IMPROVE ROBUSTNESS OF IMAGE WATERMARKING VIA ADAPTIVE RECEIVING

Xiangui Kang[1] *Jiwu Huang*[1] *Yun Q. Shi* [2]

1. Dept. of Electronics, Sun Yat-Sen University, Guangzhou 510275, P. R. China, Email: isshjw@zsu.edu.cn
2. Dept. of ECE, New Jersey Institute of Technology, NJ 07102, USA.

ABSTRACT

Almost all the existing popular watermarking schemes model the watermarking channel noises as additive noise with zero-mean. However, our experiments show that this is not reasonable in the case of channel noise introduced by image filtering. This paper presents a new quantization-based watermarking scheme with enhanced robustness via adaptive receiving and Turbo coding in addition to other measures. The present algorithm can successfully resist almost all the StirMark testing functions including both common signal processing and geometric distortions in StirMark 4.0 except for random distortion.

1. INTRODUCTION

Watermarking is often modeled as a communication system with side information [1]. In many existing watermarking schemes, including the quantization-based schemes [2, 3] and additive spread spectrum schemes [1], the watermarking channel noise is always implicitly modeled as additive noise with zero-mean (and variance σ^2). But in some cases such as image filtering, the noise in watermarking channel is additive noise with non-zero mean due to fading (decreasing or increasing), so thus obtained watermarking is essentially not robust enough, especially to some attacks such as median filtering or those resulting in a fading channel, for example, the printing-and-scanning and the attack of multiplication by a constant followed by addition of Gaussian noise [4]. It is noted that the watermark robustness against median filtering and the related test results in StirMark have not been reported in detail in the literature. According to our work, robustness to median filtering is a tough problem to handle in watermarking [5] because median filtering damages the watermark severely and the PSNR of median filtered image versus the non-filtered image is rather low. For example, the PSNRs of 3x3, 5x5, 7x7 median filtered images are as low as 31.1dB, 26.4dB and 24.0dB respectively, for Baboon image. In this paper, we develop a new adaptive receiving scheme to improve the quantization-based watermark robustness including robustness against median filtering. Note that the adaptive receiving with additive spread spectrum watermarking was first introduced to handle the fading of the watermark extracted from a corrupted marked image (resulting from the adaptive embedding, the heavier watermark fading in flat image region than in textured region and etc.) in [6]. Fading of host media feature has not been addressed in [6]. In this paper, however, we use adaptive receiving to combat the fading of host media feature, thus enhancing the watermark robustness against median filtering for the first time. We can see that the proposed watermarking technique can resist almost all test functions in StirMark 4.0 except for random distortion.

2. WATERMARK EMBEDDING

To survive attacks, we encode the message $\boldsymbol{m}$ $\{m_i;\ i=1,\ \ldots,\ L,\ m_i \in \{0,1\}\}$ with concatenated coding of DSSS coding and Turbo code. To cope with bursts of errors possibly occurred with watermarked image, 2D interleaving [7] is exploited. We embed the informative watermark (the encoded message) into LL_4 subband (32x32 in our work for 512x512 images) in DWT domain to make it robust while keeping the watermark invisible [5]. To achieve adaptive receiving, we embed a training sequence randomly together with the informative watermark.

The watermark embedding is implemented as follows. The 60-bit message $\boldsymbol{m}$ is first encoded using Turbo code with rate 1/2 [8] to obtain the message $\boldsymbol{m}_c$ $\{m_{ci};\ i=1,\ \ldots,\ L_c,\ m_{ci} \in \{0,1\}\}$ of length L_c =124. Then each bit m_{ci} of $\boldsymbol{m}_c$ is DSSS encoded using an N_p-bit bi-polar *PN*-sequence $\boldsymbol{p}=\{p_j;\ j=1,\ \ldots,\ N_p,\ N_p=7$ in our work$\}$, where "1" is encoded spreadly as $\{+1\times p_j;\ j=1,\ \ldots,\ N_p\ \}$, "0" as $\{-1\times p_j;\ j=1,\ \ldots,\ N_p\ \}$, thus obtaining a binary string $\boldsymbol{W}$.

The training sequence $\boldsymbol{T}$ $\{T_n;\ n=1,\ \ldots,\ N_T\}$, $T_n \in \{-1,1\}$ should be distributed all over the image randomly based on a key. In our work, $\boldsymbol{T}$ is composed of N_T =156 bits "1" (N_T =32x32-$L_c\times N_p$). This training sequence also helps to achieve synchronization in the watermark extraction. If the correlation coefficient between the original training sequence $\boldsymbol{T}$ and the extracted training sequence $\boldsymbol{T}^*\{T^*_n;\ n=1\ldots N_T\}$ obtained from a test image satisfies the following condition

$$\rho_{T,T^*} = \frac{1}{N_T}\sum_{n=1}^{N_T}(T_n \cdot T_n^*) \geq thresh_1$$

we regard $\boldsymbol{T}^*$ matched to $\boldsymbol{T}$ and consider synchronization is achieved.

In implementation, we put N_T -bit sequence $\boldsymbol{T}$ into a 32x32 2-D array randomly based on a key and the binary string $\boldsymbol{W}$ is filled into the remaining portion of the above-mentioned array. By applying 2-D interleaving technique [7] to this array, we obtain another 2-D array. Scanning this interleaved 2-D array, say, row-by-row, we convert it into a 1-D array $\boldsymbol{X}=\{\ x_i\ \}$ $(0\leq i<1024)$. We perform a 4-level DWT on the original image $f(x,\ y)$ by using the Daubechies 9/7 bi-orthogonal wavelet filters. The DWT coefficients in the LL_4 subband are scanned in the same fashion to form a 1-D array

Supported by NSFC (60325208, 60172067, 60133020), Funding of China National Education Ministry.

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

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

$\boldsymbol{C}$. We adopt quantization-based embedding in Equation (1) [3], to embed the binary data $\boldsymbol{X}$ into $\boldsymbol{C}$ to obtain $\boldsymbol{C}'$, where $C(i)$ and $C'(i)$ denote the i^{th} element in $\boldsymbol{C}$ and $\boldsymbol{C}'$, respectively. The quantizer $q(.)$ is a uniform, scalar quantization function of step size S, and $q(x)= kS+0.5S$, $k=\left\lfloor \frac{x}{S} \right\rfloor$ ($k \in Z$), where $\lfloor . \rfloor$ means *floor* operation. Equation (1) indicates that the proposed embedding method tries to output a $C'(i)$ value which is closest to $C(i)$ and whose corresponding embedding bit value equals to x_i .The embedding strength S can be chosen so as to make a good compromise between the contending requirements of imperceptibility and robustness. Note that the difference between $C(i)$ and $C'(i)$ is between $-0.5S$ and $+0.5S$. If $x_i=-1$, $C'(i)$ *mod* $S=0.25S$. If $x_i=+1$, $C'(i)$ *mod* $S=0.75S$. Here *mod* denotes remainder after division. In some cases such as JPEG compression, the results of corruption are assumed to be equivalent to an additive noise with zero mean. So we can extract the hidden binary data $\boldsymbol{X}^*$ according to Equation (2). Here, $C^*(i)$ is the extracted coefficient. Equation (2) indicates that if r ($r = C^*(i) \bmod S$) is in the interval $(0, 0.5S)$, then the decision is made in favor of "$x_i^*=-1$", that is, $(0, 0.5S)$ represents the hidden data bit "–1". The interval representing "+1" is $(0.5S, S)$. Performing inverse DWT on the modified image, we obtain the watermarked image $f'(x, y)$.

3. WATERMARK EXTRACTION

The watermark extraction is the inverse process of the watermark embedding. First, perform the 4-level DWT on the test image. The coefficients of the LL_4 subband are scanned according to the same way as used in data embedding and mapped into a 1-D array, denoted by $\boldsymbol{C}^*=\{C^*(i)\}$. The hidden data bit x_i is transmitted via a DWT coefficient $C'(i)$ in the LL_4 subband. The equivalent channel noise $\boldsymbol{n}\{n(i)\}$ can be presented as: $n(i) = C^*(i) - C'(i)$. The above extraction method using Equation (2) does not take into account the practically important case of the additive noise with non-zero mean due to the channel fading. Fig. 1 shows the histogram of $\boldsymbol{n}$ when the watermarked Baboon image is 3x3 median filtered. Some average noises in spatial and frequency domains before and after filtering applied to Baboon watermarked image are shown in Table 1. It is observed that the result of filtering applied to watermarked images can be viewed as introducing additive noise with non-zero mean to the watermarking channel (it is similar when filtering applied to non-watermarked image). Thus the extracted watermark using the above mentioned method is not robust enough because the above method increases the energy of the channel noise due to the assumption of the channel noise to be additive noise with zero mean. For example, as a result of the above method, the watermark with S less than 260 for Baboon image (PSNR=33.9 dB versus original image when S=260) cannot resist 3x3 median filtering.

Table 1. The average difference before and after filtering applied to Baboon watermarked image

Filtering	3x3 Median	5x5 Median	5x5 Gauss	5x5 Mean	Sharpen-ning
The average difference of the gray level value	3.23	1.76	-0.29	-0.54	0.7
The average difference of LL_4 coefficients	51.8	28.12	-4.6	-8.6	12.0

According to our experiment, the interval $(0.5S, S)$ does not contain most of r corresponding to "1" of the embedded binary data $\boldsymbol{X}$, and the interval $(0, 0.5S)$ does not contain most of r corresponding to "–1" of $\boldsymbol{X}$ because of the fading of the DWT coefficients in the LL_4 subband of the filtered image. So the above extraction method (non-adaptive receiving), that is, the interval $(0, 0.5S)$ representing "–1" and $(0.5S, S)$ representing "1" fixedly, apparently is not the best for extracting the hidden bits when the test image is filtered. According to the Equation (1), the channel can be assumed to be binary symmetric channel. Because the training sequence $\boldsymbol{T}$ is embedded randomly together with informative watermark $\boldsymbol{W}$, the distribution of r associated with $\boldsymbol{T}$ is similar to the distribution of r associated with $\boldsymbol{W}$. In the embedding, the training sequence $\boldsymbol{T}$ is composed of *only* "1", i. e. N_T bits are all "1". In the extraction, we calculate the r values associated with the N_T bits of training sequence $\boldsymbol{T}$ first, then determine the two intervals representing "1" and "–1" adaptively, respectively, according to the distribution of the r associated with $\boldsymbol{T}$ (adaptive receiving).

We search for an interval that contains the most r values associated with the training sequence bits among all the intervals with a fixed width $0.5S$. The interval 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. 2, or in the form of (c_1, d_1), where $0<c_1<d_1<S$, $c_1-d_1=0.5S$. In Fig. 2, the r values corresponding to $\boldsymbol{T}$ are sorted according to their values from small to large. That is, the horizontal axis denotes the sorting order number of r associated with training sequence $\boldsymbol{T}$, and the vertical axis denotes the value of r in the unit of embedding strength related parameter S. Interval corresponding to "1" is then determined. The remaining interval within $(0, S)$ is the interval representing "–1".

Next we extract the hidden data $\boldsymbol{X}^*$ according to which of these intervals the r values fall into. If r is in the interval representing "1", the recovered hidden data bit value is 1; if r fall into the interval representing "–1", the recovered bit value is –1. Next we perform 2-D de-interleaving, which is

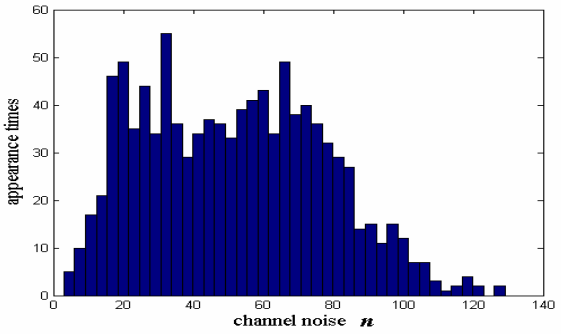

Fig.1. The histogram of the channel noise when watermarked image is 3x3median filtered (Baboon).

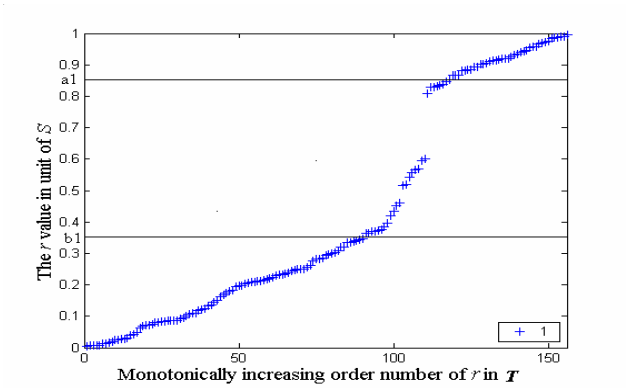

Fig.2. Determining the interval $(a_1, S) \cup (0, b_1)$ representing "1". a_1=0.85S, b_1=0.35S (Baboon, S =120).

the inverse process of 2-D interleaving [7], to $\boldsymbol{X}^*$ to obtain the binary sequences $\boldsymbol{W}^*$.

We segment $\boldsymbol{W}^*$ by N_p bits per sequence, correlate the obtained sequence with the original *PN*-sequence $\boldsymbol{p}$. The obtained correlation value is regarded as the soft decision value and is inputted to the log-MAP decoder for Turbo code [8]. The message can thus be recovered.

To demonstrate the effectiveness of adaptive receiving, we show some of the test results with the bit error rate (BER) of the extracted binary sequences $\boldsymbol{W}^*$ (the extracted watermark before decoding) and the recovered 60-bit message (the extracted watermark after decoding) with adaptive receiving or non-adaptive receiving in Table 2. We can see that adaptive receiving reduce the bit error rate (BER)

Table 2 The test results of the extrated watermark before and after decoding with BER (%) with adaptive and non-adaptive receiving.

	Baboon(*S*=120)				***Boat***(*S*=92)			
receiving method	adaptive		non-adaptive		adaptive		non-adaptive	
before or after decoding	before	after	before	after	before	after	before	after
2x2 median_filter	0.22	0	0.37	0.18	0.25	0	0.33	0.18
3x3 median_filter	0.28	0	0.68	0.73	0.29	0	0.48	0.45
5x5 median_filter	0.20	0	0.40	0.23	0.26	0	0.36	0.22
7x7 median_filter	0.25	0	0.28	0	0.29	0	0.34	0.20

of the extracted binary sequences $\boldsymbol{W}^*$ greatly, and thus can recover the hidden message with no error, especially in resisting median filtering.

4. EXPERIMENTAL RESULTS

We have tested the proposed watermarking algorithm on various images. The results on 512x512x8 Baboon, Lena and Boat images are reported here. A 60-bit watermark is embedded into each of the images. The PSNRs of marked images are larger than 41 dB. The watermarks are perceptually invisible (refer to Fig. 3). Table 3 shows the test results with our proposed algorithm using StirMark 4.0. In Table 3, "1" represents the embedded 60-bit watermark can be recovered with no error while "0" means the embedded message cannot be recovered correctly.

It is observed that our proposed technique can resist common signal processing such as JPEG compression with quality factor 12 to100, 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), sharpening very well. In particular, it can recover the watermark with no error after JPEG compression with quality factor less than or equal to 12, specifically, 8 for Baboon and 12 for Lena. We can also recover the watermark when the PSNR of the Gaussian noise corrupted image is merely 17.3 dB for Baboon and 20.2dB for Lena.

In order to combat the geometric distortion, similar to our work in [5], an additional template is embedded in the DFT domain of the obtained watermarked image $f'(x, y)$ mentioned above to recover global linear geometric transform. We recover translation by searching for the largest correlation coefficient between the original training sequence $\boldsymbol{T}$ and the extracted training sequence $\boldsymbol{T}^*$ from the DWT coefficients in LL_4 subband. However, the interval representing the hidden data bit "–1" is determined adaptively at first as we present above. In the extraction of the hidden information, we extract a data sequence $\boldsymbol{T}^*$ from the DWT coefficients in LL_4 subband of the test image $g(x, y)$, which is rescaled to the size of the original image at first (Note that we construct $\boldsymbol{T}^*$ sequence in the way as we embed the $\boldsymbol{T}$ sequence in the LL_4 subband). If, $\rho_{T,T^*} \geq thresh_1$ (refer to Section 2), we can then extract the informative watermark and recover the message from LL_4 subband directly.

(a) (b) (c)

Fig. 3. Watermarked (a) Baboon image (S=120, 41 dB), (b) Lena image (S=80, 44.5 dB), (c) Boat (S=92, 43.3dB)

Otherwise, that is, $\rho_{T,T^*} < thresh_1$, we need to resynchronize

Table 3. Experimental results with StirMark 4.0

StirMark functions	*Lena*	*Baboon*	*Boat*
JPEG_12~100	1	1	1
Rescale	1	1	1
Jitter	1	1	1
crop_75	1	1	1
RotationCrop	1	1	1
RotationScale	1	1	1
Affine	1	1	1
Gauss filtering	1	1	1
sharpening	1	1	1
Median filtering	1	1	1
SmallRandomDistortions	0	0	0

the hidden data [5] before extracting the informative watermark. In our work, we choose $thresh_1$=0.40 (empirically value). Here we can calculate the corresponding probability of false positive (false synchronization) corresponding to the threshold ($thresh_1$=0.40). When S is small, uniform distribution is a good approximation for r ($r = C^*(i) \bmod S$) [2]. For example, we choose S=80 for Lena in our work, while the mean of $C^*(i)$ for Lena is 1976.2. It is reasonable to consider the random appearance probability of bits "1" in the test sequence $\boldsymbol{T}^*$ is about 0.5, thus we have $H_{fp} = 0.5^{N_T} \cdot \sum_{k=N_T-e}^{N_T} \binom{N_T}{k} = 3.8\times10^{-7}$, which may be sufficiently low for many applications. Here $e = round\ (\frac{N_T}{2}(1 - thresh_1))$. For example, for 3x3 median filtered marked Baboon or JPEG_8 compressed marked Baboon, we can obtain ρ_{T,T^*}=0.65 or 0.42 respectively, so we can extract the informative watermark and recover the message directly.

With the blind resynchronization mentioned above, we evaluate its robustness to geometric distortion with StirMark 4.0 (Table 3). It is observed that the watermark is robust against almost all of geometrical distortion related test function, specifically, rotation (auto-crop), rotation (auto-crop, auto-scale), rescaling, jitter (random removal of rows and/or columns) and affine transform. The watermark is recovered when up to 65% of the image has been cropped.

5. CONCLUSIONS AND DISCUSSIONS

The main contributions reported in this paper are as follows.

i) By adaptive receiving using a training sequence, we model the data extraction as channel of additive noise with non-zero mean due to fading. This seems particularly promising in enhancing robustness of watermarking against median filtering (Table 2) and other distortion having fading nature. To our best knowledge, the difficulty of watermark robustness against median filtering and the success in overcoming this difficulty have not been reported in the literature.

ii) We propose a watermarking algorithm with enhanced robustness by applying adaptive receiving and Turbo code in addition to 2-D interleaving technique, resynchronization technique, the concatenated coding of Turbo code and DSSS, and the embedding strategy in the LL subband in DWT domain.

The watermark with the proposed algorithm is robust to common signal processing including Gaussian filtering, mean filtering, median filtering, sharpening, and JPEG compression with quality as low as 12.

Compared with the other existing watermarking [e.g. 2, 3, 5], the main difference is that we model the watermarking channel as channel of additive noise with non-zero mean due to fading, and the main improvement is the enhanced robustness against common signal processing including median filtering.

The idea of adaptive receiving can be combined with many existing watermarking schemes to enhance their robustness.

6. REFERENCES

[1] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, 6(2):1837-1687, 1997.

[2] M. Wu, "Joint security and robustness enhancement for quantization based data embedding," *IEEE Trans. On Circuits and Systems for Video Technology*, vol.13, no. 8, pp.831-841, Aug. 2003.

[3] 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.

[4] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *Journal of VLSI Signal Processing*, vol. 27, pp.7-33, 2001.

[5] 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.

[6] S. Voloshynovskiy, F. Deguillaume, S. Pereira, and T. Pun, "Optimal adaptive diversity watermarking with channel state estimation," *Proc. of SPIE: Security and watermarking of Multimedia content III*, vol.4314, pp. 673-685, San Jose, CA, USA, 22-25, Jan. 2001.

[7] Y. Q. Shi and X. M. Zhang, "A new two-dimensional interleaving technique using successive packing," *IEEE Trans. on Circuits and Systems, Part I*, vol. 49, no. 6, pp. 779-789, June 2002.

[8] 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.