermarks is to create ambiguity in ownership due to the presence of multiple watermarks. Let I be the actual original created by Alice, and $\hat{I}_A$ the image containing Alice's watermark, which is available to Bob. Bob can create a new image $\hat{I}_{AB}$, by embedding his signature in $\hat{I}_A$. It is very much possible that Bob obtains a much higher detection statistic (of his signature) in $\hat{I}_{AB}$ than Alice. In this case, what can prevent Bob from claiming ownership of $\hat{I}_{AB}$? If Alice's watermarking scheme is robust, it might not be possible for Bob to show any copy of an image which *originated* from I that *does not contain* Alice's signature. On the other hand, Alice's actual original I, and $\hat{I}_A$, will not contain Bob's signature. However, if Alice does not have a copy of the original (or $\hat{I}_A$) in her possession, this problem may be difficult to resolve.

3 Watermarking Algorithms

In this section we present a generalization of "conventional" watermarking algorithms. A few "unconventional" watermarking schemes that embed the watermark by introducing geometric distortions [9] or by modifying the histogram have been proposed in literature. However, the main limitation of these algorithms are lack of robustness to commonly occurring attacks like lossy compression and / or drastically reduced degrees of freedom (N). The "conventional" algorithms, rely on the assumption that if the image is altered significantly in the mean-square error (MSE) sense, then the quality of the resultant image would be so poor that it would not warrant a ownership claim. Even though this assumption is not true, conventional algorithms can be *used effectively along with regulated algorithms for undoing geometric distortions / histogram modification etc.*, or algorithms for *registration* of images.

Typically, the watermark is inserted in some transform domain. Let $\mathbf{C} = \mathcal{T}(I)$, where $\mathcal{T}$ denotes some transform and $\mathbf{C}$ are the transform coefficients of I. Generally, only a subset of $\mathbf{c} \in \Re^N$ of $\mathbf{C}$ may be modified to embed the watermark, $\mathbf{s} \in \Re^N$, which is typically a pseudo-random binary or Gaussian sequence generated from a key $\mathcal{K}$. This general model does not preclude the possibility of $\mathcal{T}$ being an identity transform, and the subset $\mathbf{c} = \mathbf{C}$. For enhanced robustness to intentional attacks, the transform (or its basis vectors) may also be generated from secret keys [10]. Let $\mathbf{C} = \mathbf{c} \cup \bar{\mathbf{c}}$, where $\mathbf{c} \cap \bar{\mathbf{c}} = \emptyset$. The coefficients $\bar{\mathbf{c}}$ are unaffected by the watermarking process. The overall embedding operation may be expressed as

$$\mathbf{C} = \mathcal{T}(I) \quad \hat{\mathbf{c}} = \mathcal{E}(\mathbf{c}, \mathbf{s}) \quad \hat{\mathbf{C}} = \hat{\mathbf{c}} \cup \bar{\mathbf{c}} \quad \hat{I} = \mathcal{T}^{-1}(\hat{\mathbf{C}}) \qquad (5)$$

Let $\tilde{I} = \hat{I} + P$ be the image in which the presence of the watermark is tested. The detection operation is now

$$\tilde{\mathbf{C}} = \mathcal{T}(\tilde{I}) \quad \tilde{\mathbf{s}} = \mathcal{D}(\tilde{\mathbf{c}}) \quad s_d = \frac{\mathbf{s}^T \tilde{\mathbf{s}}}{|\mathbf{s}||\tilde{\mathbf{s}}|} \qquad (6)$$

Table 1. Comparison of Types I, II and III

Type I	$\hat{\mathbf{c}} - \mathbf{c} = \mathbf{s}$	$\mathcal{D}(\hat{\mathbf{c}}) \neq \mathbf{s}$
Type II	$\hat{\mathbf{c}} - \mathbf{c} \neq \mathbf{s}$	$\mathcal{D}(\hat{\mathbf{c}}) = \mathbf{s}$
Type III	$\hat{\mathbf{c}} - \mathbf{c} \neq \mathbf{s}$	$\mathcal{D}(\hat{\mathbf{c}}) \neq \mathbf{s}$

Let $\mathbf{n} = \tilde{\mathbf{c}} - \mathbf{c}$ be the effect of P in the transform domain on the coefficients $\mathbf{c}$ utilized for data hiding. The watermarking algorithms that fit into the general model of Eq. (5 - 6) can further be classified into 3 types, depending on the embedding and detection operators $(\mathcal{E}, \mathcal{D})$ - Types I, II and III. For Type I methods, $(\mathcal{E}, \mathcal{D})$, take the form of linear addition. Mathematically, $\hat{\mathbf{c}} = \mathbf{c} + \mathbf{s}$. Type I methods can further be classified as non oblivious methods, where $\mathcal{D}(\tilde{\mathbf{c}}) \equiv \tilde{\mathbf{c}} - \mathbf{c}$ (for example, the method in [1]), and oblivious methods (for example, [2]), where $\mathcal{D}(\tilde{\mathbf{c}}) \equiv \tilde{\mathbf{c}}$ (no operation). Type II methods have a non-linear element in the embedding and detection operators, which usually takes the form of a quantizer with step size Δ [11, 12], or a periodic function with period Δ. In Ref [13] we presented an unification of Types I and II (called Type III). Type III methods are characterized by a period Δ and a *threshold* β. Optimal choice of Δ and β for a given permitted distortion of the host signal and additive noise variance in the channel was also discussed. For watermarking applications, optimal choice of Δ and β may be seen as choices which maximize the detection statistic s_d in Eq. (6). For Type II methods, $\mathcal{D}(\hat{\mathbf{c}}) = \mathbf{s}$. For Type I oblivious methods, however, $\mathcal{D}(\hat{\mathbf{c}}) = \hat{\mathbf{c}} \neq \mathbf{s}$. For Type III methods, in general, $\mathcal{D}(\hat{\mathbf{c}}) \neq \mathbf{s}$. The differences between the three types are tabulated in Table 1. Furthermore, when $\beta = \Delta$ Type III $\rightarrow$ Type II, and as $\Delta \rightarrow \infty$, Type III $\rightarrow$ Type I. For high SNR (ratio of permitted distortion $|\hat{\mathbf{c}} - \mathbf{c}|^2$ to the noise in the channel, $|\mathbf{n}|^2$), the optimal Type III is close to Type II ($\beta \approx \Delta$). For low SNRs, the optimal Type III is close to Type I ($\Delta >> \beta$). Typically, as watermarking schemes are expected to be very robust, $|\mathbf{n}|^2 >> |\hat{\mathbf{c}} - \mathbf{c}|^2$ (permitted distortion for embedding the watermark). Therefore, the optimal Type III for watermarking applications should utilize large values of Δ.

4 Restrictions on Choice of Signature

The restrictions for choice of signature, proposed in watermarking literature, can be classified into 3 types -

1. issued by a Watermarking Authority (Scheme I).
2. derived from a meaningful string [2] (Scheme II).
3. derived from the cover image [4] (Scheme III).

Scheme I has a major disadvantage of requiring a Watermarking Authority in possession of all "secrets". The disadvantage of the Scheme II is the following; if the method of *obtaining the signature from the meaningful string* is fixed (as it should be - otherwise the whole purpose is defeated), then it may be possible for the pirate (Bob) to "guess" the meaningful string used by Alice, thus reducing security). In addition, both Schemes I and II suffer from the *fake original* problems discussed in Section 2.2. Moreover, both these methods do not provide good solutions for the multiple watermarks problem (Section 2.3).

In Ref. [4], Craver *et. al.* suggested a novel idea (Scheme III), which at one stroke solves the fake original problem *the need for an agency* to issue signatures, and the multiple watermarks problem. They suggested that the signature be obtained *from the original image itself.* The original image is hashed by a fixed hash function. The output is used as a seed for a *fixed random sequence generator* to generate the signature. Tying up the signature to the original image in an inextricable fashion goes a long way in restricting the freedom available for the pirate to engineer counterfeit attacks. The signature is obtained as $\mathbf{s}_A = \mathcal{G}(\mathcal{H}(I))$, where $\mathcal{H}$ is a fixed hash function and $\mathcal{G}$ is a fixed random sequence generator (RSG). More importantly, $\mathcal{H}(I) \neq \mathcal{H}(\hat{I})$. The watermarking scheme is non-oblivious described as

$$\mathbf{s}_A = \mathcal{G}(\mathcal{H}(I)) \quad \hat{\mathbf{c}} = \mathbf{c} + \mathbf{s}_A \quad \tilde{\mathbf{s}}_A = \tilde{\mathbf{c}} - \mathbf{c} \quad s_d = \frac{\mathbf{s}_A^T \tilde{\mathbf{s}}_A}{|\mathbf{s}_A||\tilde{\mathbf{s}}_A|}$$

The attack for this method rests on the fact that Bob can still search for a *combination* of a fake original and its corresponding signature. Bob, who has in his possession $\hat{I}$ (or equivalently $\hat{\mathbf{c}}$), could change $\hat{I}$ significantly, in the mean-square-error sense while maintaining the "visual similarity" between the original $\hat{I}$ and the resulting (modified) image I_m. Let I_d be the difference image $I_d = I_m - \hat{I}$ and $\mathbf{c}_d = \mathbf{c}_m - \mathbf{c}$. Even though the algorithms for undoing geometric distortions / histogram modifications / pixel rescaling would not permit Bob to move very far away from $\hat{I}$, he should be able to introduce distortions such that the total power of I_d ($\mathbf{c}_d$) is much larger than that of the signature S_A ($\mathbf{s}_A$) added by Alice. Alternately, $\sum_{i=1}^{N} c_d^2(k) >> \sum_{i=1}^{N}(c(i) - \hat{c}(i))^2$. Therefore $\mathbf{c}_d = \mathbf{c}_m - \hat{\mathbf{c}} \approx \mathbf{c}_m - \mathbf{c}$. The next step for Bob is to derive his "original" (fake original) image I_f, from I_m, just by tweaking 1-2 LSBs of I_m. Let us assume that Alice, using a very sophisticated watermarking method manages to detect her signature in I_f with $P_e \approx 3 \times 10^{-7}$, (or $s_d = 5\frac{1}{\sqrt{N}}$). To obtain a comparable detection statistic of his signature in I, Bob has to *search* roughly 3.3×10^6 sequences on an average, before obtaining a suitable signature. This is certainly feasible computationally.

5 Oblivious Watermarking

We would like watermarking schemes to be able to resolve ownership unambiguously even in the face of different types of attacks like scaling, rotation, pixel scaling, histogram changes, imperceptible geometric attacks, lossy compression etc. Furthermore, we would also like the watermarking scheme to be oblivious, as it would reduce overheads necessary for detection of signature. However, an ideal oblivious watermarking scheme may never be practical. Firstly, as mentioned earlier in Section 2.2, a strictly oblivious watermarking scheme, due to its vastly reduced degrees of freedom for the choice of signature, may not be able to yield satisfactorily low false alarm probability P_e. Secondly, as seen in the previous section, deriving the signature sequence from the original is a fundamentally more secure method.

The main necessity for a oblivious watermarking scheme arises out of the impracticality for the content creator to have access to his/her original, (especially for a professional content creators who may have thousands of watermarked content) to check for the presence of his/her signature in some content which may be present at a remote location. A typical scenario where this might be required is when the content is rendered by a *compliant player* [10] (which can check for watermarks). The content provider would like to confirm that the content being rendered is indeed his/her content.

The above problem can be solved by using two watermarks; one (which is not "valid" in court) used for checking the presence of his/her signature, and a second one which is generated from the original, which can "unambiguously" resolve ownership in court. However, use of two independent signatures is a waste of bandwidth available for watermarking. A solution to this problem is to permit the content provider to have multiple keys. At least one of them however, should be derived from the original. The others *have to be registered* with an appropriate authority. As a naive example, a content provider may use 2 keys - $\mathcal{K}_1$, a content independent key registered with some authority, and $\mathcal{K}_2$ obtained from hashing the original. The key $\mathcal{K}_2$ is used to generate a ± 1 sequence of length K. The seed $\mathcal{K}_1$ is used to generate a "short" signature sequence $\mathbf{q} \in \Re^{\frac{N}{K}}$. The signature sequence $\mathbf{s} \in \Re^N$ is obtained by the Kronecker product of $\mathbf{q}$ and the ± 1 sequence. With the key $\mathcal{K}_1$, the *compliant player* can determine (by obtaining sufficiently high magnitudes of detection statistics of the short signature sequences $\mathbf{q}$ generated from $\mathcal{K}_1$) the identity of the content owner. If the content is being used in an "unacceptable" fashion, the content provider may need to take legal action, and show unambiguously (by providing the pirated content along with the actual original in court), that he/she is the true owner of the content.

5.1 Optimal Detection Statistic

Once the pirated content and the original is produced in the court by the true owner, we shall see that obtaining the detection statistic s_d as in Eq. (6) may not be the best alternative. Making the detection statistic to be a combination of the detection statistics obtained from the image in question and from the unwatermarked original, can go a long way in rendering counterfeit attacks more difficult. As the watermarking scheme is oblivious, the watermark is detected *without subtracting* the original image. But the original image is still necessary because the seed is obtained from the original image as in Scheme III. Further we also need $\mathbf{p} = \mathcal{D}(\mathbf{c})$. Let $\mathbf{S}$, $\tilde{\mathbf{S}}$, and $\mathbf{P}$ represent normalized versions of $\mathbf{s}$, $\tilde{\mathbf{s}}$, and $\mathbf{p}$ respectively. (or $\mathbf{S} = \frac{\mathbf{s}}{|\mathbf{s}|}$). The detection statistic should be

$$s_d = \mathbf{S}^T\tilde{\mathbf{S}} - b\mathbf{S}^T\mathbf{P} = \mathbf{S}^T(\tilde{\mathbf{S}} - b\mathbf{P}), \tag{7}$$

where and b is a scalar whose optimal choice shall be determined shortly.

Let $P_i, \tilde{S}_i \sim \mathcal{N}[0, \frac{1}{N}]$, $i = 1, \ldots, N$. Obviously, as $\mathbf{c}$ and $\tilde{\mathbf{c}}$ are *not independent*, P_i and $\tilde{S}_i$, $i = 1, \ldots, N$ are not independent. Let $\rho = \mathbf{P}^T\tilde{\mathbf{S}}$ and $\mathbf{W} = \tilde{\mathbf{S}} - b\mathbf{P}$. It can be easily shown that $W_i \sim \mathcal{N}[0, \sigma_W^2 = \frac{1+b^2-2b\rho}{N}]$ [7]. As we should try to minimize the false alarm probability, we should try to minimize σ_W^2. The choice of b for minimizing σ_W is $b = \rho = \mathbf{P}^T\tilde{\mathbf{S}}$. Therefore the detection statistic should be

$$s_d = \mathbf{S}^T(\tilde{\mathbf{S}} - b\mathbf{P}) = \mathbf{S}^T(\tilde{\mathbf{S}} - \mathbf{P}^T\tilde{\mathbf{S}}\mathbf{P}). \tag{8}$$

Under this condition, $\sigma_W^2 = \frac{1-\rho^2}{N}$.

The pirate, on the other hand would try to maximize σ_W^2, to simplify his search. To do this he has to reduce ρ as much as possible, by moving I_f far away from I. Bob as earlier would claim that I_f ($\mathbf{c}_f$) to be the original and would like to search for a signature $\mathbf{s}_f$ which will yield a high detection statistic with $\mathbf{c}$. In other words, Bob's detection statistic is

$$s_{d_B} = \mathbf{S}_B^T(\mathbf{P} - \mathbf{P}_f^T\mathbf{P}\mathbf{P}_f) \tag{9}$$

where, $\mathbf{p}_f = \mathcal{D}(\mathbf{c}_f)$, and $\mathbf{P}_f$ is the normalized version of $\mathbf{p}_f$. For Type I oblivious methods, where $\mathbf{p} \equiv \mathbf{c}$ and $\mathbf{p}_f \equiv \mathbf{c}_f$, it may be very difficult to achieve low values of $\rho_B = \mathbf{P}_f^T\mathbf{P}$. The original and modified image the modified image after being subjected to regulated algorithms to undo distortions that might be introduced by the pirate will be more "similar" than not "similar" even in the MSE sense. As mentioned earlier, Type III methods for watermarking are closer to Type I. Therefore it may be extremely difficult for Bob to obtain low values for ρ_B.

Note that for the genuine creator (Alice), the second term in Eq. (8) is not going to affect the signature extraction process in any way (the random signature generated from the image is very unlikely to yield a high correlation with the original image). Let us assume Alice obtains a detection statistic of $s_d = \frac{5}{\sqrt{N}}$, of her signature in Bob's image I_f. For Bob to obtain comparable detection statistic, he had to search for $\frac{1}{Q(5)} \approx 3 \times 10^6$ signatures with Scheme III. However, for the oblivious scheme suggested, Bob has to search an average of $Q(\frac{5}{\sqrt{1-\rho_b^2}})^{-1}$. For $\rho_b = 0.6$, 0.65, 0.7, 0.75 and 0.8, Bob has to search for 4.9×10^9, 4.2×10^{10} , 7.9×10^{11}, 4.9×10^{13} and 2.5×10^{16} signatures respectively.

6 Conclusions

In this paper we addressed the need for watermarking, and more importantly, the need for a restrictive protocol for watermarking. As the amount of multimedia data created is expected to increase dramatically, it may become commonplace to find say, home videos, posted insecurely on the Internet. It is very necessary in this case to give such amateur artists the choice of *not securing their content*, and still be guaranteed immunity to claims of ownership on their content by pirates. The watermarking protocol should therefore make it extremely difficult for a pirate who obtains the content (without watermarks) to claim ownership of unwatermarked content.

Sophisticated attacks on watermarks like StirMark, have demonstrated that it may be extremely difficult to demonstrate the presence of the watermark in a *completely* oblivious fashion (with no knowledge of the original), under the face of severe attacks. Though some methods [9] can overcome attacks by StirMark, it may not be able to withstand *all* forms of attacks. So it is reasonable to expect that the creator keeps a copy of the true original, which may be used in conjunction with *regulated* algorithms to "undo" the efforts of the pirate to remove the watermark or engineer counterfeit attacks. With these algorithms, "conventional" watermarking algorithms can be used effectively.

Most often, however, the content owner *may not* need to use the regulated algorithms for undoing the distortions introduced by the pirate. Under such a scenario, the overhead for the content creator (for checking for the presence of his / her signature) can be substantially reduced if the watermarking scheme is oblivious. The modifications suggested in this paper (to the one proposed in [4]) is primarily to make the detection process "oblivious". However, to unambiguously establish ownership in court, the original is still necessary. When the signature is detected without subtracting the original image, and when the original image is available, we have shown that obtaining the detection statistic as a linear combination of the statistics obtained from the image in question and the original, as in Eq. (8), can go a long way in rendering counterfeit attacks extremely difficult. The suggested protocol is a list of restrictions to be placed on watermarking methods so that they meet the end requirement, viz. unambiguous resolution of ownership. The proposed protocol aims to aid the content creator to detect a genuine watermark while simultaneously making it very difficult for a pirate to remove the watermark or engineer a counterfeit attack or show the presence of a watermark in an unwatermarked content.

References

[1] I.J. Cox, J. Kilian, F.T. Leighton, T.G. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", *IEEE Transactions on Image Processing*, vol **6** (12) pp 1673-1687, 1997.

[2] W. Zeng, B. Liu, "On Resolving Rightful Ownerships of Digital Images by Invisible Watermarks", *IEEE International Conference on Image Processing*, vol **1** pp 552-555, Santa Barbara, CA, October 1997.

[3] M. Ramkumar, A.N. Akansu, "A Robust Scheme for Oblivious Detection of Watermarks/ Data Hiding in Still Images", *SPIE Multimedia Systems and Applications*, Boston, MA, vol **3528**, pp 474 - 481, November 1998.

[4] S. Craver, N. Memon, B-L. Yeo, M.M. Yeung, "Can Invisible Watermarks Resolve Rightful Ownerships", *IS & T/ SPIE Electronic Imaging: Human Vision and Electronic Imaging*, vol **3022**, pp 310-321, February 1997. Beach, CA, USA, pp 273-278, December 1998.

[5] M. Ramkumar, A.N. Akansu, "Image Watermarks and Counterfeit Attacks : Some Problems and Solutions", *Content Security and Data Hiding in Digital Media*, Newark, NJ, May 1999.

[6] R.B. Wolfgang, E.J. Delp, "Overview of Image Security Techniques with Applications in Multimedia Systems," *Proceedings of the SPIE Conference on Multimedia Networks*, Dallas, TX, vol **3228**, pp 297-308, November 1997.

[7] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third Edition, McGraw Hill Inc., 1991.

[8] M. Kutter, F.A.P. Petitcolas. "A Fair Benchmark for Image Watermarking Systems", *Electronic Imaging*, vol **3657**, pp. 226-239, San Jose, CA, USA, January 1999.

[9] P.M.J. Rongen, M.J.J.J.B. Maes, C.W.A.M. van Overveld, "Digital Image Watermarking by Salient Point Modification: Practical Results," *Proceedings of SPIE*, San Jose, CA, vol **3657**, pp 273-282, January 1999.

[10] M. Ramkumar, A.N. Akansu, "On the Design of Robust Data Hiding Systems", presented in the 33rd *ASILOMAR Conference on Signals, Systems and Computers*, Pacific Grove, CA, October 1999.

[11] M. Wu, B. Liu, "Watermarking for Image Authentication", *IEEE International Conference on Image Processing*, Chicago, IL, USA, vol **2**, pp 437 - 441, October 1998.

[12] H.H. Yu, A. Gelman, "Information Hiding System for Digital Video Playback and Record Control", *SPIE's International Symposium on Voice, Video and Data Communications*, Boston, MA, vol **3845**, September 1999.

[13] M. Ramkumar, A.N. Akansu, "Signaling for Multimedia Steganography", submitted to the *IEEE Transactions on Signal Processing*, November 1999.