NRL Release Number 04-1221.1-0191

Correlation-based watermarking method for image authentication applications

Farid Ahmed
The Catholic University of America
Department of EECS
Washington, D.C. 20064

Ira S. Moskowitz
Naval Research Laboratory
Center for High Assurance Computer Systems—5540
Washington, D.C. 20375

Abstract. We propose a correlation-based digital watermarking technique for robust image pattern authentication. We hide a phase-based signature of the image back into its Fourier magnitude spectrum in the embedding stage. The detector computes the Fourier transform of the watermarked image and extracts the embedded signature. Authentication performance is measured by a correlation test of the extracted signature and the signature computed from the watermarked image. The quality of the watermarked image is obtained from the peak signal-to-noise ratio metric. We also furnish simulation results to show the robustness of our approach to typical image processing as found in JPEG compression.
[DOI: 10.1117/1.1763589]

Subject terms: authentication; watermark; binary phase-only filter; correlation detector; bit-plane embedding.

Paper TPR-006 received Aug. 28, 2003; revised manuscript received Jan. 19, 2004; accepted for publication Feb. 17, 2004.

1 Introduction

The rapid growth of multimedia applications has attached critical importance to digital pattern recognition and retrieval techniques. The spectacular surge of internet-based applications has made the problem even more challenging due to the inherent threats and vulnerabilities. Data over networks are constantly subject to active attacks such as blocking, spoofing, and tampering. Transmission and distribution of images over the internet, therefore, necessitates the authentication and integrity of digital media, in addition to the usual quality-of-service requirements. In this paper, we address this problem of self-authentication of digital image transmission using a signature-based information-hiding technique.

Authentication in cryptography has been addressed by attaching a digital hash-based signature to the message data during transmission and then comparing the computed and extracted signatures[1] in the receiver. An image signature, on the other hand, is an identifier of an image obtained from inherent features of the image, which can similarly be used for image authentication. Unlike digital signatures, though, image signatures are usually computed from some content-dependent features of the image.[2] Image authentication typically involves hiding this signature in the host image using a digital watermarking[3] technique. Depending on the application, a watermark can be *fragile*, *semifragile*, or *robust*. Most of the image authentication techniques proposed are semifragile,[4,5] in the sense that they are vulnerable to some malicious tampering, while at the same time they tolerate some desirable nonmalicious image modification, such as JPEG compression. There are other techniques as well that make use, with varied success, of variants of content-based extracted signatures such as robust watermarks,[6] fragile watermarks,[7] and also hybrid watermarks.[8] However, all the techniques heavily depend on the appropriate choice of signatures. In a related approach in optics, significant research has been done in optical security and authentication systems using variants of phase encoding[9] and Fourier plane encoding[10] techniques. The downside of all these approaches is that the transmitted images are usually complex-valued, hence necessitating bandwidth-intensive transmission. Also, some of these techniques require a reference image, which is often not available in a distributed system. Our method makes use of a correlation-based signature to produce a semifragile watermark for authentication applications. More precisely, we use the *binary phase-only filter* (BPOF)[11,12] of the original image for our signature, which is used in a self-authentication scenario that does not necessitate the transmission of the reference pattern.

Section 2 describes one particular motivation for our phase-based approach, followed by the watermarking procedure in Sec. 3. Simulation results are furnished in Sec. 4.

2 Motivation

Since the fragility of the signature depends on the fragility of the selected features, it is crucial to select good features. Our choice of signatures in embedding is motivated by the desire for correlation-based recognition in the detector. Correlation-based pattern recognition is a rich and established area in optical information processing. Whether it is matched filtering or a joint transform correlation setup, a reference filter is first computed from the given image(s). It is then correlated with the test image to determine detector recognition. Often the reference filter is not available. This is especially true for internet-based transmission and distribution of digital media. Consequently, in this paper we propose a self-recognition or self-authentication technique, where the reference image is embedded in the host image in a manner such that the quality of the host is only trivially degraded, while at the same time the embedded information

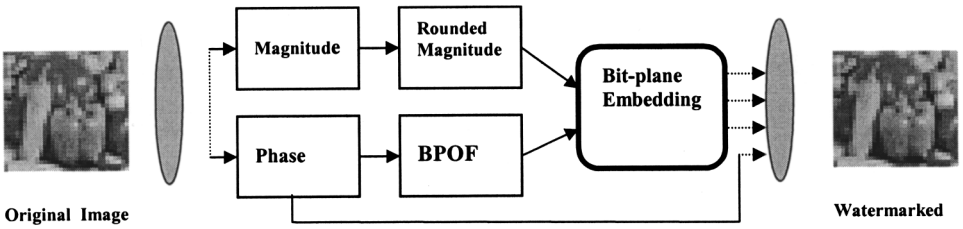

Fig. 1 The BPOF-based embedding method.

can be extracted to perform correlation tests for recognition and/or authentication. Our method also prevents spoofing of this embedded information.

The binary phase-only filter (BPOF) proposed independently by two research groups[11,12] has demonstrated optimal correlation performance, while at the same time satisfying storage and computational efficiency requirements. In a related as investigation, variants of phase-only filters were also shown to be good features for pattern recognition applications.[13] In addition, the phase quantization in the BPOF has built-in tolerance to minor changes to the image, which is a desirable aspect for a semifragile image authentication watermark. For these reasons, we propose to use the BPOF as a *signature* of the image, which will subsequently be hidden as the watermark signal.

3 The Watermarking Process

Figure 1 shows the flow chart of the watermark embedding process. The original image is transformed from the spatial domain to the frequency domain via the discrete Fourier transform (DFT). Consider an $M \times N$ original image $h(m,n)$, where m, n are the spatial indices. The DFT of $h(m,n)$ is written as $H(u,v)$, where u, v represent the spatial frequency coordinates:

$$H(u,v) = X(u,v)\exp[j\phi(u,v)]. \tag{1}$$

Here, $X(u,v)$ is the *magnitude* of the frequency coefficient, $|H(u,v)|$, and $\phi(u,v)$ is the *phase* part of the frequency $H(u,v)$, given in the standard manner as follows:

$$\phi(u,v) \text{ is the unique angle in } (-\pi,\pi] \text{ such that Eq. (1) is true.} \tag{2}$$

In our watermarking method, the phase is kept unchanged but the magnitude $X(u,v)$ is modulated. The real-valued $X(u,v)$ array is first transformed to an integer-valued array by

$$R(u,v) = \text{round}(X(u,v)). \tag{3}$$

The round $(\cdot)$ function rounds the operand to the nearest integer value, which can be represented by a fixed number q of bit planes. Hence, writing in a bit slice format, we have $R = R_{q-1}, R_{q-2}, \ldots, R_1, R_0$, where R_i is the ith bit plane of the rounded magnitude.

The second input to the bit-plane embedder is the BPOF, which is obtained by first binarizing the phase-only filter according to the following schedule:

$$b(u,v) = \begin{cases} +1 & \text{if } \cos\phi(u,v) \geqslant 0, \\ -1 & \text{otherwise.} \end{cases} \tag{4}$$

(At this stage we have therefore a bipolar binary image of the original image, which has been shown to possess a number of discriminatory features of the original image.[11,12]) Next we map this $\{+1,-1\}$ pattern respectively to a unipolar binary pattern $\{1,0\}$, which is equivalent to a 1-bit phase quantization. We denote this unipolar array as $B(u,v)$, and it is our BPOF.

As shown in Fig. 1, the embedding box now has two inputs—the rounded magnitude spectrum of the original image $[R(u,v)]$, and the BPOF signature $B(u,v)$. In order to employ bit-plane embedding, we need to decide which bit plane(s) will be modified and how. (Depending on the desired robustness, more than one bit-plane may also be modified. However, we do not discuss that possibility in this paper.) This decision is to be made from an optimal trade-off of the quality degradation of the image and the robustness of the embedded signal, which suggests one of the mid-level bit planes may work best. In the results section, we show how we came up with a good formulation of this choice. For now, it suffices to specify a particular bit plane w. The following equation then captures how this selected plane(s) is modified:

$$\tilde{R}(u,v) = C(u,v) \bullet B(u,v). \tag{5}$$

Here $\bullet$ denotes an invertible logical operation on the BPOF signature bit plane $B(u,v)$ and any other bit plane $C(u,v)$ of R. Note that the BPOF signature has the same symmetry aspects as the magnitude bit planes (from the properties of the discrete Fourier transform). The $\bullet$ operation must respect that symmetry. The $\bullet$ operation we used in our experiments was:

1. First encrypt $B(u,v)$, yielding $EB(u,v)$, with a nonavalanche[1] encryption that respects the symmetries of $B(u,v)$, and then
2. pick a w out of $\{0,\ldots,q-1\}$, and set $C(u,v)$ equal to the wth bit plane of $R = R_{q-1}, R_{q-2}, \ldots, R_1, R_0$.
3. Replace the wth bit-plane $C(u,v)$ of the rounded magnitudes with the $EB(u,v)$.

This results in a modification of the rounded bit planes from $R_{q-1}, R_{q-2}, \ldots, R_1, R_0$ to $R_{q-1}, \ldots, R_{w+1}, EB(u,v), R_{w-1}, \ldots, R_0$. We designate the modified rounded magnitudes as $\tilde{R}(u,v)$. (Note that other choices of $\bullet$ are plausible,

provided they respect the symmetry.*) Note that encryption is used to prevent spoofing. We then multiply the integer $\tilde{R}(u,v)$ by $\exp[j\phi(u,v)]$ to form $\tilde{H}(u,v)$, thus modifying the frequency representation of the original image:

$$\tilde{H}(u,v)=\tilde{R}(u,v)\exp[j\phi(u,v)]. \tag{6}$$

Next we apply the inverse discrete Fourier transformation to Eq. (6) to obtain the marked image. This will be a real-valued matrix with some of the values falling outside the range {0,...,255}. Therefore rounding, clamping, and clipping operations are performed on the marked image, resulting in the final watermarked image $h_w(m,n)$, which is representable as an unsigned 8-bit gray value. (One can also use the same method for color images by modifying only the *Y* values.) These operations, along with the subtle change of the rounded magnitudes, cause the watermarked image to differ slightly from the original image.

Optics-based authentication systems, where the watermarked image is usually a complex image, are obviously unsuitable for our needs. One of our requirements is to send a true image that is visually indistinguishable from the original image.

The detection process includes the extraction of the signature and a subsequent correlation operation. We start with a test image $t(m,n)$. The discrete Fourier transform of this yields

$$T(u,v)=|T(u,v)|\exp[j\phi_T(u,v)]. \tag{7}$$

After a rounding operation on magnitude coefficients, we extract the candidate hidden signature B' by undoing the operation • in Eq. (5). B' is then converted, in the obvious way, to the bipolar binary form $\{-1,+1\}$ designated as b'. Of course, as yet we do not know if b' is a valid signature from a legitimate watermarked image, or noise from an unwatermarked, or modified, image. This is why we refer to it as a *candidate signature*. Therefore, we must perform tests to check the validity of b' as a valid watermark signature. We do this by a correlation test.

We correlate the candidate signature b' against phase information from $t(m,n)$. The motivation behind this is that if $t(m,n)$ were a watermarked image, the watermarking process was designed so that magnitudes would be changed, but not phases. Of course, there are some minimal effects on the phase due to rounding, clipping, clamping, and also sometimes compression (if the image is stored as a JPEG). Correlation tests, though, should show a linkage of the candidate signature with the phase information from $t(m,n)$. Our experiments show that this is true. From

*If the operator is XOR and $C(u,v)$ is an all-zero bit plane, then the embedding is simply equivalent to replacing the selected bit plane with the signature plane. In another scenario, if the operation is XOR and $C(u,v)=R_i(u,v)$, and furthermore the signature does not degrade much, then Eq. (5) can be used to retrieve the original bit plane and thus the original image. This is an example of reversible watermarking. In yet another realization, $C(u,v)$ may represent any bit plane other than the embedding plane, which gives an additional degree of security.

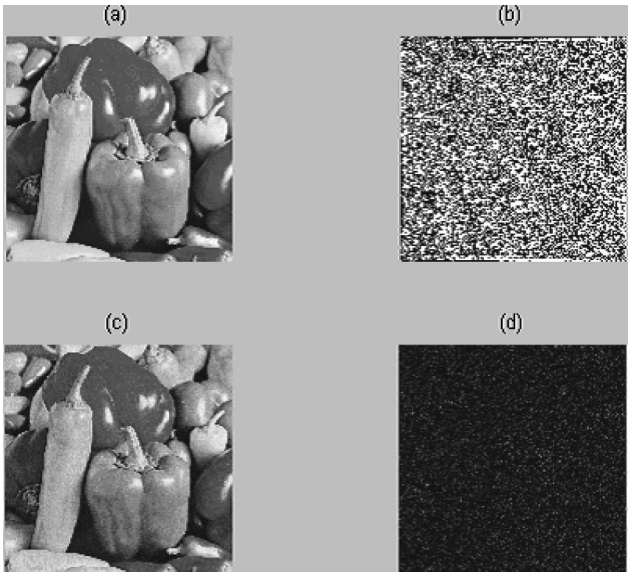

Fig. 2 (a) The original "Peppers" image, (b) the BPOF signature, (c) the watermarked image, and (d) the difference between original and watermarked image (visually enhanced).

$t(m,n)$ the phase information $\phi_T(u,v)$ is extracted. This is used to form a POF of $t(m,n)$. The POF is computed as

$$T_{\mathrm{POF}}(u,v)=\exp[-j\phi_T(u,v)]. \tag{8}$$

It is well known that the correlation of two spatial images is given by the inverse discrete Fourier transform (FT^{-1}) of the term-by-term product of the discrete Fourier transform of one image with the conjugate [hence the negative sign in Eq. (8)] discrete Fourier transform of the other image. We wish to see how similar the candidate watermarked image is to the original image; thus we would like to do an autocorrelation test. Unfortunately, we do not have the original image; we only have the watermarked image. However, we have hidden and extracted (up to noise from rounding and compression) the BPOF of the original image in the watermarked image. Therefore, based upon this and the already mentioned references on Fourier optics (especially Ref. 12), we propose the following as our correlation test for detection: We define the correlation function Corr as†

$$\mathrm{Corr}(u,v)=\mathrm{FT}^{-1}(T_{\mathrm{POF}}(u,v)\bullet b'(u,v)). \tag{9}$$

The correlation peak determines the degree of authenticity. We also use a few other metrics from the correlation plane to demonstrate the detector performance, which is enumerated in the next section.

4 Simulation and Results

Simulation of the above-mentioned algorithm is performed on two sets of images obtained from the USC SIPI Database.[14] The first set contains ten 256×256 images,

†Of course, Corr is actually a function of (u,v). Sometimes we suppress that in our notation for the sake of convenience.

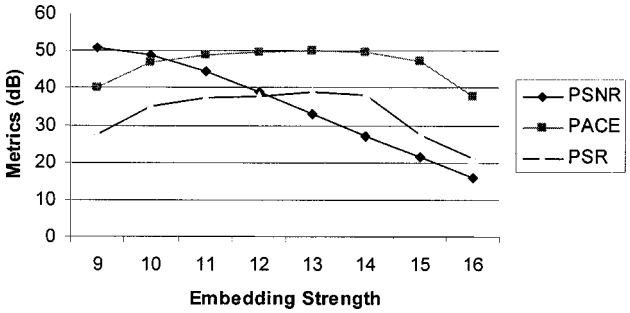

Fig. 3 Trade-off between the perceptual quality and robustness for "Peppers."

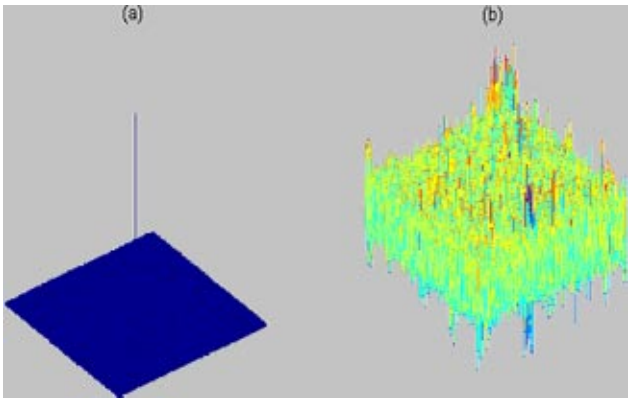

Fig. 4 Correlation output for (a) watermarked image, (b) unmarked image.

while the second set has fifteen different 512×512 images. There is a mix of natural, aerial, texture, and motion frame images in these sets. Note that some of the images are color images that we converted to grayscale in the standard manner. Although our algorithm is equally applicable to color images, the results furnished in this section are obtained from watermarking the intensity images only.

Figure 2 shows the original 512×512 grayscale "Peppers" image (a), its BPOF signature (b), the watermarked image (c), and the difference between the marked and unmarked images in (d) (which appears as noise).

Watermarking algorithms typically make a trade-off among a number of performance parameters such as robustness, perceptual quality, assurance, and detectibility. As for robustness, watermarks can be made very robust, but the resulting perceptual quality degradation of the image may make it useless. In order to compare the perceptual quality objectively, we use the widely used peak signal-to-noise ratio (PSNR).

In the current work, authentication of a watermarked image is used as measure of assurance; this in turn is measured by the degree of correlation of the extracted signature with the computed phase signature. Let $P_{\max}$ and P_{second} denote the highest and the second highest peaks of the correlation plane, respectively, as obtained from Eq. (9). Note that the second highest peak is usually calculated excluding a small (3×3 to 7×7) pixel area centered on the highest peak. Let μ be the average value of the entire correlation plane. The first metric is the ratio of the peak to average correlation energy (PACE).[15] This is a measure of the sharpness of the peak. We calculate this ratio in decibels by expressing it as

$$\text{PACE} = 20 \log_{10}\left(\frac{P_{\max}}{\mu}\right). \tag{10}$$

Next we calculate the peak-to-secondary-peak ratio (PSR), given by

$$\text{PSR} = 20 \log_{10}\left(\frac{P_{\max}}{P_{\text{second}}}\right). \tag{11}$$

This is a measure of the prevalence of false positives in a detection algorithm. A higher PSR value indicates a false positive is less likely.

Robustness of our method depends on the location of the embedding bit plane(s) and the number of bit planes being modified. For the results to follow, as discussed earlier, we only use one bit-plane embedding. Therefore, the embedding strength is directly proportional to the location of the bit plane modified. In Fig. 3, the x axis represents which bit plane $C(u,v)$ is [see Eq. (5)].

In order to select this optimal bit plane, we ran a simulation to obtain a relation of embedding strength versus detector performance metrics. Figure 3 shows this relationship for embedding planes 9 to 16 of the 512×512 "Peppers" image. The total number of bit planes in this case was 25. Note the monotonic decrease of the image quality (PSNR) with increasing embedding strength. The correlation metric PACE achieves a slight maximum at bit plane 13, while dropping down on each side. This is interesting. For a low embedding strength the extracted signature is

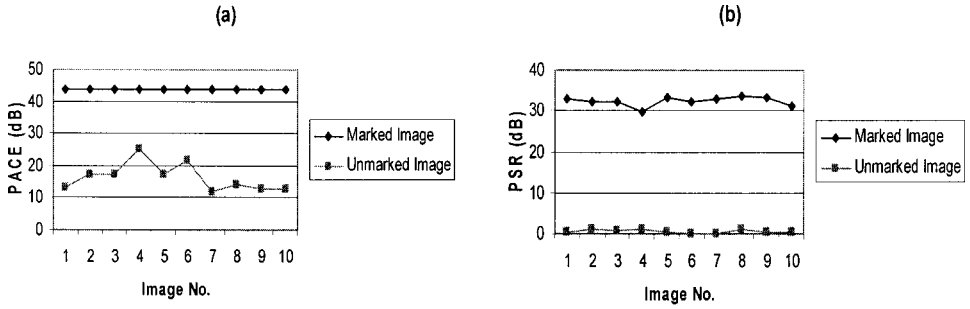

Fig. 5 Correlation performance for image set 1: (a) PACE, (b) PSR.

noisier than the computed one. For a high embedding strength, the computed one is noisier than the extracted one. Therefore, in either case the correlation becomes noisier and so do the metrics. A similar trend is also observed for the second correlation metric PSR. This experiment (done on ''Peppers'' and other images) shows that an optimal trade-off value for a selected bit plane i from a total of q planes is given by

$$i=\left\lceil\frac{q+1}{2}\right\rceil. \tag{12}$$

Using this relation in the Fig. 3 experiment, where the total number of bit planes was 25, we find that $i=13$ is the selected bit plane. For all the remaining experiments, except where noted, we use Eq. (12) to determine the selected bit plane.

Next, we look closely at the detector performance. Figure 4 shows the correlation of the extracted signature and the POF of the test image. (Of course, we are now illustrating the correlation with the standard convention of negative frequencies in order to put the ''action'' in the middle of the image.) If the test image is a marked image, we obtain a sharp correlation as shown in Fig. 4(a). For an unmarked image, the peaks appear random and there is no sharp correlation peak, as evident from Fig. 4(b).

We perform the simulation for the two sets of images and record the performance metrics as in Eqs. (10) and (11). Figure 5(a) shows the PACE values for marked and unmarked images for all ten 256×256 images. There is clear separation between them, and it is obvious that we can set a threshold that will authenticate a marked image from an unmarked one. Figure 5(b) shows the corresponding result for PSR values. The separation is even greater here, which implies a highly unlikely false-positive probability in the authentication test. Note that in calculating the second highest peak (needed to calculate PSR), we excluded a 3×3-pixel area centered on the highest peak.

Figures 6(a) and 6(b) show the corresponding metrics for set 2 images. Note that the separation between the correlation metrics of the marked and unmarked images, in general, is even more pronounced here. This characteristic difference is primarily coming from the larger correlation plane (512×512 as opposed to 256×256) in the case of Fig. 6. This also shows that our method yields better results with larger images.

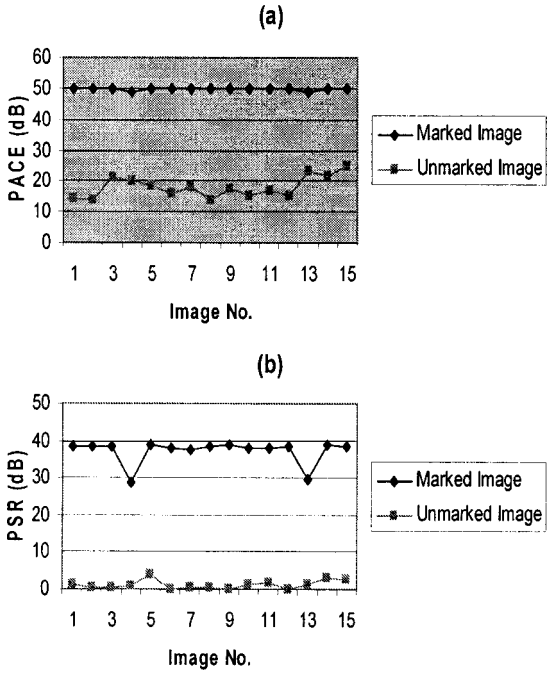

Fig. 6 Correlation performance for image set 2: (a) PACE, (b) PSR.

Let us now discuss the robustness of our semifragile watermarking method against some desirable image processing. Figures 7(a) and 7(b) show the difference images between the original and the watermarked image (accentuated for visibility) for ''Peppers.'' Note that this result is obtained from a very high embedding strength that is usually not used for typical applications. But it demonstrates one important aspect of the algorithm—that is, the embedding is context-sensitive. That means that removing our watermark will also degrade the quality of the original image seriously. This is in sharp contrast with the algorithm shown in Ref. 9, where the difference image is random.

With this result in mind, we now look into the robustness of our algorithm. Simulation results have shown that the algorithm has some built-in robustness against JPEG compression, Gaussian filtering, image enhancement, and cropping (to a certain degree). We furnish JPEG compression performance results because this is often a necessary form of image degradation that may affect the watermark.

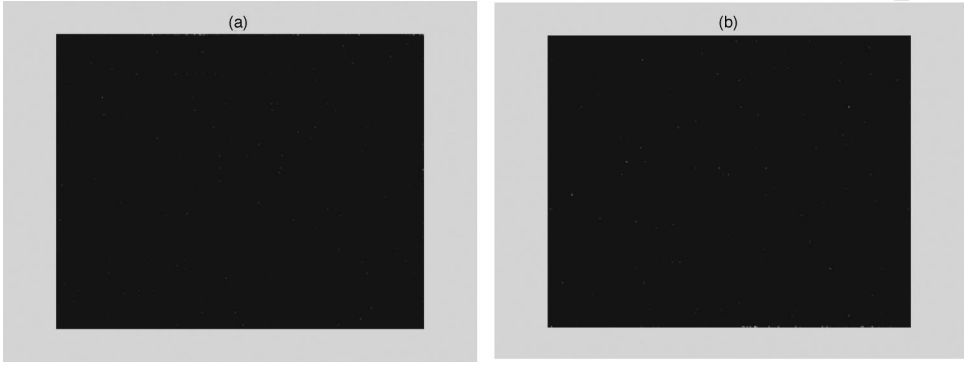

Fig. 7 Difference images with (a) 15th- and (b) 16th-bit-plane embedding.

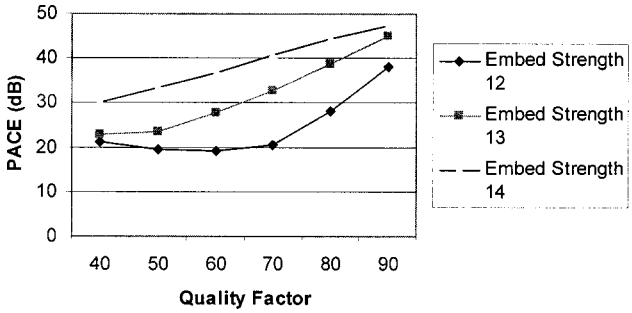

Fig. 8 Detector performance versus compression quality factor for "Peppers."

Figure 8 depicts the corresponding correlation output at different quality factors of "Peppers" and different embedding strengths. It is interesting to see that even at a quality factor of 40% (Matlab JPEG compression), the correlation metric value is 22.9. The corresponding value for an unmarked case is 17.5. Hence the marked image can still be authenticated against the unmarked one. We can even tolerate a little more compression, if the embedding plane is adjusted to the 14th bit plane as shown. For a smaller strength (12 bit planes, for example), performance degrades slowly.

5 Conclusion

We have proposed a BPOF signature-based watermarking technique for image authentication applications. Our method is extremely useful for image distribution scenarios over the internet, as there is no need to transmit complex images, thus providing a significant saving in bandwidth. Our correlation-based detector demonstrates a robust authentication of a watermarked image as distinguished from an unmarked one. Our method is also found to be particularly attractive for applications where it is desirable to tolerate image compression. As shown, the addition of the watermark is context-sensitive, to make it harder to remove.

Further work can be done on variants of the binary-phase-only signature with respect to quantization error and correlation. Multiple bit planes embedding with additional cryptographic components can also be pursued to address other assurance manifestations and robustness issues. One can also include additional (piggybacked) information in the watermarking scheme. These and other topics will be pursued in future work.

Acknowledgment

Research partially supported by the Office of Naval Research. We thank R. Heilizer and the anonymous reviewers for their helpful comments.

References

1. W. Stallings, *Cryptography and Network Security—Principles and Practices*, 3rd ed., Prentice Hall (2003).
2. H. C. Wong, M. Bern, and D. Goldberg, "An image signature for any kind of image," in *IEEE Int. Conf. on Image Processing*, pp. I-409–I-412 (2002).
3. I. Cox, J. Bloom, and M. Miller, *Digital Watermarking: Principles & Practice*, Chaps. 1–2. Morgan Kauffman Publishers (2001).
4. J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based digital signature for motion pictures authentication and content-fragile watermarking," in *IEEE Int. Conf. on Multimedia Computing and Systems*, Vol. 2, pp. 209–213 (1999).
5. J. Fridrich and M. Goljan, "Images with self-correcting capabilities," in *Proc. Int. Conf. on Image Processing*, Vol. 3, pp. 792–796 (1999).
6. C. Rey and J.-L. Dugelay, "Blind detection of malicious alterations on still images using robust watermarks," presented at *IEE Seminar on Secure Images and Image Authentication*, Apr. 2000.
7. H. Zhong, F. Liu, and L.-C. Jia, "A new fragile watermarking technique for image authentication," in *6th Int. Conf. on Signal Processing*, Vol. 1, pp. 792–795 (2002).
8. J. Fridrich, "A hybrid watermark for tamper detection in digital images," in *Proc. Fifth Int. Symp. on Signal Processing and Its Applications*, Vol. 1, pp. 301–304 (1999).
9. S. Kishk and B. Javidi, "Information hiding technique with double phase encoding," *Appl. Opt.* **41**(26), 5462–5470 (2002).
10. B. Javidi and E. Ahouzi, "Optical security system with Fourier plane encoding," *Appl. Opt.* **37**(26), 6247–6255 (1998).
11. D. Psaltis, E. Paek, and S. Venkatesh, "Optical image correlation with binary spatial light modulator," *Opt. Eng.* ■, 698–704 (1984).
12. J. L. Horner and J. R. Leger, "Pattern recognition with binary phase-only filters," *Appl. Opt.* **24**, 609–611 (1985).
13. F. Ahmed and M. A. Karim, "A filter-feature-based rotation invariant joint Fourier transform correlator," *Appl. Opt.* **34**(32), 7556–7560 (1995).
14. USC-SIPI Image Database, http://sipi.usc.edu/services/database/Database.html.
15. B. V. K. Vijay Kumar and L. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.* **29**(20), 2997–3006 (1990).

Farid Ahmed received his PhD in electrical engineering in 1996 from the University Of Dayton, OH. Currently he is an assistant professor of electrical engineering and computer science at the Catholic University of America (CUA), Washington, DC. Prior to this position he worked as a research and development engineer at Digimarc Corporation. His research interests include image and signal processing, watermarking, multimedia authentication, information security, and pattern recognition.

Ira S. Moskowitz is a mathematician at the U.S. Naval Research Laboratory. He received his PhD in mathematics in 1983 from the State University of New York at Stony Brook (SUNY). Prior to his arrival at NRL he was at Elmhurst College, Center for Naval Analysis, and Texas A&M University. His research interests include steganography, watermarking, covert channels, database inference, and anonymous communication.