Denoising and Copy Attacks Resilient Watermarking by Exploiting Prior Knowledge at Detector

Chun-Shien Lu, *Member, IEEE*, Hong-Yuan Mark Liao, *Senior Member, IEEE*, and Martin Kutter

***Abstract*—Watermarking with both oblivious detection and high robustness capabilities is still a challenging problem. In this paper, we tackle the aforementioned problem. One easy way to achieve blind detection is to use denoising for filtering out the hidden watermark, which can be utilized to create either false positive (copy attack) or false negative (denoising and remodulation attack). Our basic design methodology is to exploit prior knowledge available at the detector side and then use it to design a "nonblind" embedder. We prove that the proposed scheme can resist two famous watermark estimation-based attacks, which have successfully cracked many existing watermarking schemes. False negative and false positive analyses are conducted to verify the performance of our scheme. The experimental results show that the new method is indeed powerful.**

***Index Terms*—Attacks, denoising, oblivious detection, robustness, watermark prediction, watermarking.**

I. Introduction

WATERMARKING [8], [22], [25] is a technique which conceals one or more watermarks in a medium. Embedded watermarks can be used to declare rightful ownership (robust watermarking), to authenticate credibility (fragile watermarking) or to carry useful information (captioning watermarking). Current watermarking applications have led to either single purpose watermarking [4] or multipurpose watermarking [20]. Usually, a watermark itself can be a random signal, a meaningful message, or a company's logo. An effective watermarking scheme should satisfy a set of typical requirements, including transparency, robustness, oblivious (blind) detection, and so on. The main purpose of robust watermarking is to prevent hidden watermark(s) from being removed or destroyed so that ownership can be guaranteed. Watermarks can be detected with the help of the host media (called nonoblivious detection) or without access to the original media (called oblivious detection). Oblivious detection is practical but is still a challenge if high robustness is the major concern. Since the original source cannot be used in oblivious detection, the embedded watermark should be predicted from an attacked media. Under these circumstances, the predicted watermark values more or less deviate from their original ones. In other words, the degree of robustness will be affected. Therefore, robustness and oblivious detection are, in effect, in conflict with each other. However, if one can find a good watermark prediction scheme and then use it as part of the design methodology, then the degree of robustness degradation can be minimized. In this paper, we aim to tackle the aforementioned problem by using image watermarking as our domain.

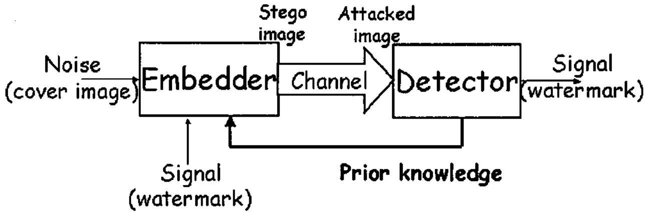


Fig. 1. Block diagram of the proposed watermarking scheme by using prior knowledge from detector.

Watermarking with oblivious (blind) detection [1], [11], [12], [26] has been extensively explored in recent years. Most of the existing methods detect watermarks by means of prediction, and this kind of strategy usually is not directly related to its hiding strategy. Therefore, robustness cannot be guaranteed. In [27], Voloshynovskiy *et al.* proposed a stochastic model to seriously address the watermark prediction problem. Since an oblivious approach usually detects watermarks by means of prediction, it is also possible that a pirate may successfully remove an embedded watermark by means of prediction. Voloshynovskiy et al. [28] called this kind of attack a "denoising and remodulation attack." In some situations, a predicted watermark may be maliciously added to another cover image that belongs to other people. This kind of attack aims to create the false positive problem. Kutter *et al.* [14] called this kind of attack a "copy attack." Since the aforementioned two attacks are very difficult to resist, any watermarking approach that claims to be *robust* may be cracked when either of the two attacks is encountered. Since a predicted watermark (for oblivious detection) may sacrifice robustness to some extent, we propose to design a watermarking system by taking both the embedding strategy and the detection strategy together into consideration. In other words, the characteristics of a predetermined detection model can be used as part of the criteria for designing a better watermarking system. In [4], Cox *et al.* proposed a new concept which views watermarking as communications with side information. This concept makes it possible to design a new watermarking method with better efficiency. In [21], Miller *et al.* adopted a similar concept [4] to design four different informed embedding strategies.

In this paper, we present a novel watermarking scheme which exploits the available information at the watermark detection (prediction) side. Fig. 1 depicts the concept of the proposed

Manuscript received March 27, 2001; revised November 26, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Naohisa Ohta.

C.-S. Lu and H.-Y. M. Liao are with the Institute of Information Science, Academia Sinica, Taipei 115, Taiwan, R.O.C (e-mail: lcs@iis.sinica.edu.tw; liao@iis.sinica.edu.tw).

M. Kutter is with the AlpVision SwissMedia Center, CH-1800 Vevey, Switzerland (e-mail: martin.kutter@alpvision.com).

Publisher Item Identifier S 1057-7149(02)01731-1.

methodology. Based on the information obtained from the detector side, we are able to use these prior information as part of the criteria for designing a robust embedder. We shall take the shrinkage-based denoising model as our watermark prediction module because it naturally leads to blind detection. Since the shrinkage-based denoising approach [5], [9], [10] adopts a soft-thresholding strategy to "gradually" decrease the magnitude of selected coefficients, it is easy to control the behaviors of denoising. Since the knowledge at the detector side is used to design an embedder, we call it a "nonblind" embedder. In sum, the proposed system is composed of a nonblind embedder and a blind detector. We shall analyze the performance of our scheme when watermark estimation-based attacks [14], [28] are encountered.

The remainder of this paper is organized as follows. In Section II, oblivious watermark detection formulated as a denoising problem is described. In Section III, the proposed scheme is described, and some performance analyses are discussed. Finally, experimental results are given in Section IV and concluding remarks are made in Section V.

II. Formulating Oblivious Detection as Watermark Prediction by Means of Shrinkage-Based Denoising

In this paper, oblivious watermark detection is formulated as a watermark prediction problem. Under the assumption that a watermark hiding/attacking process is modeled as a noise adding process, we can separate an embedded watermark from an attacked image by using the shrinkage-based denoising technique. Under the circumstances, the separated noise can be regarded as an extracted watermark, which more or less deviates from its original shape due to the execution of denoising and the effects of attacks. Based on the fact that the shrinkage operation tends to *gradually* decrease the magnitude of transformed coefficients, we propose to use shrinkage-based denoising to predict this noise (watermark). In the following, we will use the sparse code shrinkage (SCS) [9], [10] strategy to model the watermark prediction process since it is a generalization of shrinkage-based image denoising methods. In particular, some denoising algorithms (such as shrinkage-based denoising) are conducted in the transformed (wavelet or DCT) domain as most watermarking methods have done. In this section, we will investigate the relationship between watermarking and denoising in the wavelet domain. In Sections II-A and B, we shall describe in detail how to model the aforementioned processes by means of Gaussian modeling. Next, we will discuss how to use the SCS strategy to solve the denoising problem in Section II-C.

A. Gaussian Modeling of Coefficient Magnitude Update in the Hiding Process

In this paper, we only consider watermarking in the wavelet domain. Let $\mathbf{X}$ be an image in the spatial domain and let ψ be a wavelet function. Hence, the wavelet transformed image would be $\mathbf{X}^{\psi}(=\psi * \mathbf{X})$ in the space-frequency domain, where $*$ is a convolution operation. For watermarking, let $\mathbf{N}^{\psi}$ be either a multiplicative or an additive watermark (generated by a secret key) to be hidden in the wavelet domain. To facilitate analysis, $\mathbf{N}^{\psi}$ could be rewritten as $\psi * \mathbf{N}$, which implies that $\mathbf{N}^{\psi}$ embedded in the wavelet domain corresponds to $\mathbf{N}$ embedded in the spatial domain. In the watermark embedding process, the watermark $\mathbf{N}^{\psi}$ is added with the wavelet transformed image X^{ψ} to form a watermarked image $\tilde{\mathbf{X}}^{\psi}$ in the wavelet domain, which is expressed as

$$\tilde{\mathbf{X}}^{\psi} = \mathbf{X}^{\psi} + \mathbf{N}^{\psi} \tag{1}$$

where "+" is the addition operation commonly used for watermarking embedding (quantization operation is an exception). From (1), it can be derived as follows:

$$\begin{aligned}\tilde{\mathbf{X}}^{\psi} &= \mathbf{X}^{\psi} + \mathbf{N}^{\psi} = (\psi * \mathbf{X}) + (\psi * \mathbf{N}) \\ &= \psi * (\mathbf{X} + \mathbf{N}) \\ &= \psi * \tilde{\mathbf{X}}\end{aligned} \tag{2}$$

where $\tilde{\mathbf{X}}$ indicates a watermarked image in the spatial domain. The above derived result indicates that the wavelet transform (using ψ) of the watermarked image $\tilde{\mathbf{X}}$ is equivalent to the modulation of the wavelet transformed image $\mathbf{X}^{\psi}$ by adding with $\mathbf{N}^{\psi}$.

Now, suppose a watermark has been embedded into a host image in the wavelet domain. This means that the original image $\mathbf{X}$ is first wavelet transformed using ψ and then modulated using $\mathbf{N}^{\psi}$. Under these circumstances, the modulated wavelet coefficients can be modeled as the original wavelet coefficients plus Gaussian noise added in the wavelet domain. That is

$$w_{s,o}^{m}(x,y) = w_{s,o}(x,y) + n(i) \tag{3}$$

where $w_{s,o}(x,y)$ is the original wavelet coefficient, $w_{s,o}^{m}(x,y)$ is the modulated wavelet coefficient (s and o denote scale and orientation, respectively, in the wavelet domain) and $n(i)$ is the ith element of the hidden Gaussian noise-like watermark $\mathbf{N}^{\psi}$. The relationship between (x,y) and i will be defined in Section III. By means of (2), we know that the Gaussian modeling is also similarly defined for $\tilde{\mathbf{X}}$ in the spatial domain. Let $x(j)$ be the pixel intensity of an original image $\mathbf{X}$ at a position j (j denotes an index of a pixel in $\mathbf{X}$), and, let $N(j)$ be the noise value (which results from the hidden Gaussian noise-like watermark, $\mathbf{N}$). The intensity of a noisy pixel, $x^{m}(j)$, can be calculated by $x^{m}(j) = x(j) + N(j)$. Therefore, the watermarked image in the spatial domain can be modeled as

$$\mathbf{X}^{m} = \mathbf{X} + \mathbf{N} \tag{4}$$

where $\mathbf{N}$ is the noise sequence and $\mathbf{X}^{m}$ is the same as $\tilde{\mathbf{X}}$ in (2). In Section II-C, we will see how shrinkage-based denoising is conducted on $\tilde{\mathbf{X}}$.

B. Gaussian Modeling of Coefficient Magnitude Update in the Attacking Process

After watermark hiding, the watermarked image can be transmitted over the Internet and may be attacked by any process. At this time, the model of an attack is assumed to be the same as that of a modulation operation except that

1) original image $\mathbf{X}$ in (2) and (4) is replaced by the watermarked image $\mathbf{X}^{m}$;
2) $\mathbf{N}$ in (4) is resultant noise which is contributed by a hidden watermark and attacks;

3) $\tilde{\mathbf{X}}$ in (2) and $\mathbf{X}^m$ in (4) are attacked images instead of watermarked images.

To simplify the analysis, we assume that $\mathbf{N}$ is still a Gaussian distribution with variance σ. The value of σ will be small/large when the imposed attack is weak/strong.

C. SCS Technique

After conducting Gaussian modeling of coefficient magnitude change with respect to an attacked image, the next step is to separate the host image $\mathbf{X}$ from the attacked image $\tilde{\mathbf{X}}$ by denoising $\mathbf{N}$. Using the denoising operation, the estimated host image $\bar{\mathbf{X}}$ can approximate the original image, i.e., $\bar{\mathbf{X}} \approx \mathbf{X}$. In order to achieve the aforementioned goal, the independent component analysis (ICA)-based sparse code shrinkage (SCS) technique [9] is employed to model the denoising problem. An SCS-based denoising algorithm includes the following steps:

1) model the noisy image $\tilde{\mathbf{X}}$ as a set of independent components;
2) perform sparse code shrinkage on these components;
3) invert the ICA representation.

The step-by-step procedure for an ICA-based SCS denoising algorithm is given in the following. First, one has to model the host image $\mathbf{X}$ using the independent component analysis process [3]. This process decides on the major components of the host image. On the other hand, we need to consider the noise part ($\mathbf{N}$) consisting of minor components, which can be shrunk (soft-thresholded) using an adaptive soft threshold during the ICA process. In an explicit format, the host image can be modeled as $\mathbf{X} = \mathbf{A}\mathbf{s}$, where $\mathbf{A}$ is a basis matrix and $\mathbf{s}$ is the vector of independent components (ICs). Analogous to traditional transformations, such as discrete Fourier transform or wavelet transform, $\mathbf{s}$ is composed of a set of selected transformed coefficients, and $\mathbf{A}$ is a synthesis filter. Therefore, ICA has the property that different ICs are unlikely to be activated at the same time due to its sparse distributed nature (i.e., energy compaction). More specifically, an independent component will be activated if it is sufficiently large (In watermarking, watermark will be concealed into those activated ICs to maintain transparency and achieve robustness). Therefore, the noisy image $\tilde{\mathbf{X}}$ can be denoted as

$$\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{N} = \mathbf{A}\mathbf{s} + \mathbf{N}. \tag{5}$$

Suppose only the observed data $\tilde{\mathbf{X}}$ is given; the basis matrix $(\mathbf{A})$ and the ICs $(\mathbf{s})$ can be obtained by first finding a separating matrix $\mathbf{W}$ (with $\mathbf{W}^{-1} = \mathbf{A}$) via sparse coding [9]. Then, $\mathbf{s}$ can be determined by $\mathbf{s} = \mathbf{W}\mathbf{X}$, where each component $\mathbf{s}_i = \mathbf{W}_i\mathbf{X}$. After sparse coding, the noisy image $\tilde{\mathbf{X}}$ can be transformed by means of $\mathbf{W}$, and a noisy independent component, $\mathbf{s}+\tilde{\mathbf{N}}$ (in the ICA transformed domain), can finally be derived as follows:

$$\mathbf{W}\tilde{\mathbf{X}} = \mathbf{W}\mathbf{X} + \mathbf{W}\mathbf{N} = \mathbf{W}\mathbf{A}\mathbf{s} + \mathbf{W}\mathbf{N} = \mathbf{s} + \tilde{\mathbf{N}}. \tag{6}$$

In the second step, each noisy component, $\mathbf{s}_i + \tilde{\mathbf{N}}_i$, is shrunk by the denoising operation. When we use sparse code shrinkage to denoise $\mathbf{s} + \tilde{\mathbf{N}}$, we need to model the distribution of each component, $\mathbf{s}_i$, to see whether it satisfies the non-Gaussian requirement. One antecedent condition that image denoising by means of shrinkage can achieve is that each component $\mathbf{s}_i$ must be non-Gaussian so that it can be distinguished from normal Gaussian noise. Due to the energy compact representation of an ICA model, every independent component $\mathbf{s}_i$ is expected to exhibit sparse density. The second condition required for image denoising by shrinkage to function is that the variance of $\mathbf{N}$ must be assumed in advance [5]. After the sparse density of each $\mathbf{s}_i$ is modeled, their corresponding parameters can be generated to determine a suitable shrinkage function, $\mathbf{g}_i$ [9]. Then, one can shrink $\mathbf{s}_i + \tilde{\mathbf{N}}_i$ by means of $\mathbf{g}_i$ and then get the cleaned version of $\mathbf{s}$, which is represented as $\bar{\mathbf{s}}$, where

$$\bar{\mathbf{s}} = \mathbf{g}_i(\mathbf{s}_i + \tilde{\mathbf{N}}_i). \tag{7}$$

In general, the shrinkage function, $\mathbf{g}_i$, is explicitly defined [10] based on the sparse density distribution of noisy independent components to have the effects that small arguments are set to zero and the absolute value of large arguments are reduced by an amount depending on the noise level. In the third step, the approximated host image $\bar{\mathbf{X}}$ can be derived by an inverse ICA transformation: $\bar{\mathbf{X}} = \mathbf{A}\bar{\mathbf{s}}$. After the estimated host image is determined, it can be used for blind detection.

Wavelet shrinkage [5] is a good alternative to SCS-based denoising [9] due to its capability of fast computation. In wavelet shrinkage, $\mathbf{W}$ and $\mathbf{A}$ form a pair of wavelet analysis and synthesis filters. In addition, the shrinkage function used in wavelet shrinkage is fixed and is independent of the distribution of independent components. Although the denoising results obtained by applying wavelet shrinkage-based denoising are worse than those obtained by applying SCS-based denoising, their function in watermark prediction is almost the same (here, the performance of denoising is objectively derived by measuring the PSNR between the denoised image and its original version).

III. Proposed Denoising-Based Oblivious Watermarking Method

In this section, we will describe the proposed method and analyze its performance. In Section III-A, we shall describe in detail how a robust embedder can be designed by exploiting the knowledge of shrinkage-based watermark prediction. The processes of watermark embedding and watermark detection will be described as well. In Section III-B, performance analysis of the proposed scheme will be presented. In Section III-C, the relationship between our scheme and Cox *et al.*'s new watermarking concept [4] will be examined.

A. Proposed Approach: A Nonblind Embedder

In this section, we shall describe in detail the proposed watermarking system. For watermark embedding, let $k(i)$ be an element of a watermark $\mathbf{K}$ and let $k(i)$ be used to modulate a wavelet coefficient $w_{s,o}(x,y)$ as follows:

$$w^m_{s,o}(x,y) = w_{s,o}(x,y) + k(i). \tag{8}$$

After simple reorganization, we have

$$\text{sign}(k(i)) = \text{sign}\left(w^m_{s,o}(x,y) - w_{s,o}(x,y)\right) \tag{9}$$

where sign is an operator defined as

$$\mathrm{sign}(t) = \begin{cases} +1, & t \geq 0; \\ -1, & t < 0. \end{cases} \tag{10}$$

In order to maintain transparency, the sign of $w^m_{s,o}(x,y)$ has to be the same as that of $w_{s,o}(x,y)$. That is, $\mathrm{sign}(w^m_{s,o}(x,y)) = \mathrm{sign}(w_{s,o}(x,y))$.

For watermark extraction, the extracted watermark value $\tilde{k}(i)$ is obtained as $\tilde{k}(i) = w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y)$ by shrinkage-based denoising, where $\bar{w}_{s,o}(x,y)$ is the estimated original wavelet coefficient. Hence, we have $\mathrm{sign}(\tilde{k}(i)) = \mathrm{sign}(w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y))$. In addition, based on the characteristic of the shrinkage function [10] (or wavelet shrinkage function, (18)), we come out with the result that $\mathrm{sign}(w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y)) = \mathrm{sign}(w^a_{s,o}(x,y))$. By incorporating these results, the sign of an extracted watermark $\tilde{k}(i)$ can be derived by

$$\begin{aligned} \mathrm{sign}(\tilde{k}(i)) &= \mathrm{sign}\left(w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y)\right) \\ &= \mathrm{sign}\left(w^a_{s,o}(x,y)\right) \end{aligned} \tag{11}$$

which is, in essence, a denoising-based watermark detection process. As we have mentioned previously [17], the basic requirement for obtaining a higher correlation value between $k(i)$ and $\tilde{k}(i)$ is to get them to have the same sign, i.e., $\mathrm{sign}(k(i)) = \mathrm{sign}(\tilde{k}(i))$. By incorporating (9) and (11), $\mathrm{sign}(w^a_{s,o}(x,y)) = \mathrm{sign}(w^m_{s,o}(x,y) - w_{s,o}(x,y))$ holds. However, both $w^a_{s,o}(x,y)$ and $w^m_{s,o}(x,y) - w_{s,o}(x,y)$ can be either positive or negative, which makes the correlation between $k(i)$ and $\tilde{k}(i)$ hard to predict. This situation indicates that a watermarking scheme which adopts a typical spread-spectrum hiding strategy together with a shrinkage-based prediction rule cannot guarantee robustness.

From (11), we realize that the sign of an extracted watermark is dependent on the attacked wavelet coefficient due to the nature of shrinkage-based denoising. Therefore, if we can use the knowledge derived from the denoising-based prediction side, then we can design a suitable hiding strategy. In what follows, we shall discuss how to design a good hiding strategy. It is known that a pirate will not perceptually damage an image. Therefore, it is reasonable to assume that the signs of $w^m_{s,o}(x,y)$ and $w^a_{s,o}(x,y)$ are the same, i.e.,

$$\mathrm{sign}\left(w^m_{s,o}(x,y)\right) = \mathrm{sign}\left(w^a_{s,o}(x,y)\right). \tag{12}$$

By combining (9), (11), and (12), we can design the watermark embedding strategy so as to satisfy $\mathrm{sign}(w^m_{s,o}(x,y) - w_{s,o}(x,y)) = \mathrm{sign}(w^m_{s,o}(x,y))$. That is, the watermark should be embedded in order to increase the magnitudes of the chosen coefficients such that

$$\left|w^m_{s,o}(x,y)\right| > \left|w_{s,o}(x,y)\right| \tag{13}$$

holds. This derived result is exactly the same as the effect of positive modulation of cocktail watermarking [17]. Therefore, in this paper, only one watermark will be embedded in an image using positive modulation. The proposed watermarking method is described as follows.

In the watermark hiding process, suppose that $\mathbf{X}$ is a cover image and that an S-scale wavelet transform is performed on $\mathbf{X}$. Let the wavelet coefficient to be modulated be $w_{s,o}(x,y)$, where $1 \leq s \leq S$. In order to satisfy the compromise between transparency and robustness, the highest-frequency subbands will not be watermarked so that s must be larger than 1 (the finest scale). Besides, the lowest-frequency subband located at the S-scale is usually very small in size and is also nonwatermarked to preserve transparency. Fig. 2 illustrates the places (white areas) in which watermark values are embedded, where the gray areas indicate the high-frequency subbands and the black area indicates the lowest-frequency subband of a three-level wavelet decomposition. Therefore, it is not difficult to figure out that the length of a hidden watermark $(\mathbf{K})$ is about one-quarter of the original image size.

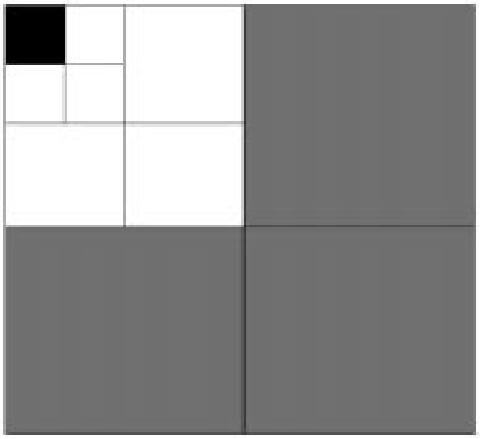


Fig. 2. Example of hiding places: A three-level wavelet decomposition with gray area indicating highest-frequency components, black area indicating lowest-frequency component, and white areas indicating the hiding places.

After wavelet coefficients have been selected to embed watermark values, an embedding rule should be designed to achieve the desired goal, i.e., (13). We have to carefully consider the relationship between the sign of a selected wavelet coefficient and that of its corresponding watermark value. Therefore, the hidden watermark sequence $\mathbf{K}$ is first sorted in an increasing order as $\tilde{\mathbf{K}}$. Let $\mathrm{top}(i)/\mathrm{bottom}(i)$ be used to denote that the watermark value $\tilde{k}(\mathrm{top}(i))/\tilde{k}(\mathrm{bottom}(i))$ is retrieved from the first/last i position of the sorted watermark sequence $\tilde{\mathbf{K}}$. The above sorted results will be recorded as

$$p(x,y) = j \tag{14}$$

where j, denoting $\mathrm{top}(i)$ or $\mathrm{bottom}(i)$, is the index of the sorted watermark sequence $\tilde{\mathbf{K}}$ and (x,y) is the location of a selected wavelet coefficient. The information of $p(x,y)$ will be required for watermark extraction. By this careful design, a watermark $\mathbf{K}$ could be hidden with each selected wavelet coefficient modulated as

$$w^m_{s,o}(x,y) = \begin{cases} w_{s,o}(x,y) + J_{s,o}(x,y) \times \tilde{k}(\mathrm{bottom}(i)) \times \alpha, \\ \quad \text{if } w_{s,o}(x,y) > J_{s,o}(x,y); \\ w_{s,o}(x,y) + J_{s,o}(x,y) \times \tilde{k}(\mathrm{top}(i)) \times \alpha, \\ \quad \text{if } w_{s,o}(x,y) < -J_{s,o}(x,y), \end{cases} \tag{15}$$

where $J_{s,o}(\cdot,\cdot)(>0)$ represents the JND values obtained from the visual model [31] and $\alpha (0 < \alpha \leq 1)$ is an image-dependent weight used to control the maximum possible modification that

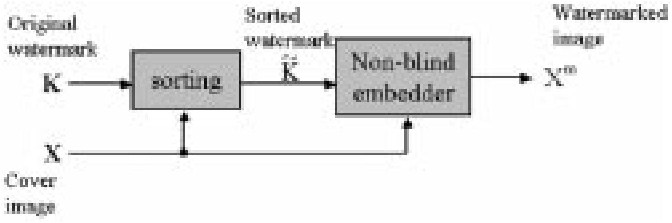


Fig. 3. Proposed watermark embedding process.

will lead to the least image quality degradation. Basically, α may be selected to satisfy perceptual transparency subjectively or to make the PNSR of a watermarked image larger than a certain value objectively. We shall see later in this section that α (no matter what value it is) will not affect the detection of watermarks. Basically, (15) implies that if a selected wavelet coefficient is positive/negative, then a positive/negative watermark value should be embedded in order to satisfy the desired goal [see (13)]. If a positive/negative watermark value is needed, one can retrieve it from the bottom/top of the sorted watermark sequence $\tilde{\mathbf{K}}$. Under these circumstances, $|w^m_{s,o}(x,y)| > |w_{s,o}(x,y)|$ is always guaranteed. Fig. 3 shows the diagram of the proposed watermark embedding process.

In the watermark detection process, an attacked image (with wavelet coefficient denoted as $w^a_{s,o}(x,y)$) is first denoised by means of a shrinkage-based denoising process. After this process is finished, the original image can be estimated (with wavelet coefficient denoted as $\bar{w}_{s,o}(x,y)$). Then, the estimated original image can be used to conduct blind watermark detection by retrieving the watermark elements $k^e(i)$ of $\mathbf{K}^e$, where

$$k^e(i) = \frac{w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y)}{J_{s,o}(x,y) \times \alpha}. \tag{16}$$

Finally, the normalized correlation value is calculated to measure the similarity between $\mathbf{K}$ and $\mathbf{K}^e$ by means of

$$\rho(\mathbf{K}, \mathbf{K}^e) = \frac{\sum_{i=1}^{||\mathbf{K}||} \operatorname{sign}(k(i)) \times \operatorname{sign}(k^e(i))}{||\mathbf{K}||} \tag{17}$$

where $||\mathbf{K}||$ denotes the length of the watermark. In (17), it can be easily checked from $\operatorname{sign}(\cdot)$ function and normalized correlation that α, being image-dependent, does not affect the watermark detection.

It is clear that the time bottleneck in the proposed system is in the sparse code calculation. Since efficiency is a major concern in watermark detection, we shall use wavelet transform to perform the shrinkage-based denoising task [5]. The wavelet shrinkage function, first proposed by Donoho et al. [5] and commonly used in denoising, is defined as

$$g(x) = \operatorname{sign}(x) \mathrm{MAX}(0, |x| - t) \tag{18}$$

where MAX is a maximum function and t is the noise level. The noise level is defined as $(\sqrt{2\log(n)}\cdot\sigma/\sqrt{n})$ [5], where the noise variance σ is usually unknown and has to be estimated, and n here is the number of samples. Based on [5], σ is approximately estimated as the median absolute deviation of the wavelet coefficients at the smallest scale divided by 0.6745. On the other hand, owing to a secret key is required to generate a hidden watermark and this hidden watermark must be sorted in the embedding process, a sorted watermark instead of a single secret key needs to be provided in the watermark detection process. This implies that our secret key (in fact, a secret sequence) is longer than those in conventional methods and each image has to be enforced to associate with a secret sequence. Fig. 4 shows the diagram of the proposed watermark extraction process.

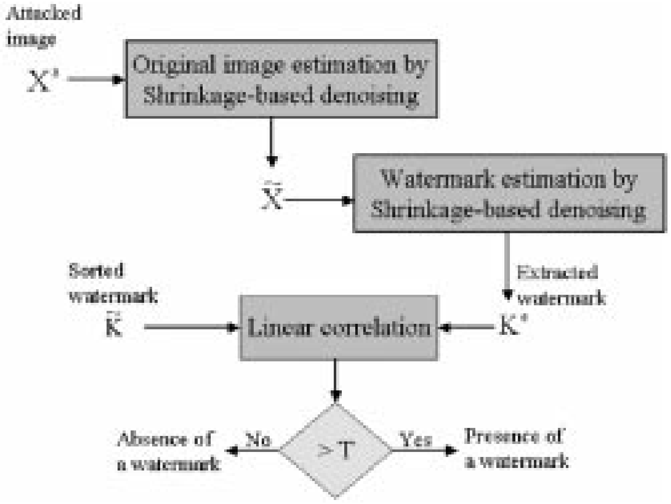


Fig. 4. Proposed watermark extraction process.

B. Performance Analysis

Some issues regarding performance evaluation of the proposed method are discussed in the following.

1) False Negative and False Positive Analysis: In our scheme, a threshold T is used to indicate the presence/absence of a watermark if a correlation value is larger/smaller than T. The error probabilities, composed of a false negative (miss detection) and a false positive (false alarm), will be used to evaluate our system. In our analysis, the distributions of the detection results with respect to attacked images (including watermarked images) and nonwatermarked images are, respectively, approximated using Gaussian probability density functions (PDFs). In fact, the detection results of attacked images are represented using a normal Gaussian distribution while those of nonwatermarked images are represented using a generalized Gaussian. The statistics of the aforementioned distributions can be estimated by means of experiments. Suppose the mean and the variance of the distribution of nonwatermarked images and those of the attacked images are μ_n, σ_n^2 and μ_a, σ_a^2, respectively, with $\mu_n < \mu_a$. The intersection area of the two distributions is defined as the error probability, and the intersection point of the above two distributions is defined as the threshold $T(-1 \leq T \leq 1)$. Then, the false negative probability can be derived as follows:

$$\begin{aligned} p_{\mathrm{fn}} &= \frac{\int_{-1}^{T} e^{-\frac{(x-\mu_a)^2}{2\sigma_a^2}}\, dt}{\int_{-1}^{1} e^{-\frac{(x-\mu_a)^2}{2\sigma_a^2}}\, dt} \\ &= \frac{\operatorname{erf}\left(\frac{(1+\mu_a)}{\sqrt{2}\sigma_a}\right) + \operatorname{erf}\left(\frac{\mu_a - T}{\sqrt{2}\sigma_a}\right)}{\operatorname{erf}\left(\frac{(1+\mu_a)}{\sqrt{2}\sigma_a}\right) + \operatorname{erf}\left(\frac{1-\mu_a}{\sqrt{2}\sigma_a}\right)}. \end{aligned} \tag{19}$$

Similarly, the false positive probability can be derived as

$$p_{\text{fp}} = \frac{\int_{T}^{1} e^{-\frac{(x-\mu_n)^2}{2\sigma_n^2}}\, dt}{\int_{-1}^{1} e^{-\frac{(x-\mu_n)^2}{2\sigma_n^2}}\, dt} = \frac{\operatorname{erf}\left(\frac{(1-\mu_n)}{\sqrt{2}\sigma_n}\right) - \operatorname{erf}\left(\frac{T-\mu_n}{\sqrt{2}\sigma_n}\right)}{\operatorname{erf}\left(\frac{(1+\mu_n)}{\sqrt{2}\sigma_n}\right) + \operatorname{erf}\left(\frac{1-\mu_n}{\sqrt{2}\sigma_n}\right)}. \tag{20}$$

False negative and false positive numerical results for different threshold values were obtained in our experiments.

2) Analysis of Denoising-Based Prediction With Different Noise Variance: For sparse code shrinkage [9] or wavelet shrinkage [5], the variance of a noise distribution, σ (relevant to the denoising capability), should be determined in advance in order to separate the original image, $\mathbf{X}$, from its embedded noise, $\mathbf{N}$. It should be noted that the value of σ is hard to predict but definitely affects the final reconstruction result. Fortunately, the major concern here is not the original image. What we are concerned about is the detected correlation values. Therefore, it is sufficient if the watermark extracted from the estimated host image is highly correlated with the hidden watermark.

In the following, we shall evaluate the performance of the proposed system when noises with different variance (σ) values are used. Here, we just summarize the final result that different σ values will not affect the correlation value significantly because the polarity of an extracted watermark value can always be kept the same as that of its original one. Please see Appendix A for more detailed analysis.

3) Resistance to Denoising Attacks: From the above analysis, we know that the predicted watermark is indeed very similar to the hidden one. Recently, Voloshynovskiy et al. [28] have presented a "denoising and perceptual remodulation attack" which is created by first predicting the hidden watermark using some denoising techniques and then removing the predicted watermark from a watermarked image by means of perceptual remodulation. Kutter *et al.* [14] also used denoising techniques to estimate a watermark. In contrast to Voloshynovskiy *et al.*'s work [28], Kutter *et al.* [14] added the estimated watermark into a nonwatermarked image to create a false alarm situation. This kind of attack is a so-called "copy attack" and can be used to challenge the concept of watermarking. From the above two works, we know that a watermark can be predicted by means of denoising and then used to create either a miss detection [28] or false alarm [14] situation. One may ask: "Does *successful* prediction of a watermark also imply that watermark removal can be done successfully?" Our answer is *NO*. We will explain why such an attack cannot successfully destroy a watermark embedded using our method.

a) Resistance to the denoising and perceptual remodulation attack: First, we will examine "denoising and remodulation attacks" [28]. Let $\mathbf{X}^a$ be an attacked image which is obtained by applying a denoising operation to a watermarked image $\mathbf{X}^m$. Suppose the denoising operation is a technique such as low-pass filtering, median filtering, Wiener filtering, or shrinkage-based denoising [5], [9]. After applying the denoising operation, we will have either $|w^a_{s,o}(x,y)| < |w^m_{s,o}(x,y)|$ or $|w^a_{s,o}(x,y)| \geq |w^m_{s,o}(x,y)|$. In fact, most coefficients will be *gradually* reduced in magnitude during denoising except when some nonshrinkage-based denoising techniques (like low-pass filtering) are used. Therefore, $|w^a_{s,o}(x,y)| < |w^m_{s,o}(x,y)|$ holds for most coefficients. In our scheme, a hidden watermark is detected in an attacked image $\mathbf{X}^a$ by means of a shrinkage-based denoising operation. Therefore, the coefficients of the estimated original image $\bar{\mathbf{X}}$ and those of the attacked image $\mathbf{X}^a$ should satisfy the following inequality:

$$|\bar{w}_{s,o}(x,y)| < \left|w^a_{s,o}(x,y)\right|. \tag{21}$$

From this analysis, we have

$$|\bar{w}_{s,o}(x,y)| < \left|w^a_{s,o}(x,y)\right| < \left|w^m_{s,o}(x,y)\right|. \tag{22}$$

In the proposed scheme, positive modulation is applied to the original image. Therefore, we can obtain that $\text{sign}(w^m_{s,o}(x,y) - w_{s,o}(x,y))$ is equal to $\text{sign}(w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y))$. This means that the overall correlation value will increase. From the above analysis, we conclude that a watermark embedded by our scheme is hard to remove using a shrinkage-based denoising algorithm.

b) Resistance to the copy attack: Next, we will examine the effect caused by the "copy attack" [14] on our scheme. Let $w^1_{s,o}(x,y)$ be the wavelet coefficient of an image $\mathbf{X}^1$ belonging to us, and let $w^2_{s,o}(x,y)$ be the wavelet coefficient of an image $\mathbf{X}^2$ belonging to someone else. Let the modulated, attacked, and denoised versions of $w^1_{s,o}(x,y)$ be denoted as $w^{1m}_{s,o}(x,y), w^{1a}_{s,o}(x,y)$, and $\bar{w}^1_{s,o}(x,y)$, respectively. Furthermore, let the hidden watermark be denoted as $\mathbf{n}^1$. Suppose a denoising technique such as Wiener filtering [15] or sparse code shrinkage [9] is applied to $\mathbf{X}^1$; the predicted watermark $\tilde{\mathbf{n}}^1$ will have the following value:

$$\tilde{k}^1(i) = w^{1m}_{s,o}(x,y) - \bar{w}^1_{s,o}(x,y) \tag{23}$$

where $1 \leq i \leq ||\mathbf{K}||$. The predicted watermark value $\tilde{k}^1(i)$ is then added to the nonwatermarked image $\mathbf{X}^2$ as

$$w^{2a}_{s,o}(x,y) = w^2_{s,o}(x,y) + \tilde{k}^1(i) \tag{24}$$

to create a counterfeit image $\mathbf{X}^{2a}$ with the wavelet coefficients $w^{2a}_{s,o}(x,y)$. Under these circumstances, we can check to see if a watermark retrieved from the counterfeit image is similar to the hidden one, i.e., $\mathbf{n}^1$. Using the proposed method, the watermark-free counterfeit image can be estimated by $\bar{w}^2_{s,o}(x,y)$, where $|\bar{w}^2_{s,o}(x,y)| < |w^{2a}_{s,o}(x,y)|$. As a consequence, the value of the predicted watermark $\tilde{\mathbf{n}}^2$ which can be calculated from $\mathbf{X}^{2a}$ is

$$\begin{aligned}\tilde{k}^2(i) &= w^{2a}_{s,o}(x,y) - \bar{w}^2_{s,o}(x,y)\\ &= w^2_{s,o}(x,y) + \tilde{k}^1(i) - \bar{w}^2_{s,o}(x,y).\end{aligned} \tag{25}$$

Due to the gradual change caused by shrinkage-based denoising, we can guarantee that

$$\text{sign}(\tilde{k}^2(i)) = \text{sign}\left(w^{2a}_{s,o}(x,y)\right).$$

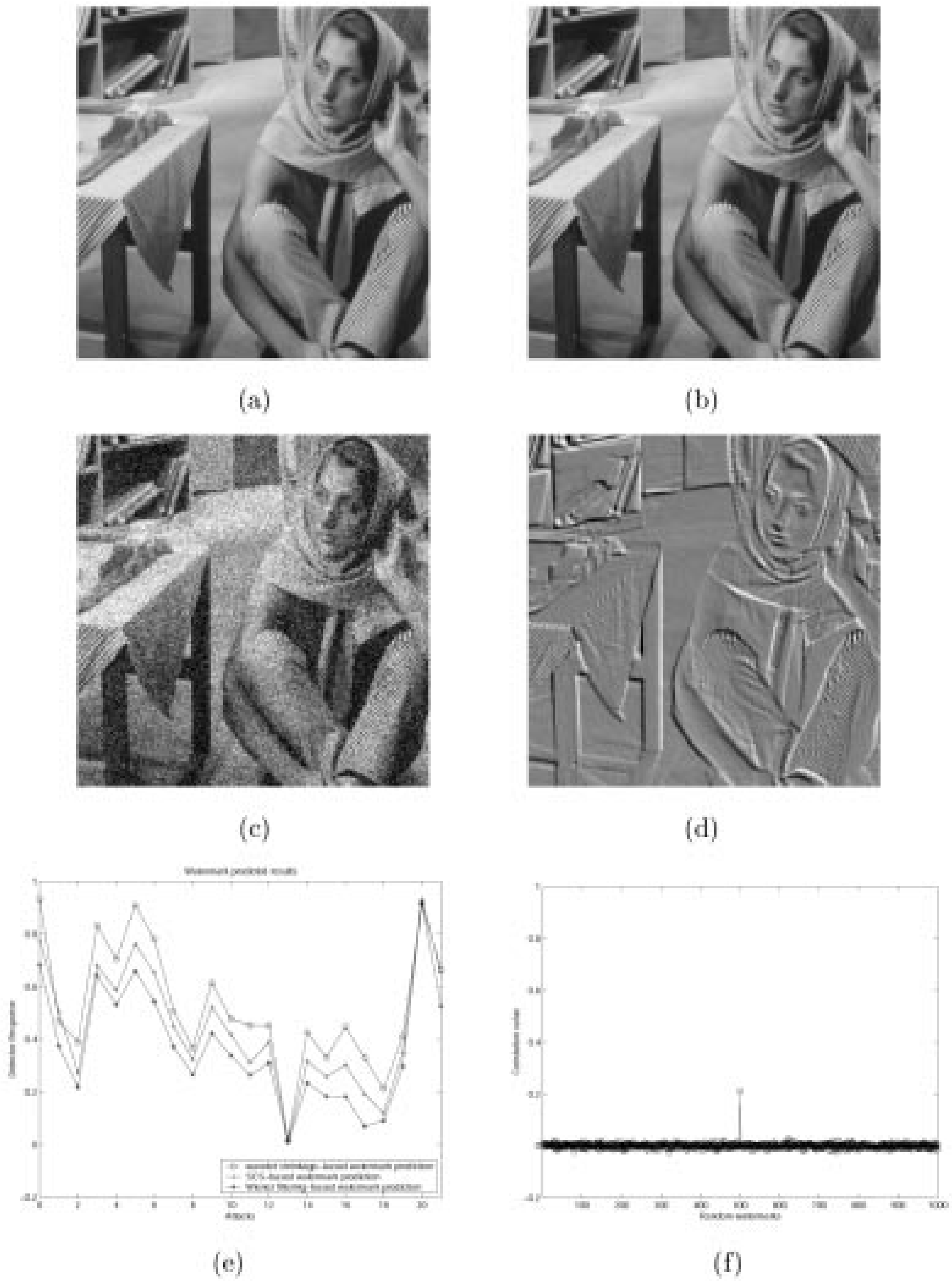


Fig. 5. Robustness test of the proposed scheme (a nonblinder embedder and a blind detector). (a) Host image; (b) watermarked image; (c) Gaussian noise added image; (d) attacked image with the shading effect; (e) comparison of detected watermarks, respectively, predicted using wavelet shrinkage, SCS, and Wiener filtering. The first response was obtained without applying any attack and the remaining results were obtained by applying the 21 attacks described previously (zeroeth attack denotes attack-free); and (f) unique watermark test for the StirMark attack.

Because $\tilde{k}^1(i)$ cannot significantly affect $w^2_{s,o}(x,y)$ from the viewpoint of transparency, we are assured that

$$\mathrm{sign}(\tilde{k}^2(i)) = \mathrm{sign}\left(w^2_{s,o}(x,y)\right). \qquad (26)$$

On the other hand, since the hidden watermark is designed to have the same sign as its corresponding wavelet coefficient $w^1_{s,o}(x,y)$, we have

$$\mathrm{sign}(k^1(i)) = \mathrm{sign}\left(w^1_{s,o}(x,y)\right). \qquad (27)$$

From (26) and (27), we find, in summary, that $\mathrm{sign}(k^1(i) \times \tilde{k}^2(i)) = \mathrm{sign}(w^1_{s,o}(x,y) \times w^2_{s,o}(x,y))$. The above conclusion indicates that the correlation value between $k^1(i)$ and $\tilde{k}^2(i)$ is directly related by the signs of their corresponding wavelet coefficients. Because the property of a nonwatermarked image is random in nature, it can be expected that the correlation value between the retrieved watermark $\tilde{\mathbf{n}}^2$ and the hidden one $\mathbf{n}^1$ will be close to zero. This means that the proposed denoising-based oblivious watermarking method (positive modulation incorpo-

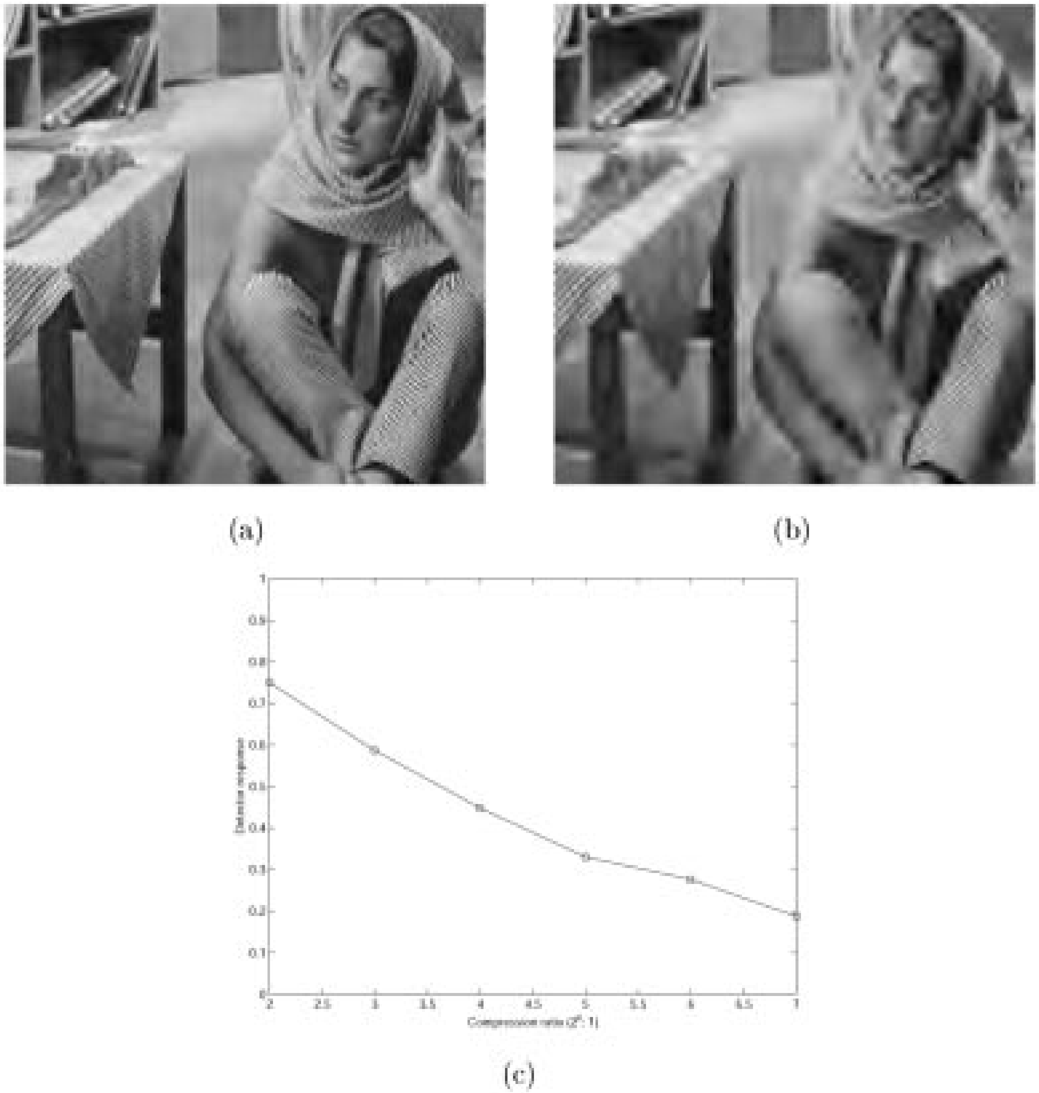


Fig. 6. Proposed scheme under the *SPIHT* compression attack: (a) the *SPIHT* compressed image with a ratio of 16 : 1; (b) *SPIHT* compressed image with a ratio of 64 : 1; and (c) detection results obtained under *SPIHT* compression at different ratios ranging from $2^2 : 1 \sim 2^7 : 1$.

rated with shrinkage-based watermark prediction) is able to resist a "copy attack" [14].

C. Relationship With the Concept of Watermarking as Communications With Side Information

In [4], Cox *et al.* proposed a new concept which views watermarking as communications with side information. In their scheme, the embedded signal $\mathbf{S}$, which is composed of an extracted signal $\mathbf{V}$ and a watermark $\mathbf{K}$, is perceptually similar to the extracted signal to achieve fidelity and is highly correlated with the hidden watermark $\mathbf{K}$ to achieve robustness. In general, $\mathbf{S}$ can be obtained as a combination of $\mathbf{V}$ and $\mathbf{K}$ by a mixing function f, i.e.,

$$\mathbf{S} = f(\mathbf{V}, \mathbf{K}). \tag{28}$$

A suboptimal way of computing $\mathbf{S}$ is defined as

$$\mathbf{S} = \mathbf{V} + \omega \cdot \mathbf{K} \tag{29}$$

where ω is a weight. Recently, four different embedding strategies (including the above one) have been proposed as "informed embedders" [21]. Their performance was compared with that of blind embedding and it was found that informed embedding is better. If our watermarking scheme is interpreted by Cox *et al.*'s concept [4], then we can derive the following result:

$$\mathrm{sign}(\mathbf{S}) = \mathrm{sign}(\mathbf{V}) = \mathrm{sign}(\mathbf{K}). \tag{30}$$

This is because our scheme attempts to keep the signs of watermark values unchanged. In this paper, robustness can be guaranteed if the signs of watermark values remain unchanged after attacks, i.e., $\mathrm{sign}(\mathbf{S}) = \mathrm{sign}(\mathbf{K})$ holds. Under the assumption that an attacked image will not be perceptually different from the original one [see (12)], $\mathrm{sign}(\mathbf{S}) = \mathrm{sign}(\mathbf{V})$ should hold. Based on this, (30) can be derived. Therefore, the hiding strategy should be designed so as to satisfy $\mathrm{sign}(\mathbf{V}) = \mathrm{sign}(\mathbf{K})$. This design has been realized by means of positive modulation [17], as expressed in (15).

IV. Experimental Results

Five standard images of size 256 × 256 were used as the host images in our experiments. Using our watermarking scheme,

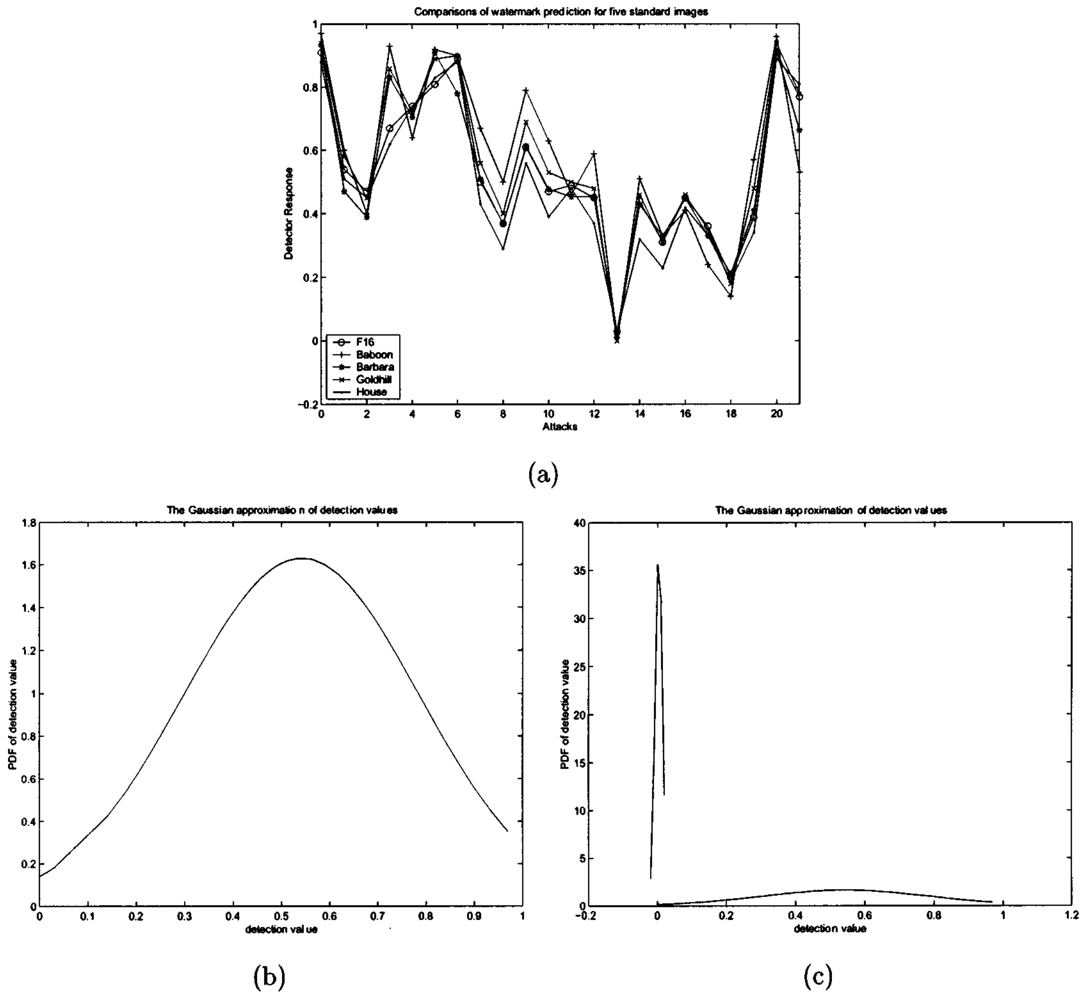
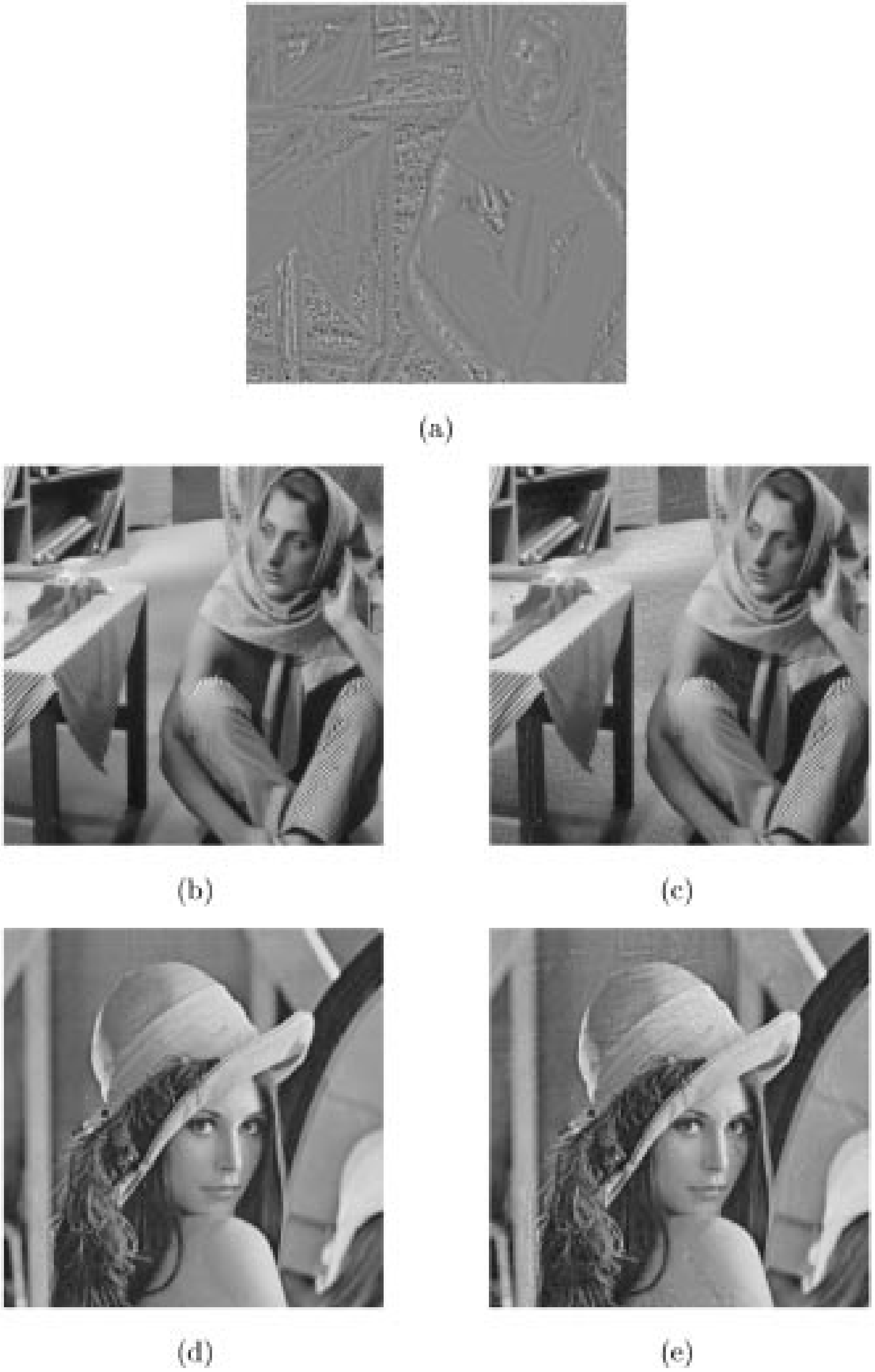


Fig. 7. Analysis of false positive and false negative: (a) comparisons of the detection results for five images under 21 attacks; (b) distribution of 100 watermarked/attacked images; (c) distribution on the right is rescaled version of (b), but the one on the left is the distribution formed by 90 nonwatermarked images.

TABLE I
PROBABILITIES OF FALSE NEGATIVE (p_{fn}) AND FALSE POSITIVE (p_{fp}) CORRESPONDING TO DIFFERENT THRESHOLDS (T)

Threshold (T)	p_{fn}	p_{fp}
0.0200	1.73×10^{-2}	2.25×10^{-2}
0.0225	1.77×10^{-2}	1.20×10^{-2}
0.0250	2.09×10^{-2}	6.10×10^{-3}

we set the length of a hidden watermark as 16 128. After watermarking was applied, the PSNR values of the five watermarked image were between 41 and 42 dB, and no perceptual distortion could be observed. 21 commonly used attacks were adopted to test the robustness of our method. These attacks included 1) blurring; 2) median filtering; 3) Wiener filtering; 4) rescaling; 5) histogram equalization; 6) sharpening; 7) and 8) Gaussian noise addition with different variance values; 9) and 10) uniform noise addition with different variance values; 11) mosaic effects; 12) texturizing; 13) shading; 14) and (15) JPEG compression with quality factors of 10% and 5%; 16) and 17) SPIHT compression with ratios of 16 : 1 and 32 : 1; (18) StirMark [24]; (19) dithering; (20) wavelet shrinkage-based denoising [5]; (21) sparse code shrinkage-based denoising [9]. Therefore, there were in total 110 attacked images (including five watermarked images). Among them, the original and the watermarked Barbara images are, respectively, shown in Fig. 5(a) and (b). The two Barbara images which were attacked, respectively, by means of Gaussian noise adding and shading are shown in Fig. 5(c) and (d). Three watermark prediction techniques, wavelet shrinkage-based denoising [5], sparse code shrinkage-based denoising [9], and Wiener filtering [15], were compared in terms of robustness. The comparison results based on the Barbara image are shown in Fig. 5(e). From Fig. 5(e), it can be found that the results obtained by applying Wiener filtering was the worst since prediction (denoising) in this case is not consistent with our modulation operation. We also found that none of the three denoising techniques could correctly predict the hidden watermark from an attacked image with the shading effect (13th attack). The reason why the shading effect attack could succeed was that the signs of most of the chosen coefficients changed. As a result, the predicted watermark values had signs which were different from their original ones. As we have noted with respect to (12), these sign changes violate our basic assumption and, thus, degrade the correlation value. Fig. 5(f) shows the result of the uniqueness test when



(a)

(b)

(c)

(d)

(e)

Fig. 8. Effects of the "denoising and remodulation attack" [28] and the "copy attack" [14]: (a) predicted watermark of Fig. 5(b) using the adaptive Wiener filter [5]; (b) and (c) watermarked images with the predicted watermark (a) removed using different weights; and (d) and (e) predicted watermark (a) added into a nonwatermarked image using different weights.

the famous StirMark attack (except for its geometric effects) was applied.

In the second group of experiments, we applied SPIHT compression with different compression ratios to see how the correlation value was affected. Fig. 6 shows a curve which reflects the change of the detector response under different compression ratios. It is apparent that when the ratio was small, its corresponding detector response was large. When the compression ratio reached $128:1$, the corresponding detector response dropped to 0.2.

In the third group of experiments, we obtain false positive and false negative numerical results. In Fig. 7(a), we compare the detection results obtained by applying 21 attacks to the five host images. We find that the five curves are very consistent. A Gaussian distribution was used to approximate these 110 detection results, as shown in Fig. 7(b). In addition, the detection results obtained from 90 nonwatermarked images were also approximated by means of another Gaussian distribution, as shown on the left-hand side of Fig. 7(c). The distribution shown on the right hand side of Fig. 7(c) was a redrawn version of

Fig. 7(b). It is clear that the distribution formed by the 90 nonwatermarked images was a sharp peak clustered around a detection value close to zero. On the other hand, the distribution formed by the five watermarked images and 105 attacked images was an obtuse curve centered at a detection value close to 0.5. According to the results of our experiments, the mean and standard deviation of the distribution formed by the 90 watermarked but nonattacked images were 0.94 and 0.04, respectively. On the other hand, the mean and standard deviation of the distribution formed by the 90 nonwatermarked images were 0 and 0.01, respectively. Based on (19) and (20), a threshold could be easily determined to obtain that both the false negative and the false positive were negligibly small under a nonattack situation. However, when attacks were imposed, the mean and standard deviation of the distribution formed by the 110 attacked images were 0.54 and 0.24, respectively. Under these circumstances, both false negative and false positive were expected to increase no matter what the threshold T was. In Table I, we show the false negative and the false positive probabilities corresponding to different threshold values.

Finally, we conducted experiments to see how a "denoising and remodulation attack" [28] and a "copy attack" [14] would affect the proposed scheme. First, the hidden watermark was predicted from the watermarked image shown in Fig. 5(b) using Wiener filtering [15]. The predicted watermark was shown in Fig. 8(a). As for the "denoising and remodulation attack," the predicted watermark was subtracted from the watermarked image to which it belonged [Fig. 5(b)]. Since our objective was to demonstrate how to remove the predicted watermark, the transparency issue was not a major concern. Therefore, the predicted watermark was directly subtracted from Fig. 5(b), and the de-watermarked image is shown in Fig. 8(b). In addition, the predicted watermark was also triplicated and then subtracted to yield a de-watermarked image, as shown in Fig. 8(c). As expected, Fig. 8(c) is less transparent than Fig. 8(b). However, the detection results obtained from Fig. 8(b) and (c) show that the hidden watermark still survived with a high correlation value ($\approx$0.79). This implies that the proposed scheme is insensitive to the weight added to the predicted watermark which is to be removed. This phenomenon clearly indicates that our scheme is able to preserve the signs of the watermark values. In addition, the predicted watermarks with different weights were added to the nonwatermarked "Lenna" image, as shown in Fig. 8(d) and (e), to examine the effect caused by a "copy attack." Similarly, the detection results reveal that no watermark was detected when our scheme was applied (the detection values were close to zero). That is, the false positive problem did not occur.

There are some critical issues that should be particularly addressed. First, the length of the current watermark is defined to be about one-quarter of an original image size. However, to claim the rightful ownership (in the application of robust watermarking) needs only about $64 \sim 80$ watermark bits (before error correction coding). In other words, it is possible to reduce the size of embedding places so that some high-frequency components cannot be watermarked. This kind of arrangement will prevent these components from being removed by an attack such as compression. Secondly, our false probability analysis was conducted with respect to all attacks instead of one particular attack. This will lead to higher but more practical false positive and false negative probabilities, as indicated in Table I. If we want to know the false probability with respect to one particular attack, some benchmarking algorithms [13], [7] should be adopted to evaluate a watermarking approach. Under these circumstances, the false probabilities will be significantly reduced.

V. Conclusion

In this paper, a novel watermarking approach, called the "nonblind" embedder, has been applied by exploiting the available information of denoising-based watermark prediction. We have found that the information obtained using shrinkage-based denoising (soft-thresholding) techniques is easy to control, and, that denoising itself is, in fact, a solution for oblivious watermark detection. The knowledge at the detector side can then be utilized to design a "nonblind" embedder, which is extremely advantageous over the common blind embedders. On the other hand, the predicted watermark can be purposely used to remove a hidden watermark or to confuse judgment about legal ownership. Therefore, we have conducted analysis to confirm that our method indeed can resist the "denoising and remodulation attack" and the "copy attack." The performance of our scheme, composed of a nonblind embedder and a blind detector, has also been analyzed regarding false negative and false positive probabilities.

At the present, it still is not possible for a watermarking scheme to resist all attacks because attackers are always smarter and one step ahead. Therefore, our first future work will focus on the problem of geometric attack resistance, which has not been treated in this paper. In the past, some techniques such as pilot signal [29], template [23], and invariant transform [16] have been developed to deal with geometric attacks. Unfortunately, resistance to both removal attacks and geometric attacks is still a challenging and contradictory problem up to now. In addition, each image should be associated with a secret key in our current watermarking design such that only semi-blind detection can achieve. This should be further improved. Finally, the problems of public-key detection [6] and other known attacks [30] should also be studied in order to obtain a mature, practical watermarking system.

Appendix A
Performance Analysis of the Proposed Scheme Under Different Variance (σ) Values

Recall that $w_{s,o}(x,y)/w^m_{s,o}(x,y)$ is the original/modulated wavelet coefficient of $\mathbf{X}/\mathbf{X}^m$ at scale s, orientation o, and position (x,y). The attacked wavelet coefficient is denoted as $w^a_{s,o}(x,y)$ with respect to $\tilde{\mathbf{X}}$. After conducting sparse code shrinkage-based denoising on $\tilde{\mathbf{X}}$, the estimated original wavelet coefficient $\bar{w}_{s,o}(x,y)$ satisfies $|\bar{w}_{s,o}(x,y)| < |w^a_{s,o}(x,y)|$ because shrinkage (i.e., soft thresholding) is an operation which gradually reduces the magnitude of a coefficient. Now, we quantitatively analyze the relationship between the SCS-based denoising process and the positive modulation process (15) as follows. According to the function of positive modulation, we know that $|w^m_{s,o}(x,y)| > |w_{s,o}(x,y)|$. When attacks are encountered, we may have three possible situations: (P1) $|w_{s,o}(x,y)| < |w^m_{s,o}(x,y)| < |w^a_{s,o}(x,y)|$; (P2) $|w_{s,o}(x,y)| < |w^a_{s,o}(x,y)| < |w^m_{s,o}(x,y)|$; and (P3) $|w^a_{s,o}(x,y)| < |w_{s,o}(x,y)| < |w^m_{s,o}(x,y)|$. To simplify the analysis, we assume that (12) holds. If (12) does not hold, then either 1) $w^m_{s,o}(x,y)$ is small or 2) the behavior caused by attacks is extremely different from that caused by the embedding process and is, thus, difficult to predict. With this basic assumption, the extracted watermark value $\tilde{k}(i)$ derived from $w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y)$ satisfies

$$\begin{aligned}\operatorname{sign}(\tilde{k}(i)) &= \operatorname{sign}(w^a_{s,o}(x,y) - \bar{w}_{s,o}(x,y)) \\ &= \operatorname{sign}\left(w^a_{s,o}(x,y)\right). \end{aligned} \tag{31}$$

Similarly, the hidden watermark value $k(i)$ satisfies

$$\operatorname{sign}(k(i)) = \operatorname{sign}\left(w^m_{s,o}(x,y) - w_{s,o}(x,y)\right). \tag{32}$$

Under situation (P1) and after applying sparse code shrinkage with different values of σ, we can get

$$|\bar{w}_{s,o}(x,y)| < \left|w^m_{s,o}(x,y)\right| < \left|w^a_{s,o}(x,y)\right| \tag{33}$$

when σ is large or

$$\left|w^m_{s,o}(x,y)\right| < |\bar{w}_{s,o}(x,y)| < \left|w^a_{s,o}(x,y)\right| \tag{34}$$

when σ is small. From (33) and (34), we know that the extracted watermark will have the same sign as the hidden watermark. It is intuitive that preservation of the same sign between the value of a hidden watermark and that of an extracted watermark will be beneficial for deriving a higher correlation value. Under the conditions that (P2) is valid and that sparse code shrinkage has been executed, we can get

$$|\bar{w}_{s,o}(x,y)| < \left|w^a_{s,o}(x,y)\right| < \left|w^m_{s,o}(x,y)\right| \tag{35}$$

whether σ is small or large. Again, (35) tends to help increase the correlation value, which is the same as in the case of (P1). Similarly, if the situation is (P3) and sparse code shrinkage has been executed, then we have

$$|\bar{w}_{s,o}(x,y)| < \left|w^a_{s,o}(x,y)\right| < \left|w^m_{s,o}(x,y)\right| \tag{36}$$

whether σ is small or large. Once again, (36) will help increase the correlation value, which is the same as in the cases of (P1) and (P2). From the above analysis, we find that different σ values will not affect the correlation value significantly because the polarity of the value of an extracted watermark can always be kept the same as that of the original watermark.

References

[1] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "Copyright protection of digital images by embedded unperceivable marks," *Image Vis. Comput.*, vol. 16, pp. 897–906, 1998.

[2] S. Baudry, P. Nguyen, and H. Maitre, "Channel coding in video watermarking: Use of soft decoding to improve the watermark retrieval," *Proc. IEEE Int. Conf. Image Processing*, vol. 3, pp. 21–24, 2000.

[3] P. Comon, "Independent component analysis—A New Concept," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[4] I. J. Cox, M. L. Miller, and A. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, pp. 1127–1141, July 1999.

[5] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *J. R. Statist. Soc. B*, vol. 57, pp. 301–337, 1995.

[6] T. Furon and F. P. Duhamel, "Robustness of an asymmetric watermarking method," *Proc. IEEE Int. Conf. Image Processing*, vol. 3, pp. 21–24, 2000.

[7] F. Perez-Gonalez, J. R. Hernandez, and F. Balado, "Approaching the capacity limit in image watermarking: A perspective on coding techniques for data hiding applications," *Signal Process.*, vol. 81, pp. 1215–1238, 2001.

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

[9] A. Hyvarinen, "Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation," *Neural Comput.*, vol. 11, pp. 1739–1768, 1999.

[10] A. Hyvarinen, P. Hoyer, and E. Oja, "Image denoising by sparse code shrinkage," in *Intelligent Signal Processing*, S. Haykin and B. Kosko, Eds. New York: IEEE Press, 2001.

[11] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proc. IEEE*, vol. 87, pp. 1167–1180, July 1999.

[12] M. Kutter, F. Jordan, and F. Bossen, "Digital signature of color images using amplitude modulation," *J. Electron. Imag.*, vol. 7, pp. 326–332, 1998.

[13] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," *Proc. SPIE*, vol. 3657, pp. 226–239, 1999.

[14] M. Kutter, S. Voloshynovskiy, and A. Herrigel, "The watermark copy attack," *Proc. SPIE*, vol. 3971, 2000.

[15] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 2, pp. 165–168, 1980.

[16] C.-Y. Lin, M. Wu, Y. M. Lui, J. A. Bloom, M. L. Miller, and I. J. Cox, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. Image Processing*, vol. 10, pp. 767–782, May 2001.

[17] C. S. Lu, S. K. Huang, C. J. Sze, and H. Y. M. Liao, "Cocktail watermarking for digital image protection," *IEEE Trans. Multimedia*, vol. 2, pp. 209–224, Dec. 2000.

[18] C. S. Lu and H. Y. M. Liao, "Oblivious cocktail watermarking by sparse code shrinkage: A regional- and global-based approach," *Proc. 7th IEEE Int. Conf. on Image Processing*, vol. 3, pp. 13–16, 2000.

[19] C. S. Lu, H. Y. M. Liao, and M. Kutter, "A new watermarking scheme resistant to denoising and copy attacks," *Proc. 4th IEEE Workshop Multimedia Signal Processing*, pp. 505–510, 2001.

[20] C. S. Lu and H. Y. M. Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Trans. Image Processing*, vol. 10, pp. 1579–1592, Oct. 2001.

[21] M. L. Miller, I. J. Cox, and J. A. Bloom, "Informed embedder: Exploiting image and detector information during watermark insertion," in *Proc. 7th IEEE Int. Conf. Image Processing*, vol. 3, Vancouver, BC, Canada, 2000, pp. 1–4.

[22] F. Mintzer and G. W. Braudaway, "If one watermark is good, are more better?," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 2067–2070, 1999.

[23] S. Pereira and T. Pun, "Fast robust template matching for affine resistant image watermarks," in *Proc. 3rd Int. Workshop on Information Hiding*, 1999, pp. 199–210.

[24] F. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," in *Proc. 2nd Workshop Information Hiding*, 1998, pp. 218–238.

[25] F. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding: A survey," *Proc. IEEE*, vol. 87, pp. 1062–1078, 1999.

[26] P. C. Su, C.-C. J. Kuo, and H. J. Wang, "Blind digital watermarking for cartoon and map images," in *Proc. SPIE Int. Symp. Electronic Imaging*, 1999.

[27] S. Voloshynovskiy, A. Herrigel, N. Baumgartner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," in *Proc. 3rd Int. Workshop on Information Hiding*, Dresden, Germany, 1999, pp. 211–236.

[28] S. Voloshynovskiy, S. Pereira, A. Herrigel, N. Baumgartner, and T. Pun, "Generalized watermarking attack based on watermark estimation and perceptual remodulation," *Proc. IEEE*, vol. 3971, 2000.

[29] S. Voloshynovskiy, F. Deguillaume, S. Pereira, and T. Pun, "Optimal adaptive diversity watermarking with channel state estimation," *Proc. SPIE*, vol. 4314, 2001.

[30] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun, "Attack modeling: Toward a second generation watermarking benchmark," *Signal Process.*, vol. 81, pp. 1177–1214, 2001.

[31] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, pp. 1164–1175, Aug. 1997.

Chun-Shien Lu (M'99) received the Ph.D. degree in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, R.O.C., in 1998.

Since October 1998, he has been a Postdoctoral Fellow with the Institute of Information Science, Academia Sinica, Taipei, Taiwan. His current research interests include multimedia security (watermarking and steganography), multimedia information processing, multimedia communication, and time-frequency analysis of signals and images.

Dr. Lu is the recipient of the Best Paper Award from the Image Processing and Pattern Recognition Society of Taiwan in 2000 for his work on digital watermarking and the Paper Award from the same society in 1997, 1999, and 2001. He organized and chaired a special session on data hiding and multimedia security at the Second IEEE Pacific-Rim Conference on Multimedia, Beijing, China, in 2001. He has been invited to organize a special session on Multimedia Security for the Third IEEE Pacific-Rim Conference on Multimedia (PCM'2002) and in the Sixth World Multiconference on Systemics, Cybernetics, and Informatics (SCI'2002). He is a member of the IEEE Signal Processing Society.

Hong-Yuan Mark Liao (S'87–M'89–SM'00) received the B.S. degree in physics from National Tsing-Hua University, Hsinchu, Taiwan, in 1981, and the M.S. and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 1985 and 1990, respectively.

He was a Research Associate with the Computer Vision and Image Processing Laboratory at Northwestern University during 1990–1991. In July 1991, he joined the Institute of Information Science, Academia Sinica, Taipei, Taiwan, as an Assistant Research Fellow. He was promoted to Associate Research Fellow and then Research Fellow in 1995 and 1998, respectively. From August 1997 to July 2000, he served as the Deputy Director of the institute. Currently, he is the Acting Director of Institute of Applied Science and Engineering Research. His current research interests include multimedia signal processing, wavelet-based image analysis, content-based multimedia retrieval, and multimedia protection. He is now the Managing Editor of the *Journal of Information Science and Engineering*. He is on the editorial boards of the *International Journal of Visual Communication and Image Representation*, *Acta Automatica Sinica*, and the *Tamkang Journal of Science and Engineering*.

Dr. Liao was the recipient of the Young Investigators' Award from Academia Sinica in 1998; the Excellent Paper Award from the Image Processing and Pattern Recognition Society of Taiwan in 1998 and 2000; and the Paper Award from the same society in 1996 and 1999. He served as the Program Chair of the International Symposium on Multimedia Information Processing (ISMIP'97) and the Program Co-chair of the Second IEEE Pacific-Rim Conference on Multimedia (2001). He also served on the program committees of several international and local conferences. He is on the editorial board of the IEEE TRANSACTIONS ON MULTIMEDIA.

Martin Kutter received the M.S. degree in electrical engineering in 1996 from the University of Rhode Island, Kingston, and the Ph.D. degree in electrical engineering in 1999 from the Swiss Federal Institute of Technology (EPFL), Lausanne.

He created the Digital Watermarking World, an Internet forum for people active in digital watermarking, and is owner of the watermarking mailing list which has more than 700 subscribers worldwide. In 2001, he co-founded AlpVision SARL, a digital watermarking company in Vevey, Switzerland. His research interests include digital watermarking of multimedia data, cryptography, data compression, image/video processing, and visual special effects.

Dr. Kutter received the Best Thesis Award from the Swiss Federal Institute of Technology in 2000 for his Ph.D. thesis "Digital image watermarking: Hiding information in images."