Robust Blind Image Watermarking Using an Adaptive Trimmed Mean Operator

Hyun Lim*, Myung-Eun Lee*, Soon-Young Park*, Wan-Hyun Cho**

* Dept. of Electronics Engineering, Mokpo National University
** Dept. of Statistics, Chonnam National University
hlim,sypark@chungkye.mokpo.ac.kr, whcho@chonnam.ac.kr

abstract

In this paper, we present a robust watermarking technique based on a DCT-domain watermarking approach and an order statistic(OS) filter. The proposed technique inserts one watermark into each of four coefficients within a 2×2 block which is scanned on DCT coefficients in the zig-zag ordering from the medium frequency range. The detection algorithm uses an adaptive trimmed mean operator as a local estimator of the embedded watermark to obtain the desired robustness in the presence of additive Gaussian noise and JPEG compression attacks. The performance is analyzed through statistical analysis and numerical experiments. It is shown that the robustness properties against additive noise and JPEG compression attacks are more enhanced than the previous techniques.

1. Introduction

Different digital watermarks have been proposed for the protection of digital multimedia data with the rapid growth of network distributions of images and video. The embedding is performed in such a way that the watermark is perceptually invisible and robust to common signal processing and intentional attacks. To achieve desired robustness, many watermarking techniques are based on a frequency domain approach where the watermark is placed in a selected set of transformed coefficients of the image so that it is remained within the image even after simple signal distortions.

The set of coefficients for the watermark to be embedded usually comes from the medium range of the frequency spectrum to offer a compromise between perceptual invisibility and robustness. For example, Barni et al. proposed to insert the watermark in the zig-zag scanned coefficients whose lower parts were skipped[1]. However, Cox et al. described a method where the watermark was inserted in large DCT coefficients, excluding the DC term[2].

To recover the watermark, some watermarking techniques require the original unwatermarked image to be compared to the watermarked one[2]. Some other techniques which can detect the watermark without access to the original unwatermarked image have been recently proposed to resolve the dispute of rightful ownership[3],[4]. In deciding the presence of a given watermark from a test image, hypothesis testing is accomplished by comparing the correlation between the coefficients of the test image and the original watermark itself. If the correlation is lower than the predefined threshold, the decoder assumes the test image does not embed the watermark while the decoder declares the presence of the watermark in that test image if the correlation is above the threshold. Therefore two types of error probabilities are involved in this hypothesis test[5].

One is the false alarm probability occurred when the decoder detects the watermark which is not really present in the test image. The other one called the watermark missing probability occurs when the decoder cannot detect the watermark from the watermarked image. As the choice of a threshold contributes to the overall detection error probability, several researches have been focused on the optimal selection of threshold and their statistical robustness analysis[6],[7].

In this paper, we present a robust watermarking technique based on a DCT-domain watermarking approach[1],[2],[8] and an order statistic(OS) filter[9]. To enhance the robustness in the presence of additive Gaussian noise and JPEG lossy compression attacks, the proposed technique inserts one watermark into each of four coefficients within a 2×2 block which scans DCT coefficients of medium frequency range in zig-zag ordering.

Then the decoder can detect the embedded watermark efficiently by using an adaptive trimmed mean operator as a local estimator of a watermark from the possibly distorted coefficients within a window. The statistical analysis and numerical experiments show that the proposed technique is very efficient in the performance of robustness.

2. The Proposed Robust watermarking algorithm

Fig. 1 shows the block diagram of the proposed

watermark embedding and detection algorithms which are robust to additive Gaussian noise and JPEG lossy compression attacks by using adaptive trimmed mean operation. The basic idea such as superimposing Gaussian random variables to DCT coefficients in the full frame medium DCT range and thresholding the correlation between the coefficients of the test image and the original watermark itself are similar to the previous method[1]. However, the major different approach of our algorithm is the use of 2×2 block based embedding and adaptive trimmed mean operator for the extraction of embedded watermarks. An adaptive trimmed mean operator was developed as a noise smoothing filter which could provide robustness against a wide range of noise possibilities[9]. The motivation for the use of trimmed mean operation in watermark detection process is to reject the possibly corrupted coefficients whose magnitude are so small that they could be vulnerable to JPEG compression while smoothing the additive white noise.

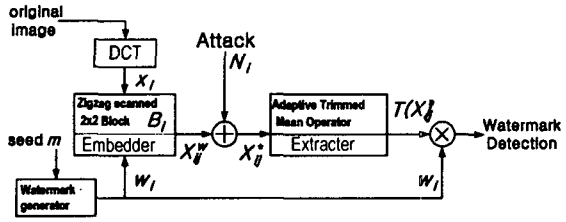

Fig. 1. A block diagram of a proposed watermarking system.

2.1 Watermark Embedding

The watermark signal is obtained by a pseudo random sequence of m real numbers that follows a normal distribution with zero mean and unit variance :

$$W=\{w_1,\cdots,w_m\},\ w_i\sim N(0,1). \tag{1}$$

We compute the $N\times N$ DCT coefficients of an gray scale image I and form the size m of 2×2 block DCT coefficients sequence by selecting a zig-zag ordering in medium frequency range of the 2-dimensional DCT-domain. This is given as

$$X=\{B_1,\cdots,B_m\} \text{ and } B_i=\{x_{i1},x_{i2},x_{i3},x_{i4}\}, \tag{2}$$

where x_{ik} is a DCT coefficient.

The watermark embedding coefficients is obtained by inserting the same watermark into each of four coefficients within a 2×2 block unlike the method in [1] where each different watermark is inserted into DCT coefficients.

$$x_{ik}^{w}=x_{ik}+\alpha\,|\,x_{ik}\,|\,w_i,\ i=1,\cdots,m,\ k=1,\cdots,4. \tag{3}$$

The positive parameter α denotes alternation strength. In order to guarantee watermark invisibility, this quantity should be restricted by some maximum value that depend on the image characteristics.

Here inserting the watermarking DCT coefficients in the zig-zag scan and applying the inverse DCT, the watermarked image I^{w} is obtained.

2.2 Watermark Detection

Given a possibly corrupted image I^{*}, the $N\times N$ DCT transform is applied to this image. Selecting the size m of 2×2 block DCT coefficients sequence by a zig-zag ordering in medium frequency range of the DCT transform domain gives us the test image $X^{*}=\{B_1^{*},\cdots,B_m^{*}\}$. Then, detection is based on the correlation between the watermark W and a test image X^{*} :

$$R=\frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{N}\sum_{k=1}^{4}T(x_{ik}^{*})w_i\right\}, \tag{4}$$

where $T(x_{ik}^{*})$ is an adaptive trimmed mean operator where the coefficients with smaller magnitude than the adaptive threshold are trimmed away and the retained coefficients with size N are averaged. Here the adaptive threshold is based on the local statistics of coefficients within a window.

3. Theoretical analysis

In this section, some statistical properties of the proposed correlation detector R are derived. Both the non-attacked watermarked image and the attacked watermarked image are considered, and also the expected value and variance of detector are estimated.

3.1 Non-attacked Watermarked Image

Let $W=\{w_1,\cdots,w_m\}$ be a watermark, and $X^{*}=\{x_{11}^{*},x_{12}^{*},\cdots,x_{m3}^{*},x_{m4}^{*}\}$ a set of DCT coefficients marked with W. We assume that the true DCT coefficients x_{ik} are independent and symmetrical distributed random variables with zero mean. To avoid complexity with respect to the trimmed operation, we restrict our analysis to the case where all coefficients within a block are retained.

Then the mean of the proposed detector R is

$$E(R)=E\left[\frac{1}{4m}\sum_{i=1}^{m}\sum_{k=1}^{4}x_{ik}^{*}w_i\right]=\frac{\alpha}{4m}\sum_{i=1}^{m}\sum_{k=1}^{4}E[\,|\,x_{ik}\,|\,] \tag{5}$$

To calculate the variance of R, we start by computing its mean square value:

$$E[R^2] = \frac{1}{(4m)^2} E\left[(\sum_{i=1}^{m}\sum_{k=1}^{4} x_{ik}^{*} w_i) \times (\sum_{j=1}^{m}\sum_{l=1}^{4} x_{jl}^{*} w_j) \right]$$
$$= \frac{1}{(4m)^2} \sum_{i=1}^{m}\sum_{k=1}^{4} E(x_{ik}^2) + \frac{3\alpha^2}{(4m)^2} \sum_{i=1}^{m}\sum_{k,l=1}^{4} E(|x_{ik}||x_{il}|) \quad (6)$$
$$+ \frac{\alpha^2}{(4m)^2} \sum_{i\neq j}^{m}\sum_{k,l=1}^{4} E(|x_{ik}||x_{jl}|)$$

where we have used the fact that $E(w_i^3) = 0$ and $E(w_i^4) = 3$.

The variance of R can now be calculated by means of the formula

$$\sigma_R^2 = E[R^2] - (E[R])^2 . \quad (7)$$

Hence, this is given as:

$$V(R) = \frac{1}{(4m)^2} (\sum_{i=1}^{m}\sum_{k=1}^{4} E(x_{ik}^2)$$
$$+ \frac{3\alpha^2}{(4m)^2} \sum_{i=1}^{m}\sum_{k,l=1}^{4} E(|x_{ik}||x_{il}|) \quad (8)$$
$$- \frac{\alpha^2}{(4m)^2} \sum_{i=1}^{m}\sum_{k,l=1}^{4} E(|x_{ik}|)E(|x_{jl}|).$$

3.2 Gaussian Noise Attacked Watermark Image

Let us consider that W be a watermark and X^* be set of DCT coefficients marked with W. We can apply the various attacks to the DCT domain, but we only think the additive Gaussian noise. When this kind of attack assume, the attacked watermarked signal is

$$x_{ik}^{a^*} = x_{ik}^{*} + \varepsilon_{ik}, \varepsilon_{ik} \sim N(0,\sigma^2), \quad (9)$$

and the correlation detector is given as

$$R = \frac{1}{4m} \sum_{i=1}^{m}\sum_{k=1}^{4} w_i (x_{ik}^{*} + \varepsilon_{ik}) . \quad (10)$$

First, if we compute the mean of the detector R, we have that

$$E(R) = \frac{\alpha}{4m} \sum_{i=1}^{m}\sum_{k=1}^{4} E[|x_{ik}|]. \quad (11)$$

This quantity is same as the case of the non-attacked image. But the variance of the detector R appears with the increase of $\frac{\sigma_\varepsilon^2}{m}$ than non-attacked one in Eq. (8) owing to the averaging operation during detection process while the noise attack increases the variance of detector with additional term σ_ε^2 in the previous method[1],[10].

4. Experiments

To evaluate the proposed watermarking technique, we applied a full-frame DCT transform to the 256×256 Lena image. We first verified the distribution of correlation statistics for two different hypothesis conditions, H_0 (the test image does not contain the watermark) and H_1 (the test image contains the watermark). For H_0, we generated 5000 different sequences of $N(0,1)$ using different seeds, and then correlated them with the original image. For H_1, we embedded 5000 different sequences to the image and correlated them with the known watermark.

Fig. 2 depicts results for the detection of watermark for the one coefficient based embedding(CBE) used in the method[1], our block based embedding with trimmed mean operator(BBE), and block based embedding without trimmed mean operator(BBE4).

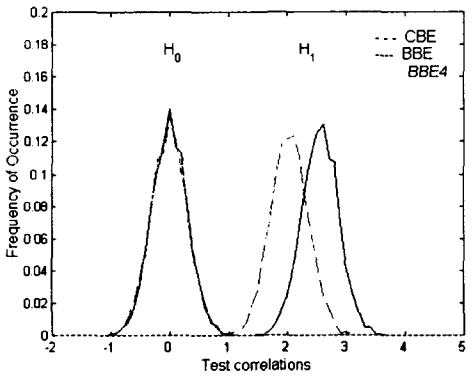

Fig. 2. Distributions of correlation statistic for Lena image.

It is obvious that the our watermarking technique is much easier to distinguish between H_0 and H_1 than the previous method. As expected, we can observe that the block-based watermarking without trimmed operation(B BE4) shows much smaller variance due to averaging of four coefficients and the one with trimmed operation(BBE) shows more distinct distribution by rejecting coefficients with smallest magnitude. This results indicate that the detection scheme with trimmed mean operator can detect the watermark with high confidence.

Next, the experiments were conducted for JPEG compression and additive Gaussian noise attacks.

Fig. 3 shows its distribution results only for two cases when the watermarked image is compressed with a quality factor of 50% and the watermarked image is corrupted by additive Gaussian noise with variance 400. The distributions of correlation for two types of attacks became more difficult in distinguishing H_1 from H_0 than the non-attacked watermarked image. However, in comparing Fig. 3 with Fig. 4, the proposed technique appears with more separated mean values or smaller variances so that

its probability error associated with two hypotheses is less than the previous method.

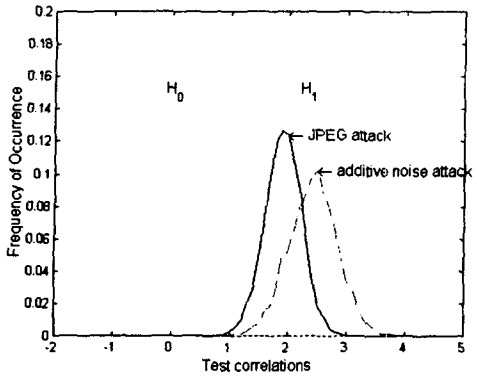

Fig. 3. The results of attacks for our watermarking technique.

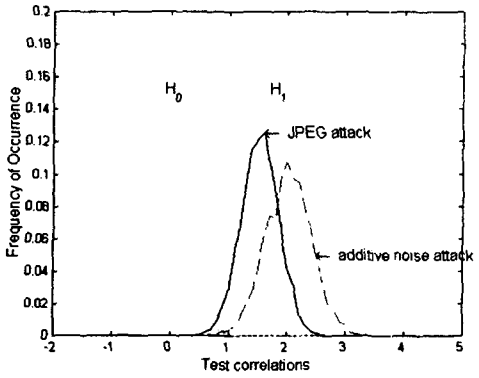

Fig. 4. The results of attacks for the previous method.

Next, we applied several JPEG compression as shown in Fig. 5. The mean values of correlation using the proposed technique are much larger than those using the previous method due to trimming operation for the coefficients with small magnitude which are very sensitive to JPEG compression.

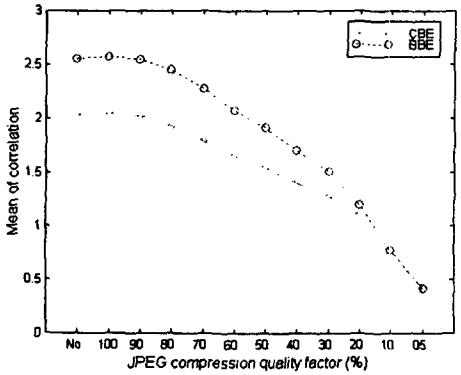

Fig. 5. Comparison for JPEG compression quality(%).

5. Conclusion

In this paper, we present a robust watermarking technique for digital images. The proposed watermarking scheme consists of block-based watermark embedding and detection algorithm with adaptive trimmed mean operation. Watermark is embedded into each of four coefficients within a 2×2 block which is scanned on a full-frame DCT coefficients in the zig-zag ordering from the medium frequency range. Then the detection algorithm can detect the embedded watermark efficiently by using an adaptive trimmed mean operator as a local estimator of a watermark from the possibly distorted coefficients within a window. The mean and variance of correlation between watermarked image and watermark are statistically analyzed and numerically verified. The results show that the proposed scheme can detect the watermark with high confidence under JPEG lossy compression and additive Gaussian noise attacks owing to adaptive trimmed mean operation.

References

[1] M. Barni, F. Bartolini, V. Cappellini, A. Piva, Robust Watermarking of Still Images for Copyright Protection. *Proceedings 13th International Conference on Digital Signal Processing DSP97*, Santorini, Greece, July 2-4, 1997, vol. 2, pp. 499-502.

[2] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, Secure Spread Spectrum Watermarking for Images, Audio and Video. Proc. *ICIP'96*, Vol III, pp. 243-246, 1996.

[3] W. Zeng and B. Liu, A Statistical Watermark Detection Technique without Using Original Images for Resolving Rightful Ownerships of Digital Images. *IEEE Trans. on Image Processing*, vol. 8, no. 11, Nov. 1999.

[4] J. J. Eggers, J. K. Su, B. Girod, Robustness of a Blind Image Watermarking Scheme. *ICIP2000 Special Session on WM*, Vancouver, Canada, Sep. 2000.

[5] A. Papoulis, *Probability \& Statistics*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[6] N. Nikolaidis, I. Pitas, Digital Image Watermarking: An Overview. *ICMCS'99*, vol. 1, pp. 1-6, July 1999.

[7] G. Voyatzis and I. Pitas, Image Watermarking for Copyright Protection and Authentication. Academic Press, 2000.

[8] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Addison Wesley, 1993.

[9] A. Restrepo and A. C. Bovik, Adaptive Trimmed Mean Filters for Images Restoration. *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1326-1337, 1988.

[10] A. Piva, *A DCT-Domain Watermarking System for Copyright Protection of Digital Images*. Ph. D. Thesis, 1998.