Collusion-Resilient Fingerprinting by Random Pre-Warping

Mehmet U. Celik, *Student Member, IEEE*, Gaurav Sharma, *Senior Member, IEEE*, and A. Murat Tekalp, *Fellow, IEEE*

***Abstract*—Fingerprinting of audio-visual content using digital watermarks is an effective means of determining originators of unauthorized/pirated copies. Watermarks embedded in content can trace the traitor responsible for piracy. Multiple users may, however, collude and collectively escape identification by creating an average of their individually watermarked copies that appears unwatermarked. We propose a novel collusion-resilience mechanism, wherein the host signal is warped randomly prior to watermarking. As each copy undergoes a distinctive warp, collusion through averaging either yields low-quality results or requires substantial computational resources to undo random warps. The method is independent of the watermarking scheme used and imposes no restrictions on the watermark signal. We demonstrate the effectiveness of this approach on digital images.**

***Index Terms*—Collusion secure, traitor tracing, watermark.**

I. Introduction

THE TERM *fingerprinting* or *traitor tracing* refers to addition of a unique mark on each copy of distributed content. If and when an unauthorized copy of the fingerprinted content is made and distributed, the mark (fingerprint) embedded in the copy uniquely identifies the traitor. Historically, fingerprinting has been mostly utilized for secret documents. The small number of users in traditional settings and relatively high value of the documents have allowed for manual embedding and detection of fingerprints, typically by making minor, but distinct, edits to the document. In large-scale digital distribution networks, however, manual editing is prohibitive.

Digital watermarking [1], [2] offers an efficient method for automatic fingerprinting of content for distribution over large networks. A fingerprint in the form of a pseudonoise pattern representing user identity is embedded in the host signal, to produce the watermarked content distributed to the user. Later, when an unauthorized copy is found, the presence of a particular watermark pattern reveals the identity of the traitor who has compromised the content. Traitors may, however, attempt to escape identification by working collectively to disable the watermark. They may use each of their fingerprinted copies to obtain a copy in which fingerprint patterns cannot be detected reliably. Such collusion often involves averaging several distributed copies and results in a high-quality untraceable copy.

Collusion-secure watermarks proposed earlier [3], [4] impose restrictions on the construction of the watermark pattern to thwart collusion attacks. In these algorithms, the watermark patterns—or payloads—are partially correlated and any subset of these patterns bears a static component. This static component not only survives the averaging attack, but also uniquely identifies a particular subset of watermarks that were averaged—revealing the group of traitors. Nevertheless, security against collusion often comes in the expense of the capacity and/or robustness of the original watermarking algorithm [4], [5] and its performance degrades as the number of traitors increases.

An alternative approach to collusion-resilient fingerprinting is proposed here. The method imposes no restrictions on the watermark pattern and may be used in conjunction with any watermarking method. In contrast with schemes that detect collusion and identify colluders, the proposed scheme is aimed at preempting collusion by preventing traitors from obtaining a high-quality copy through collusion. In our approach, the geometry of the host-signal is randomly and imperceptibly distorted prior to distribution. If a group of users collude and average multiple distinctively warped copies of the content, i.e., the mismatch between the underlying geometries, makes the resulting averaged version inferior in quality.

II. Fingerprinting and Collusion Attack

We first review spread-spectrum watermarking for fingerprinting and describe the collusion attack. Both the collusion attack and the proposed solution are equally applicable for alternative (e.g., quantization based) watermarking methods.

Consider a continuous host signal S upon which a digital fingerprint W_i (uniquely identifying a particular user) is to be imposed. The watermarking procedure superimposes the watermark pattern W_i on the host signal and yields the fingerprinted signal S_i

$$S_i = S + W_i \tag{1}$$

for all $i \in \{1, \ldots, N\}$, where N is the total number of users.

When an unauthorized copy of the host signal S_u is found, the presence of the particular watermark is checked using a correlation detector. That is W_i is present in the signal if

$$\langle S_u, W_i \rangle > \text{threshold}. \tag{2}$$

It is desirable to design watermark patterns W_i such that they are uncorrelated with the host signal and with each other. If $S_u =$

Manuscript received October 3, 2003; revised April 9, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Truong Q. Nguyen.

M. U. Celik and G. Sharma are with the Electrical and Computer Engineering Department, University of Rochester, Rochester, NY 14627-0126 USA (e-mail: celik@ece.rochester.edu; gsharma@ece.rochester.edu).

A. M. Tekalp is with the Electrical and Computer Engineering Department, University of Rochester, Rochester, NY 14627-0126 USA. He is also with The College of Engineering, Koc University, Istanbul, Turkey (e-mail: tekalp@ece.rochester.edu)..

Digital Object Identifier 10.1109/LSP.2004.835475

S_i, the correlation is randomly distributed around $\|W_i\|$ and around zero otherwise. The threshold determines the tradeoff between the false positives and misses.

To elude identification and prevent watermark detection, multiple users may obtain an average signal using their individually watermarked copies. For K colluding users, the average signal is

$$S_{\text{avg}} = \frac{1}{K}\sum_{i=1}^{K} S_i = S + \frac{1}{K}\sum_{i=1}^{K} W_i. \tag{3}$$

The correlation between the average signal S_{avg} and a particular watermark pattern W_j, $\langle S_{\text{avg}}, W_j\rangle$ decreases linearly with K, the number of copies used. By using sufficiently large K, the correlation value can be made smaller than the threshold, thereby evading detection. Moreover, the quality of the of the averaged signal is often superior when compared with the fingerprinted signal $d(S_{\text{avg}}, S) < d(S_i, S)$, where $d(\cdot,\cdot) = \|\cdot - \cdot\|$ is the Euclidean distance.

III. Collusion-Resilience by Random Pre-Warping

Consider a set of warping functions Φ such that for all functions, $\phi_i(\,\cdot\,) \in \Phi, \phi_i(S)$ is perceptually identical to S, but the Euclidean distance between the signals is significantly large. We further require that the Euclidean distance between different warped versions of the same signal is large (minimum distance constraint). The set of functions Φ may be constructed, for instance, by small local distortions on the geometry of the signal. In most cases, the human perceptual system is highly tolerant of such manipulations and the overall effect is imperceptible, despite the large mean-squared-error (mse) distortion among signals. This property has been previously exploited in Stirmark [6] where it forms the basis of de-synchronization attacks on watermarks. We exploit the same property to provide collusion resilience.

A. *Collusion Resilience for Oblivious Watermarking*

In oblivious watermarking, the watermark is detected without reference to the original. Collusion resilience is incorporated by applying a different, randomly selected warping function $\phi_i(\,\cdot\,)$ to the host signal prior to addition of the watermark W_i. The ith watermarked signal is formed as

$$S_i = \phi_i(S) + W_i. \tag{4}$$

Consider, for instance, a video stream where S is a function of two-dimensional spatial coordinates x, y, and time t. The geometric warping is applicable in the three-dimensional spatio-temporal space as

$$S_i(x,y,t) = S(x',y',t') + W_i \tag{5}$$

$$x' = \phi_i^x(x,y,t); \quad y' = \phi_i^y(x,y,t);$$
$$t' = \phi_i^t(x,y,t). \tag{6}$$

The requirement of imperceptible visual distortion then imposes a smoothness constraint on the transformation in (6).

Watermark detection is performed on a suspected copy S_u, through a correlation detector as earlier (2). As the warped host signal bears statistical characteristics similar to the original host signal, performance of the detection is not affected by the warping.

As in the earlier scheme, several users may try and collude and obtain an average host signal in order to thwart the watermark. The average signal from a collusion attack by K users is given by

$$S_{\text{avg}} = \frac{1}{K}\sum_{i=1}^{K} S_i = \frac{1}{K}\sum_{i=1}^{K}\phi_i(S) + \frac{1}{K}\sum_{i=1}^{K} W_i. \tag{7}$$

The averaged watermark terms in (7) and (3) are identical: indicating equal impairment for watermark detection. However, the corresponding "signal" terms in (7) and (3) are quite different. The averaging in (3) causes no degradation of the signal, whereas due to the warping the averaging in (7) typically produces an averaged copy often of significantly inferior quality, with perceptually significant and disturbing artifacts. For instance, if the host signal is an image the average appears to have multiple ghost images or is a blurred version. In either case, the colluded copy is a rather low quality version.

The proposed scheme does not guarantee absolute security against collusion. Given a number of copies, it is possible to undo the warping to align all signals to a common geometry and then average the registered copies. Nonetheless, this requires significant additional effort on the part of the colluders in terms of computation time. For instance, it takes approximately 4 min on a Pentium-4 system to register a 512 × 512 image using the software presented in [7].

B. *Collusion Resilience for Nonoblivious Watermarking*

The above method can also be used with nonoblivious watermarks that utilize the original signal for detection. In this case, warping may be performed either *before* or *after* watermark addition. If warping is applied a priori, the warped unwatermarked signal has to be present at the decoder. This can be achieved either by storing the warped signals or the warping parameters along with the original signal. Alternately, if warping is applied after watermarking, it must be undone before detection to prevent loss of synchronization. Though undoing random warps is challenging, it can be significantly simplified by using a pre-determined subset of all possible warping parameters. The detector performs a limited search over this small subset, unless the watermarked signal is warped again by a third party. In the latter case, the proposed scheme does not increase the computational burden. The complexity of recovering from the combination of the first and second warps is similar to that of recovering from the unknown second warp.

C. *Distinctness of Random Pre-Warping Functions*

In fingerprinting applications with a large number of users, the design of warping functions poses a tradeoff. Warping functions must introduce small or no perceptual distortion while being sufficiently different from one realization to the next (to defeat collusion). Estimation of the number of such distinct functions involves significant psychophysical testing and is not attempted here. We do note that both the perceptual quality and the Euclidean distance between the warped signals are

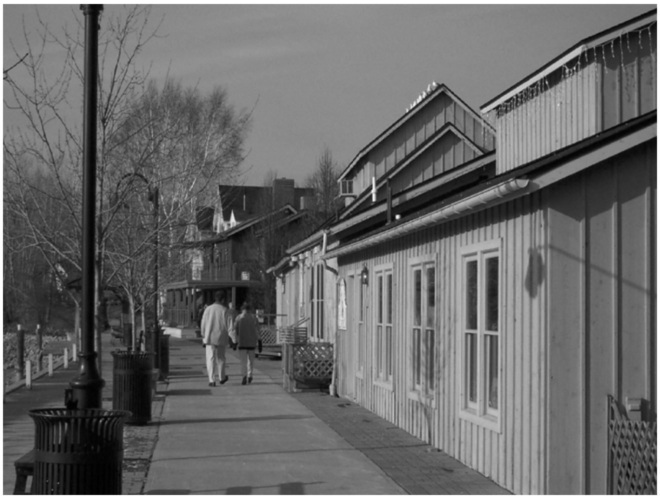

Fig. 1. Original image (800 × 600 pixels).

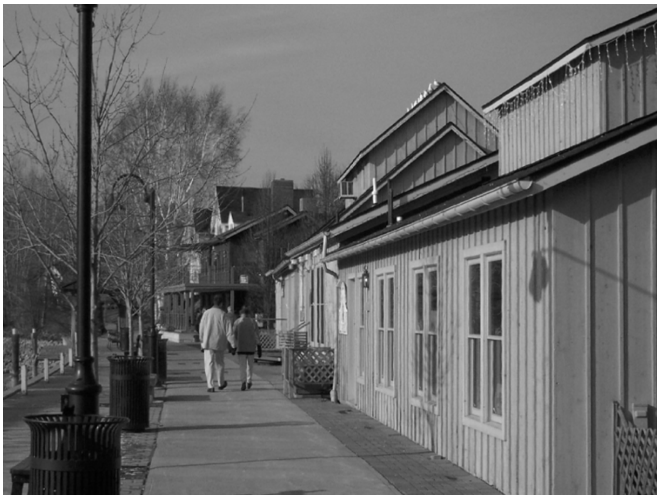

Fig. 2. Image after warping by Stirmark [6] and watermark addition. Although PSNR is low (18.1 dB), the distortion is perceptually tolerable.

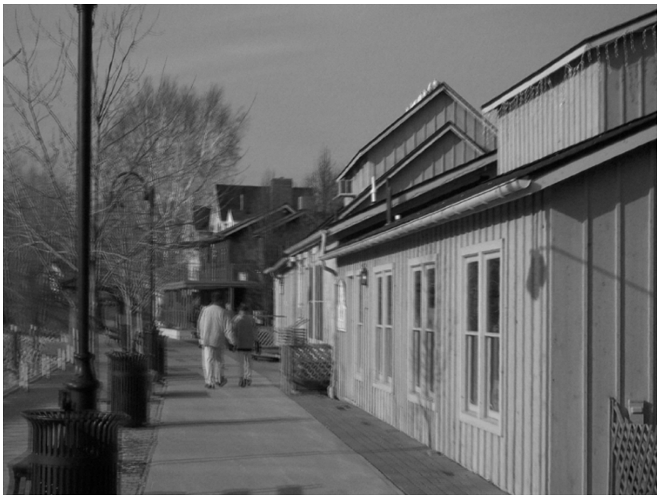

Fig. 3. Collusion result when two warped images are averaged. Ghosting around edges is visible and disturbing.

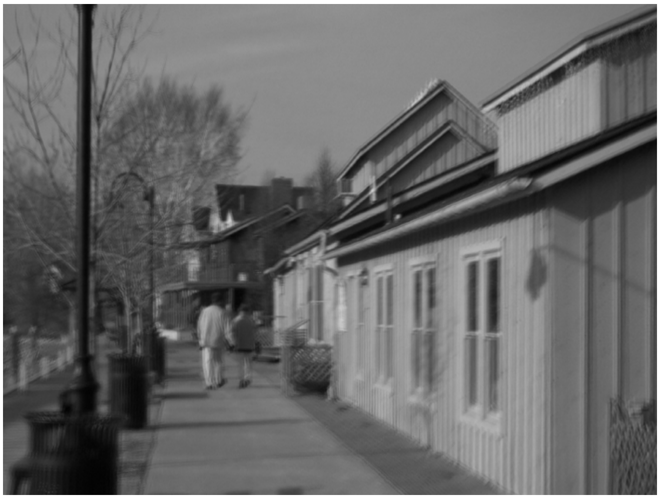

Fig. 4. Collusion result when nine (9) warped images are averaged. The image is blurred and bears little commercial value.

content dependent. For the same pre-warping, low-pass signals are more likely to have better perceived quality, however, their separation is also smaller making collusion more effective. The amount of warping may itself be adjusted based on the signal content and a test for distinctness. In Section IV, we include results for empirical tests of distinctness.

IV. Implementation and Experimental Results

We demonstrate the transparency and effectiveness of the proposed anti-collusion algorithm on digital images. We apply geometric distortions using Stirmark [6]. A simple spatial domain spread-spectrum watermarking system is chosen for illustration. The watermark is additive white Gaussian with variance σ_W^2. (σ_W^2 was set to 9, corresponding to 38.6 dB embedding distortion.) Detection is performed using normalized correlation (9), which is preceded by Wiener filtering (8)

$$\hat{W} = \frac{\sigma_W^2}{\sigma_{S_u}^2 + \sigma_W^2} \left(S_u - \mu_{S_u}\right) \tag{8}$$

$$NC = \langle \hat{W}, W \rangle / \sqrt{\langle \hat{W}, \hat{W} \rangle \langle W, W \rangle}. \tag{9}$$

Here, $\mu_{S_u}, \sigma_{S_u}^2, \sigma_W^2, \hat{W}$ are the local mean and variance of the unknown signal, variance of the watermark signal, and the watermark estimate, respectively. $\langle \cdot, \cdot \rangle$ denotes the dot product. If $\hat{W} = W$, normalized correlation is 1. However, the watermark estimate is often corrupted with noise interference from the image $\hat{W} = W_i + N$. Similarly, $\hat{W} = W_i/K + N$ when K-copies of the content are averaged. Assuming that the noise and the watermark patterns are uncorrelated and the noise-to-watermark ratio is constant, i.e., $c_0 = \langle N, N \rangle / \langle W_i, W_i \rangle$, then (9) simplifies to

$$NC = 1/\sqrt{1 + c_0 K^2}. \tag{10}$$

The visual impact of the proposed scheme is first demonstrated. Fig. 1 shows the original image. Fig. 2 shows a representative watermarked image obtained by distorting the image geometry with Stirmark (using a random seed) and adding a key-dependent pseudonoise watermark pattern. Default settings are used for StirMark (version 3.1.79), with the exception of global bending and JPEG compression, which are turned

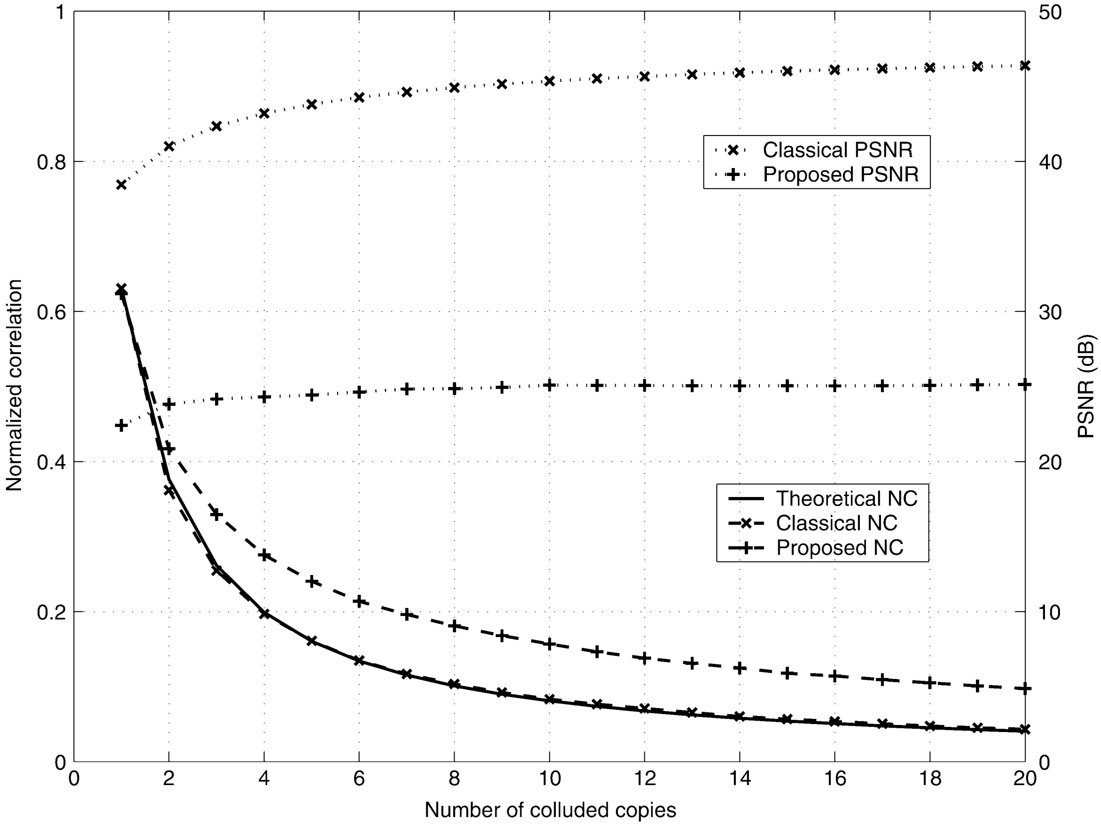

Fig. 5. Averaged PSNR and normalized correlation of the colluded copy versus number of colluders (with and without pre-warping).

off (command line options "-NOJPEG-b0.0"). The image in Fig. 2 appears identical to the original despite a rather low peak signal-to-noise-ratio (PSNR) (18.1 dB). Fig. 3 shows the image from a collusion averaging attack employing two images. The visual quality of the image is rather poor showing clear ghosting around the edges (PSNR $= 20.4$ dB). Fig. 4 shows the image from a collusion attack using nine watermarked images. The image is blurred and its perceptual quality is significantly degraded (PSNR $= 21.0$ dB). The normalized correlations corresponding to images of Figs. 2–4 are 0.39, 0.26, and 0.11, respectively.

We also test watermark performance with and without the pre-warping over 23 images (768×512) from Kodak Photo-CD. Fig. 5 summarizes PSNR and normalized correlation results averaged over all images and over 10 iterations with different keys. The PSNR increases with increasing number of colluders. Without pre-warping, this corresponds to a significant quality improvement (up to 6 dB in PSNR). In the presence of pre-warping, the PSNR metric based on pixel-wise differences is not meaningful. Despite recent efforts [8], there are no widely used quality metrics for geometrically distorted images.

Normalized correlation values in Fig. 5 decline sharply with increasing number of colluders, potentially limiting the capacity/robustness of the underlying fingerprinting system (as expected). Without pre-warping the results are in good agreement with theoretical computation based on (10). Surprisingly, the proposed anti-collusion feature improves detection—in addition to degrading visual quality of the colluded copy. This unintended gain is obtained because the Wiener filter is more effective in filtering out the blurred image produced by averaging independent warps. This reduces noise-to-watermark ratio c_0—improving normalized correlation with respect to the case with no warping. These results may not generalize to other watermark embedding and detection systems.

We also tested the distinctness of warping functions produced by the StirMark algorithm, using the same set of images. For each image, we generated hundred distinct warps using different (consecutive) seeds. For each of the $\binom{100}{2} = 4950$ possible pairs of warped images, we computed the PSNR between the pair as a measure of distinctness. The mean, standard deviation, and the 90th percentile of these pairwise PSNRs were then computed for each test image. The mean PSNR values ranged from 15–26 dB across the 23 images. As anticipated in Section III-C, the variation produced by the warping functions is dependent on the image content. The standard deviation of PSNR values for a given image, on the other hand, is around 1.5 dB, and the 90th percentile point lies within 2.5 dB of the mean. The pairwise difference between any pair of 100 randomly generated warps of an image is therefore quite comparable and is not significantly smaller than difference between the first pair. Through visual observation, we verified that the average of the first pair of warped versions produces a poor quality output for each of the 23 test images. Thus even the default parameterized warpings of StirMark allow for a reasonably large number of distinct warping functions. A system for collusion resilience based on random pre-warping is therefore scalable for large-volume distribution. If necessary, a distinctness test may be added at the embedding end to improve performance, wherein each warped image is checked against the others and used only if it satisfies preset distinctness requirements.

V. Conclusion and Discussion

We present an alternative method for collusion-resilient watermarking. Our approach is based on randomly and uniquely pre-warping each copy of the host-signal prior to distribution. Human perception is quite tolerant of the geometric distortion, and the warping therefore does not significantly affect the perceived quality of the watermarked signal. On the other hand, as the geometry of each copy is distorted independently, a collusion attack yields a low-quality signal. We show that collusion even with only two copies results in disturbing distortions—ghost edges—and visual quality does not improve when the number of copies is increased. Higher-quality collusion is only possible by undoing the warping which requires special software and substantial computational resources. Finally, the solution does not adversely affect the performance (capacity/robustness) of the underlying watermarking scheme. The approach can either replace existing collusion resistant watermarks or it can be applied in conjunction to even identify traitors who choose to use the low-quality average signals.

References

[1] G. Langelaar, I. Setyawan, and R. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Processing Mag.*, vol. 17, pp. 20–46, Sept. 2000.

[2] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proc. IEEE*, vol. 87, pp. 1079–1107, July 1999.

[3] J. Dittmann, A. Behr, and M. Stabenau, "Combining digital watermarks and collusion secure fingerprints for digital images," in *Proc. SPIE Security and Watermarking of Multimedia Content I*, vol. 3657, Jan. 1999.

[4] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, no. 5, pp. 1897–1905, Sept. 1998.

[5] J. Su, J. Eggers, and B. Girod, "Capacity of digital watermarks subjected to an optimal collusion attack," in *Proc. EUSIPCO 2000*, Tampere, Finland, Jan. 2000.

[6] F. Petitcolas, R. Anderson, and M. Kuhn, "Attacks on copyright marking systems," in *Proc. Information Hiding Workshop (IH'98)*, Portland, OR, Apr. 1998, pp. 219–239.

[7] P. Loo and N. Kingsbury, "Motion estimation based registration of geometrically distorted images for watermark recovery," in *Proc. SPIE Security and Watermarking of Multimedia Content III*, San Jose, CA, Jan. 2001, http://www-sigproc.eng.cam.ac.uk/pl201/watermarking/reg.exe.

[8] I. Setyawan, D. Delannay, B. Macq, and R. Lagendijk, "Perceptual quality evaluation of geometrically distorted images using relevant geometric transformation modeling," in *Proc. SPIE Security and Watermarking of Multimedia Content V*, San Jose, CA, Jan. 2003.