Multimedia Watermark Detection Algorithm Based on Bayes Decision Theorys

Seong-Geun Kwon, [1] Suk-Hwan Lee, [1] Kee-Koo Kwon, [1] Ki-Ryong Kwon, [2] and Kuhn-Il Lee[1]
[1] School of Electrical Engineering and Computer Science, Kyungpook National University
1370, Sankyug-Dong, Buk-Gu, Daegu 702-701, Republic of Korea
Tel: +82-53-950-5512, Fax: +82-53-950-5505
[2] Department of Electronic Engineering, Pusan University of Foreign Studies
55-1 Uam-Dong, Nam-Gu, Busan, Republic of Korea
e-mail : sgkwon@m80.knu.ac.kr

Abstract: Watermark detection plays a crucial role in multimedia copyright protection and has traditionally been tackled using correlation-based algorithms. However, correlation-based detection is not actually the best choice, as it does not utilize the distributional characteristics of the image being marked. Accordingly, an efficient watermark detection scheme for DWT coefficients is proposed as optimal for non-additive schemes. Based on the statistical decision theory, the proposed method is derived according to Bayes' decision theory, the Neyman-Pearson criterion, and the distribution of the DWT coefficients, thereby minimizing the missed detection probability subject to a given false alarm probability. The proposed method was tested in the context of robustness, and the results confirmed the superiority of the proposed technique over conventional correlation-based detection method.

1. Introduction

For multimedia copyright protection, a great deal of research has been carried out in the field of digital watermarking, which is accomplished by inserting a watermark within a host image to convey ownership information.

Watermarking algorithms can be distinguished according to the embedding domain. Generally, spatial [1],[2], frequency [3],[4], and hybrid [5],[6] domain watermarking are all included in the first category. Hybrid domain techniques operate in a space-scale domain, like a DWT domain.

Although the requirements of a watermarking system are relative to the application, a highly reliable detection of the embedded watermark is needed for all applications. The correlation-based detection scheme [7] is based on a correlation between the watermark and the image features. Due to its simplicity it has been widely used, however, its performance is still unsatisfactory as it does not utilize the characteristics of the host image being marked.

Accordingly, the current paper derives a new detection algorithm for non-additive embedding in the DWT domain. Since a DWT can efficiently manipulate a non-stationary signal, it is extensively exploited in the watermarking area. Based on the statistical decision theory, the proposed detection scheme verifies whether a given watermark is present in an image using Bayes' decision theory [8], the Neyman-Pearson criterion [9], and the pdf of the DWT coefficients. The pdf of the DWT coefficients is estimated as a Gaussian distribution and applied to the likelihood ratio based on Bayes' decision theory. The value of the likelihood ratio is compared with a decision threshold derived using the Neyman-Pearson criterion.

The validity of the proposed watermark detection algorithm is confirmed in a computer experiment by comparing its robustness with the conventional correlation-based scheme.

2. Proposed detection algorithm

Non-additive watermark embedding rule,

$$y_i = x_i + \alpha w_i x_i \tag{1}$$

is adopted where α and w_i are the embedding strength and a watermark component, respectively, and x_i is the original DWT coefficient. The watermark space W can be defined as $W_1 \cup W_0$, where W_0 and W_1 mean $\{w \neq w^*\}$ and $\{w = w^*\}$, respectively. Based on Bayes' theory, the decision rule $\delta(y)$ maps each y into 0 or 1 such as,

$$\delta(y) = \begin{cases} 1, & y : \dfrac{f_y(y \mid W_1)}{f_y(y \mid W_0)} > \lambda \\ 0, & \text{otherwise} \end{cases}, \tag{2}$$

where $f_y(y \mid W)$ is the conditional pdf and λ is the threshold. Here, $f_y(y \mid W_1)/f_y(y \mid W_0)$ is called the likelihood ratio $l(y)$.

Under the assumption that the watermark components are independent and uniformly distributed in [-1, 1], and each y_i is independent, $f_w(w)$, $f_y(y|W_0)$, and $l(y)$ can be written as

$$f_w(w) = \prod_{i=1}^{N} f_{w_i}(w_i) = \frac{1}{2^N}, \tag{3}$$

$$f_y(y \mid W_0) = \int_{[-1,1]^N} f_y(y \mid w) f_w(w) dw, \tag{4}$$

$$l(y) = \frac{\prod_{i=1}^{N} f_{y_i}(y_i \mid w_i^*)}{\frac{1}{2^N} \prod_{i=1}^{N} \int_{-1}^{1} f_{y_i}(y_i \mid w_i) dw_i}. \tag{5}$$

The pdf of the DWT coefficients is needed, which must be distributed on both the positive and negative axis, and easy to represent mathematically. Therefore, a Gaussian distribution is adopted as the pdf of the DWT coefficients:

$$f_{x_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x_i - m_i)^2}{2\sigma_i^2}\right], \quad (6)$$

where m_i and σ_i are the mean and variance of the subband to which each coefficient belongs. Therefore, $f_{y_i}(y_i \mid w_i)$ is

$$\frac{1}{(1+\alpha w_i)\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{\{y_i/(1+\alpha w_i) - m_i\}^2}{2\sigma_i^2}\right]. \quad (7)$$

As a result, the numerator of $l(\boldsymbol{y})$ is

$$\prod_{i=1}^{N} \frac{1}{(1+\alpha w_i^*)\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{\{y_i/(1+\alpha w_i^*) - m_i\}^2}{2\sigma_i^2}\right], \quad (8)$$

and the denominator of $l(\boldsymbol{y})$ is

$$\prod_{i=1}^{N} \int_{-1}^{1} \frac{1}{2(1+\alpha w_i)\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{\{y_i/(1+\alpha w_i) - m_i\}^2}{2\sigma_i^2}\right] dw_i \,. \quad (9)$$

By letting $t_i = y_i/(1+\alpha w_i)$, (9) can be simply written as follows:

$$\prod_{i=1}^{N} \frac{1}{2\alpha\sqrt{2\pi\sigma_i^2}} \int_{y_i/(1+\alpha)}^{y_i/(1-\alpha)} \frac{1}{t_i} \exp\left[-\frac{(t_i - m_i)^2}{2\sigma_i^2}\right] dt_i \,. \quad (10)$$

If it is assumed $\alpha << 1$, the integral interval of (10) is very small and centered at y_i so that the component to be integrated can be linearly approximated by Taylor's theorem as follows:

$$\prod_{i=1}^{N} \frac{1}{2\alpha\sqrt{2\pi\sigma_i^2}} \int_{y_i/(1+\alpha)}^{y_i/(1-\alpha)} \left(\frac{1}{y_i} \exp\left[-\frac{(y_i - m_i)^2}{2\sigma_i^2}\right] + \mu_0(t_i - y_i)\right) dt_i \,. (11)$$

The fact that the integration result of these two terms of (11) is $\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(y_i - m_i)^2}{2\sigma_i^2}\right]$ and about zero yields

$$\frac{1}{2^N} \prod_{i=1}^{N} \int_{-1}^{1} f_{y_i}(y_i \mid w_i) dw_i \approx \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(y_i - m_i)^2}{2\sigma_i^2}\right]. \quad (12)$$

The likelihood ratio becomes

$$\prod_{i=1}^{N} \frac{1}{1+\alpha w_i^*} \exp\left[\frac{(y_i - m_i)^2}{2\sigma_i^2} - \frac{\{y/(1+\alpha w_i^*\} - m_i)^2}{2\sigma_i^2}\right]. \quad (13)$$

In addition, if the logarithm is taken to (13) for simplicity, then the log likelihood ratio $\ell(\boldsymbol{y})$ is

$$\sum_{i=1}^{N} \left[\frac{(y_i - m_i)^2}{2\sigma_i^2} - \frac{\{y_i/(1+\alpha w_i^*) - m_i\}^2}{2\sigma_i^2}\right] - \sum_{i=1}^{N} \ln(1+\alpha w_i^*), \quad (14)$$

and the decision rule can be given in the form

$$\ell(\boldsymbol{y}) > \ln\lambda \,. \quad (15)$$

Inserting (14) into (15) yields

$$\sum_{i=1}^{N} \left[\frac{(y_i - m_i)^2}{2\sigma_i^2} - \frac{\{y_i/(1+\alpha w_i^*) - m_i\}^2}{2\sigma_i^2}\right] > \ln\lambda + \sum_{i=1}^{N} \ln(1+\alpha w_i^*) \,. \quad (16)$$

The term on the left is the sufficient statistic for the decision test. By letting

$$v_i = \frac{1}{2\sigma_i^2}\left[(y_i - m_i)^2 - \left(\frac{y_i}{1+\alpha w_i^*} - m_i\right)^2\right], \quad (17)$$

$$\lambda_1 = \ln\lambda + \sum_{i=1}^{N} \ln(1+\alpha w_i^*), \quad (18)$$

the result test is defined by the inequality

$$z(\boldsymbol{y}) = \sum_{i=1}^{N} v_i > \lambda_1 \,. \quad (19)$$

When attacks are taken into account, a threshold selected to minimize the error probability P_e will produce bad results so that the missed detection probability P_{MD} becomes considerably larger than the false alarm probability P_{FA}. Therefore, to solve this problem, the Neyman-Pearson criterion is included in the selection of a decision threshold so that the threshold is chosen in such a way that P_{MD} is minimized subject to a given $\overline{P_{FA}}$. Among the thresholds satisfying the constraint $P_{FA} \le \overline{P_{FA}}$, the one minimizing P_{MD} was selected in the current study. The relation between $\overline{P_{FA}}$ and λ_1:

$$\overline{P_{FA}} = P(z(\boldsymbol{y}) > \lambda_1 \mid W_0) = \int_{\lambda_1}^{+\infty} f_{z(\boldsymbol{x})}(z(\boldsymbol{x})) dz(\boldsymbol{x}), \quad (20)$$

where $z(\boldsymbol{x})$ is represented by:

$$z(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{1}{2\sigma_i^2}\left[(x_i - m_i)^2 - \left(\frac{x_i}{1+\alpha w_i^*} - m_i\right)^2\right]. \quad (21)$$

Based on the central limit theorem, the pdf of $z(\boldsymbol{x})$ is assumed to be Gaussian and the mean and variance of $z(\boldsymbol{x})$ are

$$m_{z(x)} = \sum_{i=1}^{N} m_{v_i} = \sum_{i=1}^{N} [a_i(m_i^2 + \sigma_i^2) - b_i m_i], \quad (22)$$

$$\sigma_{z(x)}^2 = \sum_{i=1}^{N} \sigma_{v_i}^2 = \sum_{i=1}^{N} [\sigma_i(2a_i m_i - b_i)]^2, \quad (23)$$

respectively, where a_i and b_i are $\frac{\alpha w_i^*(\alpha w_i^* + 2)}{2\sigma_i^2(\alpha w_i^* + 1)^2}$ and $\frac{\alpha m_i w_i^*}{\sigma_i^2(\alpha w_i^* + 1)}$, respectively. In results, (20) is rewritten as

$$\overline{P_{FA}} = \int_{\lambda_1}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_{z(x)}^2}} \exp\left[-\frac{(z - m_{z(x)})^2}{2\sigma_{z(x)}^2}\right] dz(x)$$
$$= \frac{1}{2} erfc\left(\frac{\lambda_1 - m_{z(x)}}{\sqrt{2\sigma_{z(x)}^2}}\right), \quad (24)$$

if $\overline{P_{FA}}$ is 10^{-9}, then λ_1 is $4.24\sqrt{2\sigma_{z(x)}^2} + m_{z(x)}$.

As such, the proposed algorithm uses the pdf of the DWT coefficients as a Gaussian pdf and applies this pdf to the decision rule based on Bayes' decision rule. Also, to minimize the missed detection probability subject to a given significance level, the Neyman-Pearson criterion is included in the threshold selection. As a result, the proposed detection algorithm can overcome the disadvantages of the correlation-based scheme.

3. Experimental results

Experiments were conducted to confirm that the proposed detection algorithm was an improvement on the conventional correlation-based algorithm. In particular, LENA and CABLECAR with a pixel size of 512×512 were used as the test images and a biorthogonal filter was used for the DWT. For an objective comparison, the two detection algorithms were applied to the same marked images where that watermark had been embedded into the absolutely largest 5000 DWT coefficients that were not in the baseband or the highest frequency. Plus, α and $\overline{P_{FA}}$ were 0.3 and 10^{-9}, respectively.

To test the robustness, attacks involving common image processing, geometric transformations, and lossy compression were conducted. Each type of attack was applied several times, while increasing the strength of the attack. For each attack, the detector responses were related to the actually embedded watermark. In the following figures, the left axis represents the proposed method, while the right axis represents the correlation-based method.

3.1. Robustness against common image processing

The robustness of the watermark was tested against blurring and noise addition. Blurring was applied while increasing the size of the Gaussian window. Whereas, Gaussian noise with 256 gray levels was added randomly into the watermarked image with added region ranging from 10 to 90% of the image size. The results obtained are represented in Figs. 1 and 2.

3.2. Robustness against geometric transformation

Cropping and warping were conducted as geometric transformations and the results are represented in Figs. 3 and 4. The cropped image was composed of the watermarked image and original image as the center and a border, respectively. Warping was carried out with varying warp sizes ranging from 10 to 90% of the image size in the center area.

3.3. Robustness against lossy compression

Lossy compression is ultimately necessary to reduce the bit rate in the Internet environment. Therefore, JPEG compression was iteratively applied to the watermarked images, while decreasing the quality factor. The results are represented in Fig. 5.

4. Conclusions

To improve the performance of the conventional correlation-based algorithm, a new watermark detection algorithm was proposed relying on the statistical theory. The proposed algorithm utilizes the pdf of the DWT coefficients, Bayes' decision theory, and the Neyman-Pearson criterion to minimize the missed detection probability subject to a given false alarm probability. Experimental results verified the validity of the proposed algorithm and confirmed that the proposed algorithm could produce a higher detection rate than conventional correlation-based detection method.

References

[1] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," *Proc. of SPIE*, vol. 2420, pp. 164-173, Feb. 1995.

[2] I. Pitas, "A method for watermark casting on digital images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 6, pp. 775-780, Oct. 1998.

[3] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673-1687, 1997.

[4] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," *Proc. of ICIP*, vol. 1, pp. 520-527, Oct. 1997.

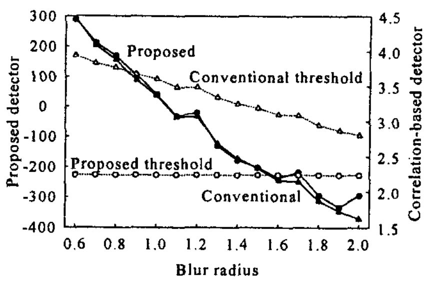
[5] X. Xia, C. Boncelet, and G. Arce, "A multiresolution watermark for digital images," *Proc. of ICIP*, vol. 1, pp. 548-551, Oct. 1997.

[6] W. Zhu, Z. Xiong, and Y. -Q. Zhang, "Multiresolution watermarking for images and video," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 9, no. 4, pp. 545-550, June 1999.

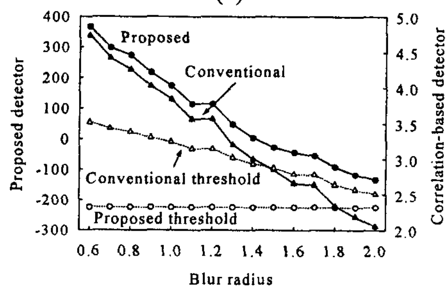
[7] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, vol. 66, pp. 357-372, May 1998.

[8] L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[9] J. Franco and W. Rubin, *Radar Detection.* Norwood, MA: Artech House, 1980.

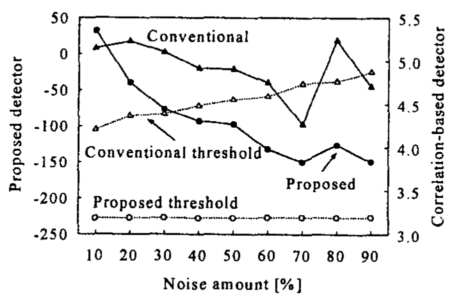


(a)

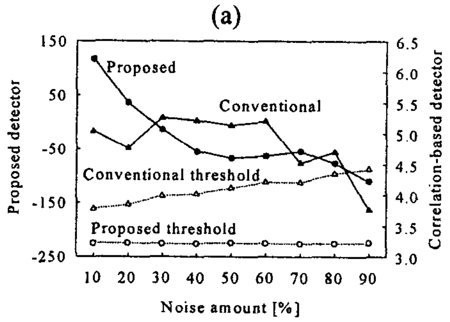


(b)

Fig. 1. Robustness of (a) LENA and (b) CABLECAR against blurring.

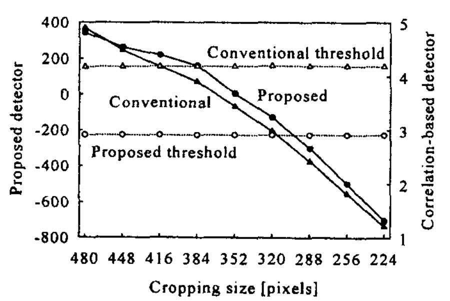


(a)

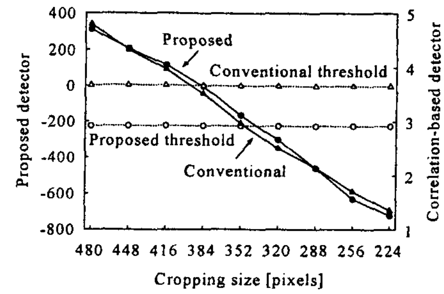


(b)

Fig. 2. Robustness of (a) LENA and (b) CABLECAR against noise addition.

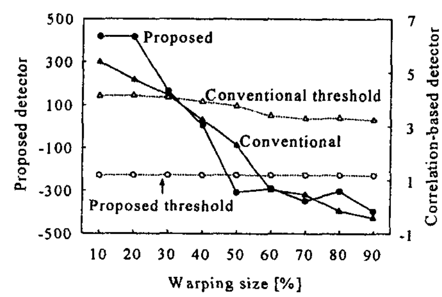


(a)

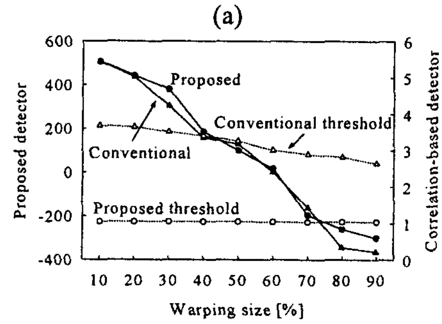


(b)

Fig. 3. Robustness of (a) LENA and (b) CABLECAR against cropping.

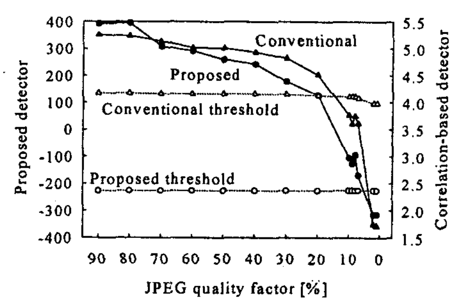


(a)

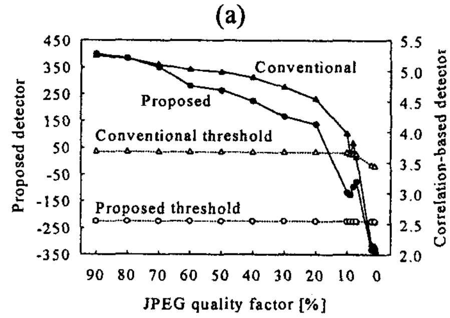


(b)

Fig. 4. Robustness of (a) LENA and (b) CABLECAR against warping.



(a)



(b)

Fig. 5. Robustness of (a) LENA and (b) CABLECAR against JPEG coding.