정지영상의 Tamper Proofing을 위한 워터마킹

황 희 근, 이 동 규, 이 두 수
한양대학교 전자공학과

Watermarking for Tamper Proofing of Still Images

Hee-Keun Hwang, Dong-Gyu Lee and Doo-Soo Lee
Dept. of Electronic Engineering, Hanyang University
E-mail : tosarx@ihanyang.ac.kr

Abstract

In this paper, we propose a robust and fragile watermarking technique for tamper proofing of still images. Robust watermarks are embedded by quantization with a robust quantization step-size, and it is imperceptible value for human visual system. Fragile watermarks are embedded by thresholding and quantization with EZW(Embedded Zerotree Wav-elet) algorithm. The proposed method enables us to distinguish malicious change from non-malicious change. Futhermore this technique enables us to find tampering regions and degrees.

I. 서론

컴퓨터 네트워크의 급속한 확장과 진보된 멀티미디어 기술의 발전은 디지털 데이터의 불법적인 복사와 위조를 가능하게 하였다. 이러한 불법적인 데이터의 조작을 막기 위한 방법으로 디지털 워터마킹 기법이 제안되었다. 디지털 워터마킹이란 저작권이나 인증정보를 디지털 데이터에 삽입하는 것으로, 삽입된 정보를 워터마크라 하며, 저작권 침해(piracy)나 위조(tampering)여부를 확인하는 방법이다[1-2].

삽입된 워터마크는 다음의 특징을 가져야 한다.

(1) 원본신호(host signal)에 삽입되어질 때, 눈에 띄지 않아야 한다.
(2) 저자권자의 워터마크 추출은 용이해야 하며, 또한 추출된 마크는 신뢰성을 주어야 한다.
(3) 불법적인 워터마크 추출이 불가능해야 한다.
(4) 적용에 따라서 필터링, 압축, 외부잡음 인가, 기하학적 변형 등의 신호왜곡에 강인하거나, 또는 깨지기 쉬워야한다.

워터마킹 기법은 강인한 워터마킹(robust watermar-king)과 깨지기 쉬운 워터마킹(fragile watermarking) 기법으로 나누어진다. 전자는 고의적인 어택(attack)에 대해서 삽입된 워터마크를 제거하기가 어렵도록 고안되었다. 어택이란 추출된 워터마크의 신뢰도에 영향을 미치는 워터마킹된 신호의 조작이다. 반대로 후자는 신호 자체에 대한 신뢰도를 주기 위하여 고안되었다. 즉, 신호 안에서의 어떠한 변화나 변화된 영역을 찾을 수 있게 한다. 디지털 영상인 경우, 다양한 소프트웨어를 이용하여 영상의 일부분을 변경하거나 위조하기가 쉽기 때문에, 영상 안에서의 어떠한 변화나 변화된 부분을 찾는다는 것이 매우 중요하다. 이는 법적 증거나 보험금 청구, 신문사진 등의 위조여부를 증명하게 해준다[3].

본 논문에서는 영상의 위조여부를 증명하기 위한 방법으로, 두 가지의 강인한 워터마크와 깨지기 쉬운 워터마크를 DWT영역에 삽입한다.

II. 제안된 워터마킹 기법

원본영상을 DWT을 사용하여, N 스케일 레벨로 분해한다. 각 레벨별로 $3n$ 세부영상 즉, LH_n, HL_n, HH_n(n =1 ,2 ,…, N)과 가장 거친(coarse)레벨에서 근사화 영상 LL_N을 얻는다.

분해된 영상에 대해 근사화 영상과, 가장 정교한 세부영상을 제외한 중간 주파수 영역에 양자화 과정을 통하여 워터마크를 삽입한다. 양자화 과정은 키 값에 대응하는 웨이블릿 계수를 양자화하는 것으로, 다음절에서 자세히 설명될 것이다.

마지막으로 워터마킹된 영상에 대하여 IDWT한다.

2.1 Robust Watermarking 기법

강인한 워터마크는 선택된 웨이블릿 계수들에 대해 양자화 과정을 수행함으로서 삽입된다[4]. 양자화 과정은 다음과 같다.

선택된 웨이블릿 계수들을 $X = [x_k]$라 놓고, 강인

한 워터마크를 $R = [r_k]$라 놓는다. 여기서 r_k는 저자의 서명이나 ID로부터 의사랜덤(pseudo-random)하게 만든 K개의 이진워터마크이다. 즉, $r_k = \{-1, 1\}$이며, $k = 1, 2, \cdots, K$ 이다.

단계1: 우선 $q_k = x_k / \Delta$을 계산한다. 여기서 Δ는 양자화 간격이다. Δ값이 클수록 더욱 강인한 워터마크를 삽입할 수 있으나, 그만큼 원본영상을 훼손시키게 된다. 그러므로 강인한 워터마크를 삽입할 경우에는, 인간시각시스템이 인식하지 못하는 범위의 값인 Δ_R값으로 한다[5].

단계2: 구해진 q_k값을 r_k값으로 양자화한다.

만약 $r_k = -1$이면, $\widehat{q_k}$= 우수(even),

$\widehat{q_k} \in (q_k - 1,\ q_k + 1]$.

만약 $r_k = 1$이면, $\widehat{q_k}$= 기수(odd),

$\widehat{q_k} \in (q_k - 1,\ q_k + 1]$.

단계3: 양자화된 값 $\widehat{q_k}$부터 웨이블릿 계수 값을 얻는다. $\widehat{x_k} = \widehat{q_k} \cdot \Delta$

양자화 과정은 그림 1과 같다.

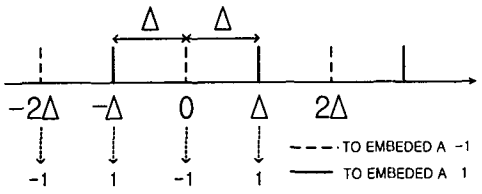


그림 1. 워터마크를 삽입하는 양자화 과정

워터마크의 추출은 삽입과정과 같이 대상영상을 DWT한 후, 키 값에 대응하는 웨이블릿 계수에 대해 삽입과정에서 쓰인 Δ값으로 $\widetilde{r_k} = round\,[\,\widetilde{x_k} / \Delta]$을 수행한다. 여기서 $\widetilde{X} = [\widetilde{x_k}]$는 워터마킹된 영상에 대한 키 값에 대응하는 웨이블릿 계수이고, 함수 $round\,[\,\cdot\,]$는 round-off 연산이다.

추출된 워터마크 $\widetilde{r_k}$와 삽입된 워터마크 r_k의 상관도 $\rho(R, \widetilde{R})$가 미리 정의된 임계치 T_R보다 크면 워터마크는 검출된다. 검출 조건은 식(1)과 같다[4].

$$\rho(R, \widetilde{R}) = \frac{\sum r_k \widetilde{r_k}}{\sum r_k^2 \sum \widetilde{r_k}^2} \geq T_R \qquad (1)$$

이러한 방법으로 삽입된 워터마크는 JPEG압축, 가우시안 잡음인가, 필터링과 같은 일반적인 신호처리에 더 강인함을 보인다[4].

2.2 Fragile Watermarking 기법

깨지기 쉬운 워터마크는 거친 레벨에서의 세부영상들의 웨이블릿 계수 $X = [x_l]$에 대해 EZW(embedded zerotree wavelet)알고리즘을 도입하여 삽입된다[6-7]. 즉, 웨이블릿 계수는 임계치 T_F와 비교된 후, 양자화과정을 통하여 삽입된다.

깨지기 쉬운 워터마크를 $F = [f_l]$라 놓는다. 여기서 $f_l = \{-1, 1\}$이며, $l = 1, 2, \cdots, L$ 이다.

단계1: 임계치 T_F 값을 설정한다.

단계2: 계수 x_l의 절대값과 T_F와 비교한다.

$|x_l| > T_F$이면, 2.1절에서 설명한 방법으로 양자화한다. (단, $\Delta = \Delta_F$)

$|x_l| \leq T_F$이면, 단계3으로 넘어간다.

단계3: m ($m < T_F$)값을 삽입한다.

$f_l = -1$이면, $x_l = -m$,

$f_l = 1$이면, $x_l = m$.

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	Tampering Region		30	31	32
33	34	35			38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64

그림 2. tamper proofing을 위해 64개의 블록으로 나눈다.

워터마크의 검출과정은 다음과 같다. 우선 영상을 그림 2와 같이 8×8블록으로 나눈다. 그리고 k번째 블록의 위조여부를 확인하기 위해 TAF(tamper assessment function)을 계산한다[3]. TAF는 식(2)와 같다.

$$TAF_k = \frac{1}{M} \sum_{i=1}^{M} f_k(i) \oplus \widetilde{f_k}(i) \qquad (2)$$

여기에서 f_k와 $\widetilde{f_k}$는 k번째 블록에서 삽입된 워터마크와 추출된 워터마크이고, M은 한 블록안에 삽입된 워

터마크의 길이이며, $\overline{\oplus}$는 not exclusive-or 연산이다. TAF_k 값은 '0'과 '1'의 범위에 존재하는데, 영상의 위조여부는 TAF_k 값으로 알 수 있다.

- $TAF_k = 1$, unaltered.
- $\delta \le TAF_k < 1$, non-malicious.
- $TAF_k < \delta$, malicious.

여기서 δ값은 사용자 정의 값으로, 보다 더 신뢰도가 필요한 경우에는 δ값이 클수록 좋다[8].

III. 실험결과

제안된 기법의 성능을 평가하기 위해서 그림 3(a)의 512×512 "Lena" 영상을 사용하였다. 사용된 DWT는 선형위상 9/7 필터를 사용하였다[5]. 그림 3(b)는 파라미터 $N=4$, $\Delta_R=14.16$, $\Delta_F=7$, $T_F=6.999$, $m=2$ 값으로 워터마크가 삽입된 영상이다. 이 경우에, 강인한 워터마크는 ($K=256$) *LH4* 영역에 삽입하고, 깨지기 쉬운 워터마크는($M=20$, $L=1280$) *HL4*, *HL3*영역에 삽입하였다.

강인한 워터마크에 대한 실험으로, 그림 3(b)의 영상에 가우시안 잡음인가(ADWGN), JPEG압축의 어택을 가했다. 표 1은 각각의 어택에 대한 $\rho(R, \widetilde{R})$값을 나타낸다.

표 1. ADWGN, JPEG 어택에 대한 $\rho(R, \widetilde{R})$

ADWGN (PSNR[dB])	$\rho(R, \widetilde{R})$	JPEG (압축비율)	$\rho(R, \widetilde{R})$
39.16	1.0	7.76	1.0
36.55	0.9453	9.14	0.9141
33.41	0.7578	12.8	0.6797
22.26	0.1484	17.07	0.1094

깨지기 쉬운 워터마크에 대한 실험으로 그림 3(b) 영상에 대해, 그림 2의 정사각형 위조영역(120×120 pixels)을 변경한다. 우선 세가지의 위조된 영상을 만든다. 그림 4(a)는 워터마크가 삽입된 영상중 위조영역에 대해 5×5 mean 필터링한 그림이다. 그림 4(b), (c)는 그림 3(b)의 워터마킹된 영상에 대해 각각 7.76의 압축비율로 압축, PSNR이 39.16[dB]가 되게 하는 가우시안 잡음 인가를 한 후, 위조영역에 대해 원본영상 즉, 워터마크가 삽입되지 않은 영상으로 대체한 그림이다.

그림 4(d), (e), (f)는 64블록에 대한 TAF_k값을 나타낸다. 그림 4(d)에서, 만일 $\delta=1$라 한다면, 제안된 방법은 영상에서의 변화된 모든 부분을 찾을 수 있게한다. 즉 27, 28, 35, 36 블록이 위조되었다. 그림 4(e), (f)인 경우, $\delta=0.75$라 한다면, 이는 잡음인가와 JPEG압축과 같은 non-malicious 변화로부터 malicious 변화를 구별할 수 있게 한다. 또한 TAF_k값이 더 작을수록, 블록의 변화가 더 많음을 알 수 있다.

IV. 결론

본 논문은 정지영상의 위조여부를 확인하기 위하여, 강인한 워터마크와 깨지기 쉬운 워터마크를 삽입하는 기법을 제안하였다. 제안된 기법은 영상의 변화된 위치와 변화된 정도를 확인할 수 있게 해주며, 또한 JPEG 압축, 가우시안잡음 인가와 같은 non-malicious 변화와 malicious 변화를 구별할 수 있게 한다.

Reference

[1] S. Katzenbeisser, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House.

[2] W. Bendor, D. Gruhl and N. Morimoto "Techniques for Data Hiding," *Proc. SPIE*, Feb. 1995.

[3] D. Kundur, "Digital Watermarking for Telltale Tamper Proofing and Aunthentication," *Proc. IEEE*, vol. 87, no. 7, pp. 1167-1180, July 1999.

[4] D. Kundur and D. Hatzinakos, "Digitalwatermark -ing using multiresolution wavelet decomposition, " *Proc. IEEE Int. Conference on*, vol. 5, 1998.

[5] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *Image Processing, IEEE Trans on*, vol. 6, pp. 1164-1175, 1997.

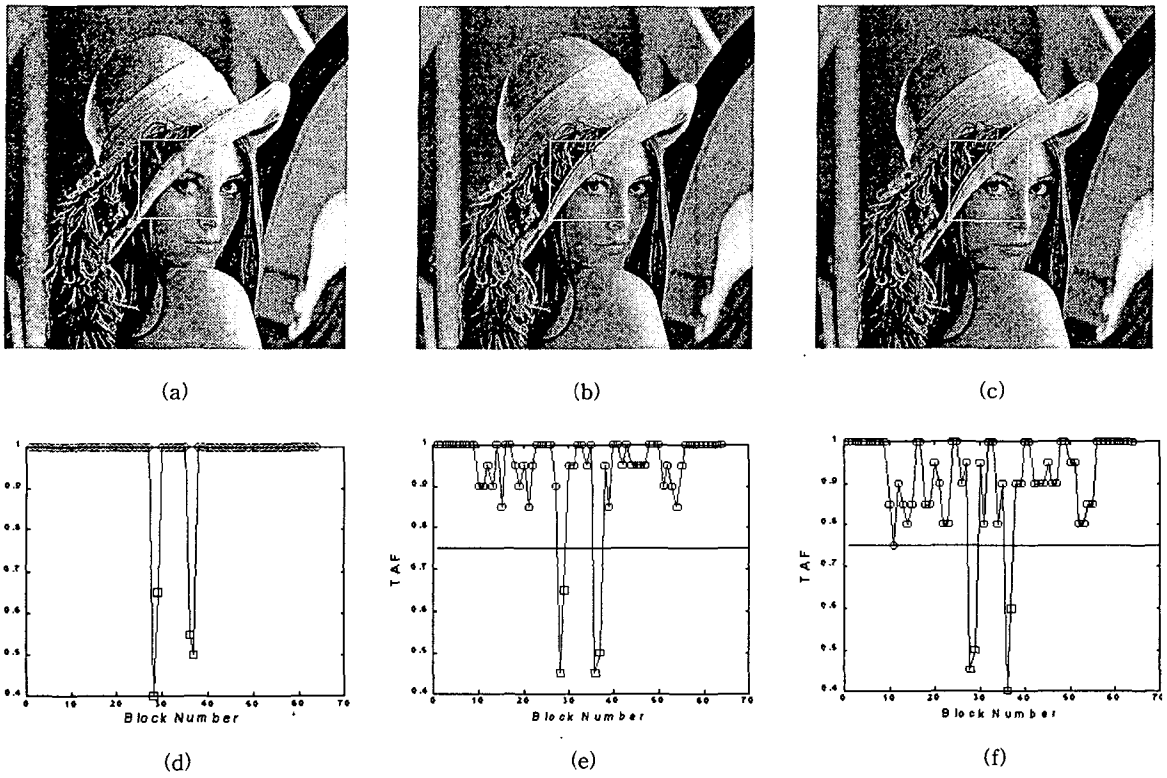
[6] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Trans. on, Signal Processing*, vol. 41, no.12, pp. 3445 -3462, Dec. 1993.

[7] H. Inoue, A. Miyazaki, A. Yamamoto and T. Ktsura, "A digital watermark based on the wavelet transform and its robustness on image compression," *Proc. IEEE Int. Conference on Image Processing*, Chicago, vol.2, pp.391-395, Oct. 1998.

[8] H. Inoue, A. Miyazaki, T. Katsura, "Wavelet-Based Watermarking for Tamper Proofing of Still Images," *Image Processing*, vol. 2, Jan. 2001.

(a) (b)

그림 3. (a)"Lena" 원본영상, (b)워터마크가 삽입된 영상(PSNR=40.56dB)



(a) (b) (c)

(d) (e) (f)

그림 4. (a)는 위조영역에 5×5 mean 필터링한 영상이고, (b)는 그림 3(b)의 워터마킹된 영상에 대해 JPEG압축한 후에 위조영역에 원본영상을 대체한 영상이며, (c)는 그림 3(b)에 대해 가우시안 잡음을 인가한 후에 위조영역에 원본영상으로 대체한 영상이다. (d), (e), (f)는 각각의 위조영상에 대한 TAF값을 나타낸다.