

Improving Robustness of Quantization-Based Image Watermarking via Adaptive Receiver

Xiangui Kang, *Member, IEEE*, Jiwu Huang, *Senior Member, IEEE*, and Wenjun Zeng, *Senior Member, IEEE*

Abstract—In this paper, the watermarking channel is modeled as a generalized channel with fading and *non-zero mean* additive noise. In order to improve the watermark robustness against the generalized channel, we present an optimized watermark extraction scheme by using an adaptive receiver for quantization-based watermarking. In the proposed extraction scheme, we adaptively estimate the decision zone of the binary data bits and the quantization step size. A training sequence is embedded to the original image together with the informative watermark. The estimation of the decision zone takes advantage of the response function of the training sequence. Compared to those watermarking schemes without receiver adaptation, the main improvement is the enhanced robustness against median filtering, image intensity Direct Current (DC) change, histogram equalization, color reduction, image intensity linear scaling, image intensity non-linear scaling such as Gamma correction etc.

Index Terms—Watermark, Adaptive receiver, Quantization

I. INTRODUCTION

Digital watermarking has emerged as a potentially effective tool for multimedia copyright protection, fingerprinting, tamper proofing, access control, annotation, and authentication [1][2][3]. In general, a watermark embedding scheme adopts one of the following two basic embedding mechanisms: Type-I and TYPE-II [4]. Type-I is the spread spectrum based watermarking, in which a noise-like watermark is added to a host image and is detected via a correlator [2] [3][5][6][7]. It is a host-interference non-rejecting method [4]. Type-II is the relationship enforcement based data embedding [4][8]-[14]. Most of them belong to quantization-based embedding. It is a host-interference rejecting method, meaning host-interference can be eliminated in watermark detection if the watermarking system is well designed [9]. In this paper, we focus on the quantization index modulation (QIM) methods. As discussed in [9], QIM methods are “provably good” against arbitrary bounded attacks, which arise in several copyright applications. In particular, they achieve provably better rate-distortion-robustness trade-offs than the spread-spectrum based methods (Type-I). In some other applications such as ownership protection, fingerprinting, and access control, the robustness of the watermarks may be critical. Fewer than 100 information bits (for example, 60 bits, according to [14]-[16]),

may be sufficient to represent authorship information, a time stamp, copyright information, or control information. The watermarked data may encounter a variety of legitimate processing as well as malicious attacks. A serious problem that limits some practical applications of digital watermarking technology is the insufficient robustness of the existing watermarking algorithms.

Digital watermarking is often modeled as a communication system with side information [2]. In many existing watermarking schemes, especially quantization-based schemes, the watermarking channel noise is always implicitly modeled as additive noise with a *zero-mean* (and a variance of σ^2) [4][9], or as watermark amplitude scaling followed by addition of *zero-mean* noise [12][13], where the amplitude scaling factor can be estimated by Fourier analysis method or maximum likelihood method [12]. But in some cases such as median filtering, a commonly used signal processing procedure for denoising, the noise in the watermarking channel is additive noise with a *non-zero* mean, as will be demonstrated in Section III. As a result, thus derived watermarking schemes (with the assumption of additive noise with a *zero-mean*) is typically not robust enough, especially to median filtering [7][14][16], image intensity DC change, histogram equalization etc. It was shown that robustness to median filtering is a difficult problem to handle in digital watermarking, for both TYPE-I [7] and TYPE-II watermarking [14][16], because median filtering damages the watermark severely. For example, the PSNRs of the filtered images with 2x2, 3x3, 5x5, 7x7 median filters are as low as 25.9 dB, 31.1 dB, 26.4 dB and 24.0 dB respectively, for the “Baboon” image. In order to address a general channel with fading and *non-zero* mean additive noise [17], we use a training sequence to adaptively estimate the decision zone after the estimation of the amplitude scaling factor using the Fourier analysis [12], thus significantly enhancing the performance in watermark extraction.

Note that there are some prior works in Type-I watermarking [15][19] or Type-II watermarking [8] that adopt a training sequence (or called reference watermark) to facilitate the proper detector settings in order to enhance the robustness of watermark extraction. [8][15][19] adopt the diversity technique, in which the watermarking channel is divided into multiple sub-channels for better local channel characterization. The reference watermark is used to estimate the sub-channel state, such as the BER (bit error rate) which can then be used to weight the contribution of the sub-channel to watermark extraction [8], or to estimate the presumably *zero mean* local noise parameters such as noise variance and fading parameters of the extracted watermark [15][19]. The main difference between our work and these prior works is in the channel model

Manuscript received MMM DD 2005; revised MMM DD, 2008. This work is supported by NSFC (90604008, 60633030, 60403045), NSF of Guangdong (04205407, 04009742) and 973 Program (2006CB303104).

Xiangui Kang and Jiwu Huang are with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275, China. (e-mail: isskxg@mail.sysu.edu.cn, isshjw@mail.sysu.edu.cn)

Wenjun Zeng is with the Dept. of Computer Science, University of Missouri-Columbia, MO65211, U. S. A.(e-mail: zengw@missouri.edu).

assumption, i.e., we assume the watermarking channel is *non-zero* mean additive noise. Furthermore, dividing the watermarking channel into multiple sub-channels tends to consume more training sequence bits, thus encroaching the embedding space of the watermark message. The method has no advantage given a global additive noise, as acknowledged in [8]. As will be shown in Section IV, the proposed scheme achieves much better robustness to additive noise corruption, JPEG compression, median filtering, and simultaneously achieves much better watermark invisibility than the scheme proposed in [8].

The rest of the paper is organized as follows. In Section II, we introduce the model of watermark embedding. Section III presents the proposed watermark extraction scheme with the adaptive receiver. Experimental results are presented in Section IV to compare the proposed scheme to both non-adaptive extraction schemes and others that do incorporate reference watermarks. Section V draws the conclusions.

II. MODEL OF WATERMARK EMBEDDING

To survive various attacks, we encode the message $\mathbf{m} \{m_i; i=1, \dots, L_m, m_i \in \{0,1\}\}$ with concatenated coding of Turbo code and direct sequence spread spectrum (DSSS) coding. To cope with possible bursts of errors in DSSS decoding, 2D interleaving [20] is exploited to interleave the DSSS coded bits \mathbf{W} . As discussed by Cox *et al.* [2], watermark should be embedded in the low frequency AC coefficients in the DCT domain due to their large perceptual capacity and robustness to signal processing. Huang *et al.* [7] extended this strategy to embed the watermark in the DC coefficients of block-based DCT transform. The strategy has also been extended to the DWT (discrete Wavelet transform) domain and has been shown to provide good performance in terms of robustness and invisibility [14]. We embed the *informative watermark* (i.e., the encoded message) into the LL_4 subband, e.g., the approximation band with a size of $(L_x/16) \times (L_y/16)$ for $L_x \times L_y$ images after a 4-level DWT, to make it robust while keeping the watermark invisible [14][16]. To achieve adaptive extraction, we embed a training sequence into the image

The watermark embedding process is shown in Fig. 1. The embedding is implemented as follows. A L_m -bit message \mathbf{m} is first encoded using a Turbo code with rate $1/2^1$ [22] to obtain the message $\mathbf{m}_c \{m_{ci}; i=1, \dots, L_c, m_{ci} \in \{0,1\}\}$ of length L_c . Then each m_{ci} of \mathbf{m}_c is DSSS [21] encoded using an N_p -bit bi-polar PN-sequence $\mathbf{p}=\{p_j; j=1, \dots, N_p\}$, where a bit of “1” is encoded as \mathbf{p} , and a bit of “0” as $-\mathbf{p}$, i.e., $m_{ci} \xrightarrow{\text{coding}} \mathbf{W}_i \{w_j; w_j \in \{-1,1\}, 1 \leq j \leq N_p, 1 \leq i \leq L_c\}$. A binary string \mathbf{W} is thus generated from \mathbf{m}_c . The training sequence $\mathbf{T} \{T_n; n=1, \dots, N_T, T_n \in \{-1,1\}\}$ is then distributed in random positions all

¹ A stronger FEC code can be used to improve the robustness, at the cost of more redundancy (thus smaller message embedding rate) and complexity. The focus of this paper, however, is to investigate the performance improvement provided by the proposed adaptive receiver without introducing additional FEC overhead.

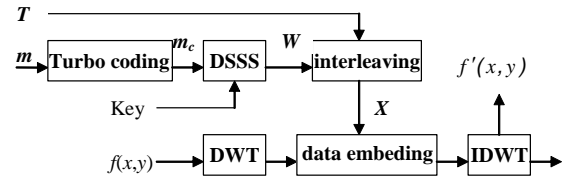


Fig.1. The watermark embedding process

over the image based on a key. In our experiments, \mathbf{T} is composed of $N_T/2$ bits of “1” and $N_T/2$ bits of “-1”.

In the implementation, we put the N_T -bit sequence \mathbf{T} into random positions of a $(L_x/16) \times (L_y/16)$ 2-D array based on a key and then the informative watermark \mathbf{W} is filled into the remaining portion of the above-mentioned array. By applying the 2-D interleaving technique [20] to this array, we obtain an interleaved 2-D array. Scanning this interleaved 2-D array in raster-scan order, we convert it into a 1-D array $\mathbf{X}=\{x_i\}$. We perform a 4-level DWT on the original image $f(x, y)$. The DWT coefficients in the LL_4 subband are then scanned in the same fashion to form a 1-D array \mathbf{C} . We adopt quantization-based embedding as shown in Eq. (1) ([1][8][9]) to embed the binary data \mathbf{X} into \mathbf{C} to obtain \mathbf{C}' , where $C(i)$ and $C'(i)$ denote the i^{th} element in \mathbf{C} and \mathbf{C}' , respectively. The quantizer $q(\cdot)$ is a uniform, scalar quantization function with a step size of S , and $q(\zeta)=kS+0.5S$, $k=\text{floor}(\zeta/S)$, where *floor* denotes the *floor*

$$\begin{cases} C'(i) = q(C(i) - \frac{1}{4}S) + \frac{1}{4}S, & \text{if } x_i = 1 \\ C'(i) = q(C(i) + \frac{1}{4}S) - \frac{1}{4}S, & \text{if } x_i = -1 \end{cases} \quad (1)$$

$$\begin{cases} x_i^* = 1, & r = C^*(i) \bmod S > \frac{S}{2} \\ x_i^* = -1, & \text{otherwise} \end{cases} \quad (2)$$

operation. The embedding strength S may be chosen so as to achieve a good compromise between the contending requirements of imperceptibility and robustness. If $x_i = -1$, $C'(i) \bmod S = 0.25S$. If $x_i = 1$, $C'(i) \bmod S = 0.75S$. Here *mod* denotes the modulus after division. By performing inverse DWT on the modified image, we obtain the watermarked image $f'(x, y)$.

In some cases such as JPEG compression, the effects of corruption are assumed to be equivalent to an additive noise with a *zero mean*. So one can extract the hidden binary data \mathbf{X}^* based on Eq. (2) with the best decision level chosen as $0.5S$. Here, $C^*(i)$ is the extracted coefficient. Eq. (2) suggests that if r ($r = C^*(i) \bmod S$) is in the interval of $(0, 0.5S)$, then the decision is made in favor of “ $x_i^* = -1$ ”, that is, $(0, 0.5S)$ represents the hidden data bit of “-1” or the decision zone of

the hidden data bit “-1” is $(0, 0.5S)$. The decision zone of the hidden data bit “1” is $(0.5S, S)$. However, in some other attack scenarios, this conventional watermark detector may not work well, as discussed in the next section.

III. THE PROPOSED WATERMARK EXTRACTION SCHEME USING ADAPTIVE RECEIVER

A. Channel Model and the Response Function of the Training Sequence

The watermark extraction is the inverse process of the watermark embedding. First, we perform the 4-level DWT on the test image. The coefficients of the LL_4 subband are scanned in the same way as used in the embedding process and mapped into a 1-D array, denoted as $C^* = \{C^*(i)\}$. A hidden data bit x_i is carried via a DWT coefficient $C'(i)$ in the LL_4 subband. The equivalent channel can in general be represented as [8][15][19]:

$$C^*(i) = \beta C'(i) + n(i). \quad (3)$$

That is, the channel is a fading channel with a fading factor β and an additive noise $n\{n(i)\}$. It is observed that the channel noise n is typically an additive noise with a *non-zero* mean for some signal processing and image manipulation (see Table I). For example, the mean of the channel noise ($n(i) = C^*(i) - C'(i)$ here) is 51.8 and 20.2 for the filtered “Baboon” image with the 3x3 and 2x2 median filter, respectively. Additional average noises in the spatial and frequency domains after median filtering are shown in Table I. It is observed that the effect of median filtering can be viewed as introducing additive noise with a *non-zero* mean to the watermarking channel (the same is true when median filtering is applied to non-watermarked images). The extraction method discussed above (in Eq. (2)) does not take into account this practically important case. Thus the extracted watermark using the method illustrated in Eq. (2) is not robust enough, as it effectively results in the increase of the energy of the channel noise due to the unrealistic assumption that the channel noise is additive noise with a *zero* mean.

For a better demonstration of the impact of channel noise with *non-zero* mean, we consider the case in which no fading exists, such as JPEG compression and median filtering. Fig. 2 and Fig. 3 show the distribution of r (refer to Eq. (2)) associated with a special training sequence T , half of which ($N_T/2=78$ in

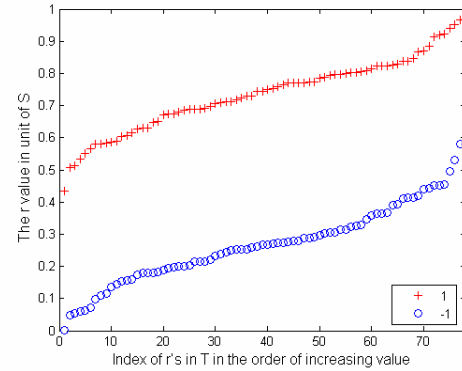


Fig. 2. The distribution of the training sequence in the JPEG_15 compressed Baboon image.

our case) is composed of “1” and the other half is composed of “-1”. In these figures, the r values corresponding to “1” and “-1” of T are sorted and indexed in increasing order, respectively, according to their values. That is, the horizontal axis denotes the index of the r values associated with the training sequence T in increasing order, and the vertical axis denotes the value of r in the unit of the embedding strength related parameter S . Because the training sequence is embedded in random positions together with the informative watermark and undergoes similar processing, the distribution of r

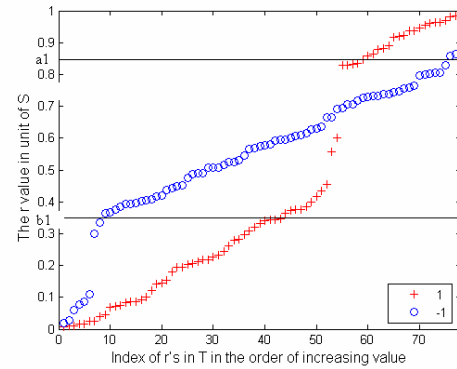


Fig. 3. The distribution of the training sequence in the filtered Baboon image with 3x3 median filter, and the interval $(a_1, S) \cup (0, b_1)$ that represents “1”. $a_1=0.85S$, $b_1=0.35S$.

TABLE I. THE AVERAGE NOISE INTRODUCED BY VARIOUS FILTERING

Test image	Lena						Baboon					
	2x2 Median	3x3 Median	5x5 Median	5x5 Gauss	5x5 Mean	Sharpening	2x2 Median	3x3 Median	5x5 Median	5x5 Gauss	5x5 Mean	Sharpening
The average noise of the gray level value	0.9	1.3	0.62	-0.3	-0.56	0.77	1.3	3.23	1.76	-0.29	-0.54	0.7
The average noise of the LL_4 coefficients	10.6	21.2	10.0	-4.8	-8.9	12.3	20.2	51.8	28.12	-4.6	-8.6	12.0

associated with the training sequence is expected to be similar to the distribution of r associated with the informative watermark. Fig. 2 shows the distribution of the training sequence in a JPEG compressed image with a quality factor of 15 (denoted as JPEG_15, and in general as “JPEG_quality-factor” in the rest of the paper). We can see that the interval containing most of the r 's corresponding to the “1”s of T is $(0.5S, S)$, and the interval containing most of the r 's corresponding to the “-1”s of T is $(0, 0.5S)$. Fig. 3 shows the distribution of the training sequence in a filtered image with a 3×3 median filter. We can see that because of the introduced *non-zero* mean noise mentioned above to the filtered image, the interval $(0.5S, S)$ does not contain most of the r 's corresponding to the “1”s of T any more, and the interval $(0, 0.5S)$ does not contain most of the r 's corresponding to the “-1”s of T . So Eq (2) apparently is not the best strategy for extracting the hidden bits when the test image is median filtered. We therefore propose to use a *sliding window* based method to determine the decision zone *adaptively* to handle the case of *non-zero* mean noise

B. Extraction Using Adaptive Receiver

In the extraction process, we calculate the r values associated with the N_T bits of the training sequence first, then we search for an interval that has the largest correlation value c_r with the training sequence among all the intervals with a fixed width of $0.5S$. The correlation value c_r is calculated as follows. If an interval contains N_1 r 's corresponding to the “1”s of T and N_{-1} r 's corresponding to the “-1”s of T , then

$$c_r = N_1 - N_{-1} - (N_T/2 - N_1) + (N_T/2 - N_{-1}) = 2(N_1 - N_{-1}) \quad (4)$$

The interval (i.e., *sliding window*) can be in the form of $(a_1, S) \cup (0, b_1)$, where $0 < b_1 < a_1 < S$, $b_1 + S - a_1 = 0.5S$, as shown in Fig. 3, or in the form of (c_1, d_1) , where $0 \leq c_1 < d_1 < S$, $d_1 - c_1 = 0.5S$. The resulting interval is then treated as the decision zone

of “1”. The remaining interval of $(0, S)$ is the decision zone of “-1”.

In the implementation, we divide the entire interval $(0, S)$ into N subdivisions, i.e., $(0, S/N), \dots, ((N-1)S/N, S)$, each with a width of S/N . Here, S/N is equivalent to the search step size when we search for an interval that has the largest correlation value c_r . We choose $N = 40$ in our work. Larger N may be chosen to improve the accuracy, but at the cost of search time. Our experiment shows that the equivalent channel noise introduced by some operations such as JPEG compression etc. is additive noise with *near-zero* mean, in which case $(0.50S, 1.00S)$ is the interval representing “1”. To better take advantage of such knowledge, only an interval with a width of $0.5S$ that satisfies the following two conditions may be chosen as the interval representing “1” to replace the default interval $(0.50S, 1.00S)$. The first condition is that it has the largest correlation value c_r . The second condition is that the c_r associated with it, is $0.025 \times N_T$ ($0.025 \times N_T = 4$ in our simulation case) larger than the c_r associated with the interval $(0.50S, 1.00S)$. The value of 0.025 is chosen empirically and is believed to be a reasonable value to allow random fluctuation.

Now we consider a general case with channel fading which is shown in Eq. (3). In this case, we need to replace S with $S' = \beta S$, which may be estimated by the Fourier analysis [12] using all coefficients C^* . According to Eq. (1) and Eq. (3), it is observed that the histogram of the coefficients C^* of the LL_4 subband shows prominent local maxima with a distance $S/2$ [12]. Note that no such structure is visible in the histogram of image spatial pixel values. Denote the maximum value and minimum value of C^* as C_{\max} and C_{\min} , and let the histogram of C^* with values in the range of $[C_{\min}, C_{\max}]$ has L_b bins, so each bin has a width of $\nabla = (C_{\max} - C_{\min}) / L_b$. We perform 1-D FFT of length L_f (we choose $L_b = L_f = 2048$ empirically in our work, larger L_b may result in more precise estimation but takes more time) to the histogram of C^* , then the first dominating peak [12] away from the zero frequency of the DFT magnitude

TABLE II. THE TEST RESULTS WITH BER FOR ADAPTIVE RECEIVING AND NON-ADAPTIVE RECEIVING (BEFORE AND AFTER MESSAGE DECODING).

	<i>Baboon</i> ($S=120$)		<i>Boat</i> ($S=100$)		<i>Peppers</i> ($S=100$)		<i>Lena</i> ($S=90$)									
	adaptive	non-adaptive	adaptive	non-adaptive	adaptive	non-adaptive	adaptive	non-adaptive								
BER before decoding and whether message recovered after decoding	before	after	before	after	before	after	before	after								
3x3 median_filter	0.28	yes	0.68	no	0.24	yes	0.49	no	0.17	yes	0.46	no	0.16	yes	0.28	yes
5x5 median_filter	0.19	yes	0.40	no	0.21	yes	0.36	no	0.17	yes	0.22	yes	0.09	yes	0.16	yes
color reduction	0.01	yes	0.49	no	0.23	yes	0.51	no	0.08	yes	0.52	no	0.24	yes	0.49	no
IDCC(-80)	0	yes	0.93	no	0	yes	0.16	yes	0	yes	0.16	no	0	yes	0.17	no
IS(1/8)	0	yes	0.49	no	0	yes	0.50	no	0	yes	0.48	no	0.03	yes	0.51	no
IDCC(+80)+IS(1/8)	0	yes	0.51	no	0	yes	0.52	no	0	yes	0.49	no	0.03	yes	0.51	no
Gamma Correction (1.2)	0.09	yes	0.18	yes	0.16	yes	0.48	no	0.16	yes	0.63	no	0.21	yes	0.50	no
Gamma Correction (0.8)	0.06	yes	0.68	no	0.14	yes	0.54	no	0.18	yes	0.39	no	0.22	yes	0.66	no

“IDCC” stands for “intensity DC change”; “IS” stands for “intensity scaling”.

spectrum is always at the location k_m where the corresponding normalized frequency is as follows.

$$k_m / L_f = (1/0.5S') / (1/\nabla) = \nabla / 0.5S' \quad (5)$$

So S' can be calculated as:

$$S' = 2 \times (L_f / k_m) \times (C_{\max} - C_{\min}) / L_b = 2 \times (C_{\max} - C_{\min}) / k_m \quad (6)$$

Based on the estimated S' and the derived decision zone, we extract the hidden data \mathbf{X}^* , and then perform 2-D de-interleaving to \mathbf{X}^* to obtain the sequences \mathbf{W}^* , which is composed of “1”, “-1”. We segment \mathbf{W}^* into the sequences of N_p bits and correlate the obtained sequence with the original PN -sequence \mathbf{p} . The obtained correlation value is treated as the *soft* decision value and is input to the log-MAP decoder for Turbo code [22]. The watermark message can thus be recovered.

To demonstrate the effectiveness of the proposed adaptive method which includes adaptively estimating the decision zone and the quantization step size, we show some of the test results with the bit error rate (BER) of the extracted binary sequence \mathbf{W}^* (the watermark before message decoding) with the proposed adaptive receiver and with non-adaptive method of that shown in Eq. (2). Note that information about whether the embedded message can be recovered without any error after Turbo decoding is also included in Table II. We can see that the proposed method with an adaptive receiver reduces the BER of the extracted binary sequence \mathbf{W}^* significantly, and thus allows us to recover the hidden message more faithfully, especially under the attack of median filtering, intensity DC change (IDCC in short), image intensity scaling (IS in short), color reduction, and gamma correction. Note that color reduction reduces the 256 grayscales (i. e. 8 bit image) to 16 shades of gray (i. e. 4-bit image) [8]. In Table II, the change introduced to the DC in intensity DC change is -80, and a scaling factor of 1/8 is used for intensity scaling. The combination attack of intensity DC change and intensity scaling includes intensity DC change with a DC change of +80 and intensity scaling with a scaling factor of 1/8. Both intensity DC change and intensity scaling may degrade the image quality severely, and fail the watermark extraction if non-adaptive receiver is used. With the proposed adaptive method, however, the watermark can still be extracted as shown in Table II.

IV. EXPERIMENTAL RESULTS

We have tested the proposed watermarking scheme on a database of 1000 images. Some representative results are reported here. A 60-bit message is embedded into each of the 512x512 gray scale images. We adopt $L_c = 124$, $N_T = 156$, $N_p = 7$ and adopt the Daubechies 9/7 bi-orthogonal wavelet in our work. Based on our many experiments on various images, S can be varied between 80–120 according to the texture complexity of the host image, and $S = 100$ is typically used for the test images as the PSNRs of the resulting watermarked images have an acceptable value of about 42.5 dB and the visual quality of

TABLE III. THE S VALUES AND THE PSNRs OF THE MARKED IMAGES W.R.T. THE ORIGINAL IMAGES

	<i>Lena</i>	<i>Baboon</i>	<i>Boat</i>	<i>Pepper</i>
S	90	120	100	100
PSNR (dB)	43.5	41.1	42.5	42.6

the watermarked images are not degraded. The parameter S for some images that we used in our experiments and the PSNRs of the marked images with respect to the original images are shown in Table III. The PSNRs of all watermarked images are larger than 41 dB. The watermarks are perceptually invisible.

Our experiments show that, for the above four test images (“Lena”, “Baboon”, “Boat”, and “Pepper”), the proposed technique can resist common signal processing such as JPEG compression with the quality factor from 10 to 100, median filtering (2×2, 3×3, 5×5, 7×7), and convolution filtering including Gaussian filtering (3×3, 5×5, 7×7), mean filtering (3×3, 5×5, 7×7), FMLR (Frequency Mode Laplacian Removal, a Stirmark function [18]) and sharpening very well. The embedded 60-bit watermark message can be recovered exactly for all the above test functions. In addition, we also test a few other functions. Crop_25 is a test function in StirMark that crops 25% of the pixels in each dimension, which means that a portion with a size of 384×384 is retained from the original image of a size of 512×512 (cropped by 43% of the total pixels). Color reduction reduces the 256 grayscales to 16 shades of gray (i.e. 4-bit image)[8]. In intensity DC change, a reasonable DC change range between -80 and +80 is used. A scaling factor ranging from 1.8 to 1/8 is used in intensity scaling, and in the combination attack of intensity DC change and intensity scaling, DC change is +80 and scaling factor is 1/8. The proposed watermarking scheme is robust to all the above attacks, i.e., the embedded 60-bit watermark message can be recovered without any bit error. For image cropping, we assume that the watermark extractor is applied to a resynchronized image (the cropped portion is padded with zero values). Especially, the watermark also survives non-linear amplitude scaling such as gamma correction with the gamma γ being a value between 0.8 and 1.2. That is, we perform the following modification to the image intensity y .

$$y' = 255 \times (y / 255)^\gamma \quad (7)$$

The proposed scheme can also recover the message exactly when the SNR of the image corrupted by Gaussian noise is merely 16.9 dB for “Lena” and 14.3 dB for “Baboon”.

The comparison between the proposed scheme and the scheme proposed in [8][14] on the “Baboon” image is shown in Table IV. A total of 1024 bits, including $N_T = 156$ bits reference watermark and 868 bits informative watermark \mathbf{W} (generated from the same 60-bit message using Turbo encoding and DSSS coding), are embedded in the DWT detail subbands using the scheme proposed in [8] and the LL subband using

TABLE IV. PERFORMANCE COMPARISON BETWEEN THE PROPOSED SCHEME AND THAT IN [8] AND [14].

	PSNR of marked image	JPEG_8~18	noise corrupted marked image, SNR14.3~23.5dB	2×2, 5×5 MF	histogram equalization	color reduction	IDCC	IS	Gamma correction
The proposed scheme	41.1dB	1	1	1	1	1	1	1	1
[14]	41.1dB	1	1	0	0	0	0	0	0
[8]	39.1dB	0	0	0	1	1	1	1	1

“MF” stands for “median filtering”; “IDCC” stands for “intensity DC change”; “IS” stands for “intensity scaling”.

the scheme proposed in [14] respectively. The extracted informative watermark W^* by [8] and [14] are further decoded to obtain a 60-bit message respectively. For a fair comparison, DSSS and Turbo encoding/decoding are applied to all schemes, so the three schemes have the same embedding payload (60 bits message payload) and the same embedding rate ($N_T = 156$ bits reference watermark versus 868 bits informative watermark W). Note that “1” in Table IV indicates that the embedded message can be recovered without error, and “0” indicates that the embedded message can not be recovered correctly. It is observed that the proposed scheme is robust to median filtering, color reduction, IDCC, IS, gamma correction, and histogram equalization. Histogram equalization changes the contrast of the image, and results in a visually similar image. While the scheme in [14], in which *zero-mean* channel noise is implicitly assumed and no fading is considered, is not robust to these attacks. Compared to the scheme in [8] that also uses a training sequence, the proposed scheme achieves much better robustness performance to additive noise corruption, JPEG compression, median filtering, and simultaneously achieves much better watermark invisibility because the PSNR of the watermarked image using the proposed scheme is 2 dB larger than that using the scheme in [8].

Note that we require that the 60-bit message be extracted without any error, so the false positive rate of the watermark is $2^{-60} = 10^{-18}$, assuming that the extracted bits are independent. We did not find any false positive of watermark on the 1000 test images in our experiments. Note that Table II ,IV do not include the test results of geometric distortion except for cropping because the main focus of this paper is to optimize the watermark extractor, not to provide a systematic watermarking system. Readers who are interested in watermark robustness to geometric distortions are referred to the recent works [3] [11] [14] that addresses watermark resynchronization.

V. CONCLUSIONS AND DISCUSSIONS

In this paper, by adaptively estimating the decision zone exploiting a training sequence and estimating the quantization step size using the Fourier analysis method, we model the data extraction process as one associated with a generalized channel of additive noise with a generally *non-zero* mean and fading. This appears to be particularly promising in enhancing the robustness of the watermarking system against median filtering, intensity DC change, color reduction, intensity linear scaling, non-linear intensity modification such as Gamma

correction etc. The proposed scheme is robust to common signal processing including Gaussian filtering, mean filtering, median filtering, sharpening, and JPEG compression with a quality factor of as low as 10. Compared to the watermarking scheme described in [14] in which no adaptation is employed, the main improvement is the enhanced robustness against median filtering, intensity DC change, intensity linear scaling, color reduction, histogram equalization and intensity non-linear scaling, etc. Compared to the scheme proposed in [8] which also uses a training sequence, the proposed scheme achieves much better robustness to additive noise corruption, JPEG compression, median filtering, and achieves much better watermark invisibility simultaneously. The proposed adaptive estimation of the decision zone can be combined with many existing watermarking schemes [7][14] to enhance their robustness.

ACKNOWLEDGMENTS:

The authors are grateful to the anonymous reviewers for their constructive comments, which have helped to significantly enhance the quality of this paper.

REFERENCES

- [1] P. Moulin and R. Koetter, “Data hiding codes,” *Proceedings of IEEE*, vol.93, no. 12, pp. 2083- 2126, Dec. 2005.
- [2] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoan, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Processing*, vol. 6, no. 12, pp.1837-1687, Dec. 1997.
- [3] C.-S. Lu, S.-W. Sun, C.-Y. Hsu, P.-C. Chang, “Media hash-dependent image watermarking resilient against both geometric attacks and estimate attacks based on false positive-oriented detection.” *IEEE Trans. on Multimedia*, vol. 8, no. 4, pp. 668-685, Aug. 2006.
- [4] M. Wu, B. Liu, “Data hiding in images and video: Part I: Fundamental Issues and Solutions,” *IEEE Trans. on Image Processing*, vol. 12, no.6, pp. 685-695, June, 2003.
- [5] Podilchuk, C., I. and Zeng, W., “Image-adaptive watermarking using visual models,” *IEEE Journal on Selected Areas in Communications*, vol.16, no.4, pp.525-539, 1998.
- [6] C.-Y. Lin, M. Wu, Y. M. Lui, J. Bloom, M. Miller, I. Cox, “Geometric Distortion Resilient Public Watermarking for Images,” *IEEE Trans. on Image Processing*, vol. 10, no.5, pp. 767-782, may 2001.
- [7] J. Huang, Y. Q. Shi, “Reliable information bit hiding,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol.12, no.10, pp. 916-920, 2002.
- [8] D. Kundur and D. Hatzinakos, “Diversity and attack characterization for improved robust watermarking,” *IEEE Trans. Signal Processing*, vol. 49, no. 10, pp. 2383-2396, 2001.
- [9] B. Chen, and G., W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1423- 1443, May 2001.

- [10] K. Solanki, U. Madhow, B. S. Manjunth, and S. Chandrasekaran, "Estimating and undoing rotation for print-scan resilient data hiding," *2004 IEEE International Conference on Image Processing (ICIP)*, pp. 39-42, Singapore, 2004.
- [11] D. He and Q. Sun, "A RST resilient object-based video watermarking scheme," *2004 IEEE International Conference on Image Processing (ICIP)*, pp. 737-740, Singapore, 2004.
- [12] I. D. Shterev, R. L. L. Lagendijk, and R. Heusdens, "Statistical amplitude scale estimation for quantization-based watermarking," *SPIE Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, San Jose, Jan. 2004.
- [13] J. Wang, I. D. Shterev, R. L. L. Lagendijk, "Scale estimation in two-band filter attacks on QIM watermark," *SPIE Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 6072, pp. B1-B10, San Jose, Jan. 2006.
- [14] X. Kang, J. Huang, and Y. Q. Shi, Y. Lin, "A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.13, no. 8, pp.776-786, Aug. 2003
- [15] F. Deguillaume, S. Voloshynovskiy, and T. Pun, "Secure hybrid robust watermarking against tampering and copy attack," *Signal Processing*, vol. 83, pp. 2133-2170, 2003.
- [16] X. Kang, J. Huang, and Y. Q. Shi, "An image watermarking algorithm robust to geometric distortion," In: *Proc. of 2002 Int. Workshop on Digital Watermarking (IWDW2002)*, Lecture Notes in Computer Science, vol. 2613, pp. 212-223, Seoul, Korea, 2002.
- [17] H. V. Zhao, M. Wu, Z. J. Wang, K. J. R. Liu, "Forensic analysis of nonlinear collusion attacks for multimedia fingerprinting," *IEEE Trans. on Image Processing*, vol. 14, no.5, pp. 646-661, May, 2005.
- [18] Fabien A. P. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 58-64, September 2000.
- [19] S. Voloshynovskiy, F. Deguillaume, S. Pereira, and T. Pun, "Optimal adaptive diversity watermarking with channel state estimation," In *Proc. of SPIE: Security and watermarking of Multimedia content III*, vol.4314, pp. 673-685, San Jose, CA, USA, 22-25, Jan. 2001.
- [20] Y. Q. Shi and X. M. Zhang, "A new two-dimensional interleaving technique using successive packing," *IEEE Transactions on Circuits and Systems, Part I: Fundamental Theory and Application*, vol. 49, no. 6, pp. 779-789, June 2002.
- [21] P.-C. Chen, Y.-S. Chen, and W.-H. Hsu, "A digital image watermarking system: modeling, performance analysis, and application," *Journal of Computers*, vol. 13, no. 1, pp. 1-10, 2001.
- [22] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. on Communications*, vol. 44, no. 10, pp. 1261-1271, Oct. 1996.
- [23] M. Wu, *Multimedia Data Hiding*, Ph. D. Thesis, Princeton University, Jun. 2001.

Xiangui Kang (M'00) received the B.S., M.S., and Ph.D. degrees from Peking University, in 1990, Nanjing University in 1993, Sun Yat-Sen University, in China, in 2004 respectively. He is currently an associate professor with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China. His research interests include watermarking, multimedia communications and security. He is a member of the IEEE ComSoc's Multimedia Communications Technical Committee.



Jiwu Huang (SM'00) received the B.S. degree from Xidian University, China, in 1982, the M.S. degree from Tsinghua University, China, in 1987, and Ph.D. degree from Institute of Automation, Chinese Academy of Science in 1998. He is currently a Professor with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China. His current research interests include multimedia security and data hiding.



Dr. Huang serves as a member of IEEE CAS Society Technical Committee of Multimedia Systems and Applications

Wenjun Zeng (S'94-M'97-SM'03) received his B.E., M.S., and Ph.D. degrees from Tsinghua University, China, in 1990, the University of Notre Dame in 1993, and Princeton University in 1997, respectively, all in electrical engineering.



He has been an Associate Professor with the Computer Science Department of University of Missouri, Columbia, MO since Aug. 2003. Prior to that, he had worked for PacketVideo Corporation, San Diego, CA, Sharp Labs of

America, Camas, WA, Bell Laboratories, Murray Hill, NJ, and Matsushita Information Technology Lab, Panasonic Technologies Inc., Princeton, NJ. He has also consulted with Microsoft Research, Huawei Technologies, and a couple of start-up companies. From 1998 to 2002, He was an active contributor to the MPEG4 Intellectual Property Management & Protection (IPMP) standard and the JPEG 2000 image coding standard, where four of his proposals were adopted. He has been awarded 12 patents. His current research interests include multimedia communications and networking, content and network security, wireless multimedia, and distributed source and channel coding.

Dr. Zeng has served as an Organizing Committee Member and Technical Program Committee Chair/Member for a large number of IEEE international conferences. He is an Associate Editor of the *IEEE Transactions on Multimedia*, and is on the Editorial Board of *IEEE Multimedia Magazine*. He co-guest edited the *Proceedings of the IEEE's Special Issue on Recent Advances in Distributed Multimedia Communications* published in January 2008, and was the Lead Guest Editor of *IEEE Transactions on Multimedia's Special Issue on Streaming Media* published in April 2004. In the recent past, he has served as the TPC Chair for the 2007 *IEEE Consumer Communications and Networking Conference (CCNC)*, the TPC vice-Chair for CCNC 2006, the TPC Co-Chair, *Multimedia Communications and Home Networking Symposium, 2005 IEEE Inter. Conf. Communication*. He is a member of the IEEE Signal Processing Society's Multimedia Signal Processing Technical Committee and the IEEE ComSoc's Multimedia Communications Technical Committee.