WUJNS
Wuhan University Journal of Natural Sciences

Vol. 10 No. 1 2005 279-283

Article ID: 1007-1202(2005)01-0279-05

Wavelet-Based Denoising Attack on Image Watermarking

□ **XUAN Jian-hui[1], WANG Li-na[1,2], ZHANG Huan-guo[1†]**

1. School of Computer, Wuhan University, Wuhan 430079, Hubei, China;
2. National Laboratory of Pattern Recognition, Chinese Academy of Science, Beijing 100080, China

Abstract: In this paper, we propose wavelet-based denoising attack methods on image watermarking in discrete cosine transform (DCT) or discrete Fourier transform (DFT) domain or discrete wavelet transform (DWT) domain. Wiener filtering based on wavelet transform is performed in approximation subband to remove DCT or DFT domain watermark, and adaptive wavelet soft thresholding is employed to remove the watermark resided in detail subbands of DWT domain.

Key words: watermarking; watermarking attack; wavelet denoising; Wiener filter

CLC number: TP 309

Received date: 2004-05-10

Foundation item: Supported by the National Natural Science Foundation of China (90104005, 60373087, 60473023)

Biography: XUAN Jian-hui (1977-), male, Ph. D candidate, research direction: digital watermarking, digital rights management. E-mail: xuanmian@etang. com

† To whom correspondence should be addressed. Email: liss@ whu. edu. cn.

0 Introduction

Digital watermarking was primally proposed as a means of copyright protection in 1990's, since then, the promising technique has become a hot research field. Although a large number of watermarking algorithms are proposed in recent years, so far almost all those algorithms have vulnerabilities to certain watermarking attacks. Security of watermarking has raised more and more attentions nowadays.

A variety of attacks on watermarking have been proposed to evaluate the security of the present watermarking algorithms. The existing attacks can be divided to four categories: removal attacks, geometrical attacks, cryptographic attacks, protocol attacks[1]. The removal attacks namely are intend to remove the watermark from the watermarked media completely. The category of attacks include denoising attacks, lossy compression, remodulation, quantization and collusion attacks[1]. The denoising attack is based on the assumption that the watermark can be modeled noise statistically. In fact, embedding watermark means slight modification on the original image, and the watermarked image can be thought of as the noisy image. Image denoising aims to remove the noise from the noisy image while preserving the characteristic components of the image. In recent years, wavelet-based thresholding as a new denoising method has a superior performance. Related to wavelet-based denoising attack on watermarking, Ref. [1] only gives theory analysis in terms of estimation but no practical strategy. This paper presents wavelet-based denoising attacks on existing spread spectrum (SS) watermarking schemes based on the transformation domain such as DCT and DWT.

This paper is organized as follow. In Section 1, existing wavelet-based denoising approaches are presented. Watermarking attacks based on wavelet denoising are proposed in Section 2. Section 3 gives the experiment results and discussion. A conclusion is drawn in Section 4.

1 Wavelet-Based Image Denoising

1.1 Image Denoising

An image is often corrupted by noise in its acquisition and transmission. The goal of denoising is to remove noise while retaining as much image information as possible. The formulation of image denoising as following:

$$X = S + N \tag{1}$$

$$\hat{S} = D(X) \tag{2}$$

Here X denotes observed noisy image, S and N denote original image and noise respectively. $D(\ \cdot\ )$ is a linear or non-linear function which can obtain an estimation $\hat{S}$ of original image from the noisy image.

1.2 Wavelet Thresholding

In recent years, there are a lot of researches on wavelet thresholding and threshold selection, in the setting of additive white Gaussian noise. Thresholding is a non-linear technique , replacing coefficients below a certain threshold by zero. Due to the energy compact property of wavelet transform, the wavelet domain becomes a sparse space after thresholding. After inverse wavelet transform, a noise-suppressed image is reconstructed.

Suppose W and W^{-1} denote wavelet transform and inverse wavelet transform respectively, and Y , T , t and C denote wavelet coefficients, thresholding operator, threshold and thresholded coefficients respectively. Then wavelet thresholding can be formulated as:

$$Y = WX \tag{3}$$

$$C = T(Y,t) \tag{4}$$

$$\hat{S} = W^{-1}C \tag{5}$$

Wavelet thresholding includes tow types: hard thresholding and soft thresholding. The hard thresholding can be defined as:

$$T(Y,t) = \begin{cases} Y, & |Y| \geqslant t \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

And the soft thresholding can be defined as:

$$T(Y,t) = \operatorname{sgn}(Y)\max(0,\ |Y| - t) \tag{7}$$

To select a appropriate threshold is a core task in wavelet thresholding and many literature like Ref. [2-6] are concerned with thresholding selection. There are several famous wavelet thresholding methods such as Visu Shink[3], SureShink[4] and BayesShink[5]. Dohono and Johnstone[4] proposed VisuShink with a universal threshold, $\sigma\sqrt{2\lg M}$, where σ is the noise standard deviation and M is the sum of pixels of an image. Obviously, in the setting of additive white Gaussian noise, the threshold depends on M. Due to the universal threshold, VisuShink is found to yield an overly smoothed estimate. SureShink proposed by Dohono and Johnstone is an subband-adaptive thresholding method by the principle of minimizing the Stein Unbiased Risk Estimate(SURE) for threshold estimate, and has been shown to perform better than VisuShink. BayesShink can provide a optimal estimate in a sense of Minimum Mean-Square Error (MMSE) and performs better than SureShink. BayesShink performs soft thresholding, with the data-driven, subband-dependent threshold $\hat{T}(\hat{\sigma}_X) = \hat{\sigma}^2/\hat{\sigma}_X$, where $\hat{\sigma}^2$ and $\hat{\sigma}_X$ are the estimates of noise variance and signal standard deviation , respectively.

2 Watermarking Attack Based on Wavelet Denoising

Watermarking in transform domains, which are dominantly DCT,DFT or DWT domain, usually has superior robustness than ones in spatial domain. Spread spectrum watermarking, where the watermark spreads through a large range of spectrum, is the most typical watermarking method. Here we aim to attack typical SS watermarking methods.

2.1 Attack on DCT-Based or DFT-Based Watermarking (ADW)

The DCT-based watermarking algorithms are essentially similar with DFT-based ones. Most existing DCT-based or DFT-based watermarking algorithms embed watermark in middle or low frequency.

Cox *et al*'s algorithm[7] is a classical DCT-based watermarking method, so here try to attack it as an example. Our objective is to decrease the correlation value to below a certain threshold while preserving the acceptably perceptual quality. In Cox *et al*'s method[7], an independent and identically distributed (i. i. d.) Gaussian random vector is embedded in the most biggest coefficients of DCT or DFT domain, which performs well in robustness against common image processing, such as lossy compression, filtering , requantization and so on. Because the biggest coefficients in DCT domain correspond to the smooth region of an image, which covers most of the energy of an image, so in the case of Cox *et al*'s algorithm, watermark can survive low-pass filtering. If wavelet thresholding is performed in high-pass subbands of an watermarked image, the watermark will survive the processing, because the watermark resides in low-pass components. However, if a filtering is performed in the approximation subband under a certain wavelet decomposition level, the watermark should be removed partly or mostly depending on the decomposition level. We choose the Wiener filter because it is the optimal one in a sense of MMSE and performs well when noise is weaker. Of course, other filter, such as mean filter and median filter can also be adopted, but Wiener filter could be more suitable due to its adaptability. To summarize, the attack consists of three main steps:

1) Wavelet transform of the watermarked image;

2) Wiener filtering in approximation subband, preserving detail subbands unchanged;

3) Inverse wavelet transform to reconstruct the estimated image.

To the attacked images, the lower the correlation between the extracted watermark and the original watermark is, the worse the visual quality is. To get the tradeoff between visual quality and correlation, the wavelet decomposition level and window size of Wiener filter need be selected through experiments.

2.2 Attack on Wavelet-Based Watermarking (AWW)

Many wavelet-based watermarking methods were proposed due to the multi-resolution property of wavelet decomposition. Wavelet-based watermarking methods can be divided into two categories: the approximation image methods and the detail subband methods[8].

In the first category of methods, an approximation image is treated as a general image and the watermark is embedded as usual in spatial domain or DCT or DFT domain. To attack this category of methods, the same steps as performed in Section 2.1 are needed. In the second category of methods, watermark is embedded in detail subbands. Those methods differ in the selection of the coefficients and the subbands. One method is to select the biggest coefficients in detail subbands, and another one is to select those coefficients above a certain threshold. Compared to the noisy image scenario, the watermark in detail subbands is like the high-frequency noise. So we can employ wavelet soft thresholding to remove watermark within the constraint of acceptable visual quality because the soft thresholding affects all coefficients in a certain subband and usually performs well than hard thresholding. The attack involves three main steps:

1) Wavelet transform of the watermarked image;

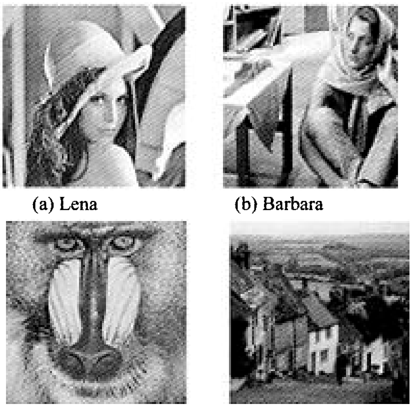
2) Subbands-adaptive soft thresholding in detail subbands, preserving approximation subband unchanged;

3) Inverse wavelet transform to reconstruct the estimated image.

To balance the visual quality and the correlation between the extracted watermark and the original watermark, the wavelet decomposition level need be selected through experiments.

3 Experiment Result and Discussion

The 512×512 grayscale images "Lena","Baboon", "Barbara" and "Goldhill" (as Fig. 1) were used to test our attack methods.



(a) Lena (b) Barbara

(c) Baboon (d) Goldhill

Fig. 1 Experimental images

In order to test the effect of the ADW, Cox *et al*'s algorithm[7] was implemented in DCT domain by MATLAB, where the strength factor $\alpha=0.1$, the length of watermark sequence was 1000 and the correlation threshold is 6. The Daubechies wavelet with four vanishing moments (i. e. db. 4) was adopted to implement the ADW. The MATLAB function wiener2 (I, [m n]), where m and n are window size, was used as Wiener filter. The Peak Signal Noise Ratio (PSNR) was computed by Eq. (8) to measure the image quality, where f_{ij}, F_{ij} and N denote the pixel intensity of input image, that of the output image and the size of image, respectively. The PSNR (dB) and the correlation between the extracted watermark and the original watermark at different levels of wavelet decomposition and size of Wiener filter window are given from Table 1 to Table 3.

$$\mathrm{PSNR} = 20\lg \frac{255}{\sqrt{(\sum_{i,j=0}^{N-1}(f_{ij} - F_{ij})^2)/N^2}} \qquad (8)$$

In order to test the effect of the AWW, two DWT-based watermarking approaches [9,10] were implemented by MATLAB. In the approach of Ref. [9], the watermark is embedded in detail subbands, while in the approach of Ref. [10], the watermark is embedded in the approximation image in DFT domain similar to Cox *et al*'s approach[7]. In both approaches, the parameters were set as follows: the strength factor $\alpha = 0.1$, the length of watermark sequence was 1 000 and the correlation threshold was 6 as Ref. [7]. The Daubechies wavelet with four vanishing moments (i. e. db. 4) was adopted to implement the AWW. The BayesShink was implemented by MATLAB to carry out denoising attack. The PSNR and correlation are shown in Table 4.

Table 1 The result of ADW with 2×2 Wiener filter window

Level of wavelet decomposition	Correlation				PSNR /dB			
	Lena	Barbara	Baboon	Goldhill	Lena	Barbara	Baboon	Goldhill
1	22.71	22.51	20.64	23.43	34.44	31.55	27.08	33.51
2	13.92	13.72	12.01	12.03	31.35	30.30	27.66	31.56
3	6.67	6.00	7.77	4.04	28.18	26.89	27.97	29.58
4	3.67	4.79	7.45	5.06	25.16	23.63	27.25	27.38
5	4.90	5.73	6.82	7.99	22.39	21.62	24.84	24.83
6	6.80	8.87	9.84	10.76	21.34	21.83	22.58	22.97

Table 2 The result of ADW with 3×3 Wiener filter window

Level of wavelet decomposition	Correlation				PSNR /dB			
	Lena	Barbara	Baboon	Goldhill	Lena	Barbara	Baboon	Goldhill
1	28.45	28.60	19.68	25.60	35.56	30.61	25.78	32.34
2	13.71	14.73	8.51	10.30	31.78	30.45	26.33	30.80
3	5.07	3.73	4.45	2.37	28.28	26.48	26.97	27.84
4	2.76	3.06	4.55	3.79	24.88	22.78	25.99	25.56
5	6.76	4.56	5.08	5.90	21.57	20.38	23.48	23.91
6	8.72	8.18	7.80	9.68	20.66	20.59	19.68	21.96

Table 3 The result of ADW with 4×4 Wiener filter window

Level of wavelet decomposition	Correlation				PSNR /dB			
	Lena	Barbara	Baboon	Goldhill	Lena	Barbara	Baboon	Goldhill
1	19.21	19.4	13.04	16.03	32.52	29.07	24.34	30.28
2	7.61	7.27	5.55	4.86	29.06	27.64	24.90	27.99
3	2.74	1.95	3.37	1.11	25.49	23.64	25.28	25.71
4	1.63	3.00	3.46	3.47	22.02	20.36	23.76	23.54
5	5.63	4.32	5.51	5.59	19.43	18.57	20.49	21.15
6	8.09	7.78	7.34	9.10	19.37	17.92	18.61	19.66

Table 4 The result of AWW

Level of wavelet decomposition	Correlation				PSNR /dB			
	Lena	Barbara	Baboon	Goldhill	Lena	Barbara	Baboon	Goldhill
1	31.96	5.95	26.89	28.64	35.65	25.52	22.58	31.87
2	6.08	3.35	7.72	8.48	29.40	23.34	19.37	27.77
3	−0.58	1.05	1.91	2.46	23.33	21.74	18.32	25.02
4	−0.58	1.00	1.83	2.46	22.42	19.67	17.54	22.46
5	−0.53	1.04	1.82	2.44	19.70	17.63	16.67	20.73
6	−0.52	1.01	1.82	2.43	17.26	15.74	15.61	19.52

From Table 1 to Table 3, it can be observed that the ADW is effective, while the level of wavelet decomposition should not be too high. On the one hand, too high decomposition level will result in the increasing of correlation, instead of decreasing, because the approximation subband can not contain most energy of the image if its size is too little. On the other hand, the higher the decomposition level is, the lower the PSNR is. The size of Wiener filter window also has significant impact on the correlation and PNSR.

From Table 4, it is observed that the AWW can effectively remove the watermark in detail subbands. The result of attack on the watermarking approach of Ref. [10] is very close to that of attack on Cox *et al*'s method[7]. In fact, the method of Ref. [10] is essentially similar with Cox *et al*'s method[7], but the frequency in which the former embeds watermark is lower than that in which the latter does.

The parts of experimental result images are given in Fig. 2. The result image of ADW preserves good details, but there are some changes in smooth region, the reason is that low frequency is filtered but high frequency is unchanged. On the contrary, the image denoised by BayesShink shows that there are blur and ringing artifacts in the neighborhood of image edges, it is a inherent result of wavelet thresholding like other transform-based filters. A question encountered here is that if all the subbands are inserted watermark, each individual attack method above can't remove watermark completely. To solve the question, both methods above must be performed orderly. As a result, the result image could have poorer visual quality.

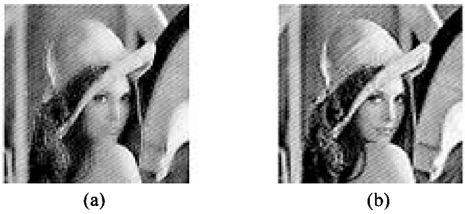


Fig. 2 Attack results

(a) The result of ADW(three-level decomposition, 3×3 Wiener filter window); (b) The result denoised by BayesShink(three-level decomposition)

4 Conclusion

Wavelet-based thresholding is employed to remove watermark, which is based on the idea that the watermark can be considered as additive noise. Experiment results show that wavelet-based Wiener filtering is considerably effective to remove watermark in DCT or DFT domain, wavelet denoising can remove the watermark resided in wavelet detail subbands. But there are artifacts in attack result images, to improve the image quality, other improved wavelet-based denoising methods need to be considered.

References

[1] Voloshynovskiy S, Pereira S, Iquise V, *et al*. Attack Modelling: Towards a Second Generation Benchmark. *Signal Processing*, 2001, **81**(6):1177-1214.

[2] Donoho D L. De-Noising by Soft-Thresholding. *IEEE Trans Inform Theory*, 1995,**41**(3):613-627.

[3] Donoho D L, Johnstone I M. Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika*, 1994, **81**:425-455.

[4] Donoho D L, Johnstone I M. Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Assoc*, 1995, **90**(432):1200-1224.

[5] Chang S G, Yu B, Vetterli M. Adaptive Wavelet Thresholding for Image Denoising and Compression. *IEEE Transactions on Image Processing*, 2000, **9**(9):1532-1546.

[6] Scharcanski J, Jung C R, Clarke R T. Adaptive Image Denoising Using Scale and Consistency. *IEEE Transactions on Image Processing*, 2002,**11**(9):1092-1101.

[7] Cox I J, Kilian J, Leighton T, *et al*. Secure Spread Spectrum Watermarking for Multimedia. *IEEE Transactions on Image Processing*, 1997,**6**(12): 1673-1687.

[8] Meerwald P, Uhl A. A Survey of Wavelet-Domain Watermarking Algorithms. *http://www.cosy.sbg.ac.at/∼pmeerw/Watermarking/WaveletSurvey/wm_wavelet_survey.zip*, June 2003.

[9] Zhu Wen-wu, Xiong Zi-xiang, Zhang Ya-Qin. Multiresolution Watermarking for Images and Video: a Unified Approach. *ICIP* '98, Vol. 1. New York: IEEE Computer Society, 1998. 465-468.

[10] Liang J, Xu P, Tran T D. A Universal Robust Low Frequency Watermarking Scheme. *http://citeseer.ist.psu.edu/340428.html*, May 2003.

□