

특이치 분해를 이용한 비가시적 워터마크 기법

정희원 유주연*, 유지상**, 김동욱***, 김대경****

An invisible watermarking scheme using the SVD.

Joo-Yeon Lyu*, Ji-Sang You**, Dong-Wook Kim***, Dai-Gyoung Kim****, Regular Members

요 약

본 논문에서는 특이치 분해를 이용한 워터마크 기법을 제안하였다. 워터마킹의 단계는 먼저 웨이블릿 변환에 의한 저해상도 LL대역의 영상을 3×3 블록단위로 나누고, 엔트로피와 조건수를 기준으로 선택한 블록을 특이치 분해하여 워터마크를 삽입하는 것이다. 이는 워터마크 추출을 안정적으로 수행할 수 있는 저 대역 워터마크 삽입 기법을 제공한다. 본 논문에서 워터마크를 효율적으로 추출하기 위해 영상에 가해진 공격을 국소적으로 모델링하여 공격연산자를 근사적으로 구하고, 워터마크 삽입 시 수행된 특이치 분해와 추정된 공격연산자를 적용하여 삽입 위치에 따라 워터마크 집단을 결정한다. 각 워터마크 집단 내에서 최적의 워터마크를 구하고 T검정을 이용하여 워터마크의 유무를 검정한다. 제안된 워터마크 기법이 여러 단계 JPEG 공격에 견고하다는 것을 수치실험에서 확인하였다.

Key Words : digital watermarking; singular value decomposition; T-testing; wavelet transform

ABSTRACT

In this paper, we propose a new invisible digital watermarking scheme based on wavelet transform using singular value decomposition. Embedding process is started by decomposing the lowest frequency band image with 3×3 block among which we define the watermark block chosen by a key set; entropy and condition number of the block. A watermark is embedded in the singular values of each watermark blocks. This provides a robust watermarking in lowest possible time-frequency domain. To detect the watermark, we are locally modeling an attack as 3×3 matrices on the watermark blocks. Combining with the SVD and the attack matrices, we estimate watermark set corresponding to the watermark blocks. In each watermark block, we determine an optimal watermark which is justified by the T-testing. A numerical experiment shows that the proposed watermarking scheme efficiently detects the watermarks from several JPEG attacks.

I. 서론

최근 들어 정보가 디지털화 되어감에 따라 디지털 콘텐츠의 유통에 있어 저작권의 보호가 중요하며, 이를 바탕으로 한 디지털 콘텐츠 구

성이 필요하게 되었다. 현재 이에 대한 기술적인 방법 중에 하나로 국내외적으로 디지털 워터마킹이 이용되고 있으며, 멀티미디어 영상, 동영상, MP3 등 디지털 콘텐츠의 다양성에 따라 많은 비가시적인 워터마킹의 기법이 제안되고 있다. 특히, 실시간 서비스를 위해 신속성과

*한양대학교 자연과학대학 수학과 (lema99@empal.com)

**광운대학교 전자정보대학 전자공학과(isyoo@daisy.kw.ac.kr)

***광운대학교 전자정보대학 반도체 및 신소재 공학과 (dwkim@daisy.gwu.ac.kr)

****한양대학교 과학기술대학 응용수학과 (dekim@hanyang.ac.kr)

논문번호: 030279-0701, 접수일자: 2003년 7월 1일

*이 논문은 한국과학재단 목적기초연구(과제번호 : RO1-2002-000-00350-0)의 일부지원으로 이루어졌음.

견고성이 강조되는 워터마크알고리즘의 ((1), (2)) 개발이 증가하고 있다.

현재 워터마킹 기법의 추세는 웨이블릿 변환을 기반으로 한 많은 워터마킹 기법이 주류를 이룬다. ((1), (2), (3)) 이는 웨이블릿 변환이 데이터의 시간 및 주파수 성분을 동시에 효율적으로 국소할 수 있기 때문이다. ((5))

웨이블릿 기반의 비가시적인 워터마킹 기법은 주로 영상의 경계 부분의 정보를 갖고 있는 저주파 및 고주파영역인 LH 대역 (low-high frequency band)에 워터마크를 삽입하는 방법((3))을 택하고 있다. 이는 압축이나 잠음 같은 공격에 민감하여 삽입된 워터마크를 효율적으로 복원하기 어렵다는 단점이 있다. 최근에는 견고한 워터마킹 삽입을 위해 저주파영역인 LL 대역(low frequency band)에 워터마크를 삽입하는 기법((4))이 제안되었다.

본 논문에서는 보다 견고한 워터마킹 기법을 위해 웨이블릿 영역의 LL 대역에서 특이치 분해 (singular value decomposition, SVD)를 이용한 워터마크 기법을 제안하였다. 워터마크의 삽입은 먼저 웨이블릿 변환에 의해 선택된 저해상도인 LL 대역의 영상을 3×3 블록단위로 나누고, 키값으로 이용할 엔트로피 (entropy)와 조건수(condition number)를 기준으로 워터마크가 삽입될 블록을 선택함으로써 이루어진다. 여기서 블록에 대한 엔트로피와 조건수는 경계를 포함한 균일한 영상 블록을 택하는 역할을 한다. 제안된 방법은 선택한 블록을 특이치 분해하여 워터마크를 삽입하는 것이며, 이는 워터마크 추출을 안정적으로 수행할 수 있는 저대역 워터마크 삽입기법을 제공한다. 주어진 워터마크 블록의 SVD는 Karhunen-Loeve 변환과 유사하여, SVD의 유니터리(unitary) 행렬은 영상 블록을 표현하는 공분산 행렬에 의존한다. 또한 대응되는 특이치는 공분산 행렬의 고유치와 관련이 있는데, 가장 큰 특이치는 영상 블록의 저주파 스펙트럼 정보를 갖고 있다. 본 논문에서는 이 값을 변형하여 영상 블록의 저주파 정보에 워터마크를 삽입한다.

워터마크의 추출은 두 단계로 이루어진다. 먼저 주어진 영상의 LL 대역에서 워터마크 삽입 시 선택한 블록을 중심으로 공격연산자를 3×3 행렬 A 로 모델링한다. 이때 블록 주변의 웨이블릿 계수 값을

이용하여 국소적으로 공격 연산자를 추정한다. 다음 단계로 워터마크 삽입 시 수행된 특이치 분해와 추정된 공격연산자를 이용하여 워터마크 삽입 위치에 따른 워터마크 집단을 결정한다. 각 워터마크 집단 내에서 최적화된 워터마크를 추정하여 T검정을 이용하여 워터마크의 삽입 유무를 검정하는 것이다. 본 논문의 수치실험 결과는 제안된 워터마크 기법이 여러 단계의 JPEG 공격에 견고하다는 것을 보여주고 있다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 웨이블릿 영역에서 SVD를 이용한 워터마킹의 삽입 기법을 제안하였고, 3장에서는 공격연산자를 국소적으로 추정하여 워터마크를 추출하는 알고리즘을 다루었다. 4장에서는 워터마크 삽입 유무를 효율적으로 결정하기 위한 통계적 검정을 살펴보고, 마지막으로 5장에서는 수치실험 결과를 제시하고 이 결과를 토대로 6장에서는 본 논문의 결론을 맺는다.

II. SVD를 이용한 워터마크의 삽입기법

이차원 분리형 웨이블릿 변환을 이용하여 주어진 영상 데이터를 네 단계까지 분해한다. 첫 번째 단계에서 저주파-저주파영역인 LL_1 밴드, 고주파 및 고주파영역인 HH_1 밴드, 고주파 및 저주파영역인 HL_1 밴드, 그리고 저주파 및 고주파영역인 LH_1 밴드로 영상을 분해한다. LL_1 밴드를 같은 방법으로 분해하여 다시 LL_2 , HH_2 , LH_2 , HL_2 로 분해하여, 마지막으로 네 번째 단계에서 LL_4 밴드를 선택한다.

워터마크를 삽입할 영역을 정하기 위해 위하여 3×3 행렬블록으로 LL_4 밴드 영상을 분해한다. 블록 열중에서 주어진 모수 μ 와 ρ 에 대하여 각각 엔트로피가 크고 조건수가 작은 값을 갖는 블록 X_o 를 선택한다. 여기서 조건수 $\kappa(X_o)$ 는 블록행렬의 조건수(condition number)이며 엔트로피는 다음과 같이 주어진다.

$$H(X_o) = -\sum p_i \log p_i, p_i = \frac{x_i}{|x|} \quad (1)$$

이때 x_i 는 X_o 의 i 번째 픽셀 값이다. 즉

$x(X_o) < \mu$ 이고 $H(x) > \rho$ 이다. μ 와 ρ 는 워터마크의 위치와 개수를 결정하는 중요한 모수이다. 이 모수를 임계값으로 결정되는 픽셀을 중심으로 주변 여덟 개의 픽셀을 선택한다. 이와 같이 선택된 블록을 워터마크 블록으로 지정하고, 블록의 중심 픽셀 위치를 첨자로 하여 키(KEY)를 정한다.

이제 워터마크를 삽입하기 위해 워터마크 블록인 3×3 행렬 X_o 를 특이치 분해를 한다. 즉,

$$X_o = U \Sigma_o V = U \text{diag}(\sigma_1, \sigma_2, \sigma_3) V^T \quad (2)$$

이때, $\text{diag}(\sigma_1, \sigma_2, \sigma_3)$ 는 $\sigma_1, \sigma_2, \sigma_3$ 를 원소로 하는 대각행렬이며, Karhunen-Loeve 변환에의 한 분해와 같이 세 개의 특이치 $\sigma_1, \sigma_2, \sigma_3$ 는 영상블록에서 각각 상대적인 저주파, 중간주파, 고주파 성분을 표현한다. 워터마크 삽입을 위하여 저주파 성분인 σ_1 에 대한 상대오차 E_w 를 이용해 워터마크를 $\sigma_w = \sigma_1(1 - E_w)$ 와 같이 생성한다. 여기서 상대오차는 워터마크의 강도를 나타내며, 이는 하나의 키 값으로 이용한다. 위의 워터마크를 특이치 대각 행렬에 삽입하여 새로운 대각행렬 $\Sigma_w = \text{diag}(\sigma_w, \sigma_2, \sigma_3)$ 를 이용해 워터마크가 삽입된 3×3 블록 $X_w = U \Sigma_w V^T$ 를 생성한다. 마지막단계로 각 X_w 집합이 워터마크 블록에 대치되어 웨이브렛 변환을 통하여 워터마크가 삽입된 영상을 재구성한다.

III. 워터마크 검출 알고리즘

본 논문에서 제안한 검출 알고리즘의 특징은 워터마크의 추출을 효율적으로 구현하기 위해 삽입과정의 역과정 중 공격연산자를 국소적으로 추정하는 것이다. 먼저 워터마크가 삽입된 것으로 추정되는 영상을 삽입과정과 같이 웨이브렛 분해 한 후 삽입시 정해진 키를 통하여 워터마크가 삽입된 블록을 취한다.

고정된 블록에 대하여 가해진 공격을 하나의 연산자로 모델링하여 공격 연산자를 3×3 행렬 A 로 간주한다. 이때 공격연산자에 대한 사전 정보는 없다고 가정한다. 따라서 공격연산자

A 를 추정하는 것은 원 영상의 해당 블록의 특이치 분해의 결과를 정의역으로 하고 공격당한 영상의 해당 블록을 치역으로 하는 연산자 A 를 구하는 역문제로 모델링 된다. Y 를 공격당한 영상의 LL_1 밴드에 있는 3×3 블록이라 하고, 이로부터 추정될 블록을 X 라 하자. 즉, $AX = Y$ 라 하면 $AX = A(U \Sigma V^T) = Y$ 이므로

$$(A U) \Sigma = Y \quad V = Z. \quad (3)$$

이때 $\Sigma = \text{diag}(\sigma, \sigma_2, \sigma_3)$ 이며 σ 는 구해야 할 σ_1 의 추정치가 된다. 식 (3)에서 공격연산자 A 를 구하기 위해, U 의 열벡터를 u_1, u_2, u_3 이라 하고,

$$b_1 = Au_1, \quad b_2 = Au_2, \quad b_3 = Au_3, \quad e = Ae \quad (4)$$

라 놓자. 이때 $e = (1, 1, 1)^T$ 는 A 를 추정하기 위해 가정한 가상의 벡터이다. 본 논문에서는 공격에 민감하지 않은 LL 밴드에서 $e = Ae$ 로 가정하여 공격연산자를 모델링 하였다. 더욱이 공격한 후의 결과는 주어진 블록 주변의 픽셀 값의 변화와 관계가 있으므로 (4)의 마지막 식에서 e 대신에 주어진 블록의 주변의 오른쪽, 왼쪽 벡터를 이용하여 공격연산자를 모델링 하였다. 결과적으로 공격연산자는 세 가지 방법으로 모델링하였다.

이제 Z 의 열벡터를 z_1, z_2, z_3 라 하고

$$B = [e, b_2, b_3], \quad C = \left[e, \frac{1}{\sigma_2} z_2, \frac{1}{\sigma_3} z_3 \right] \quad (5)$$

라 놓으면, 공격연산자를 $A = C B^{-1}$ 와 같이 구할 수 있다. 따라서 σ_1 를 다음 관계식으로부터 추정할 수 있다.

$$A u_1 \sigma = z_1 \quad (6)$$

식 (6)에서 σ_1 를 추정하는 방법을 다음과 같이 세 가지로 정리 할 수 있다.

(i) 벡터의 방향비: z_1 과 Au_1 의 각 요소의 비율을 결정한다. 즉,

$$\sigma^1 = \frac{z_{11}}{b_{11}}, \quad \sigma^2 = \frac{z_{12}}{b_{12}}, \quad \sigma^3 = \frac{z_{13}}{b_{13}} \quad (7)$$

(ii) 노름(norm)의 보존:

$$\sigma^4 = \frac{|z_1|}{|b_1|} \quad (8)$$

(iii) 최소제곱방법:

$$\sigma^5 = \frac{z_1^T b_1}{|b_1|} \quad (9)$$

위의 다섯 가지 추정 값을 세 가지로 모델링된 공격연산자에 각각 적용하면 σ_1 에 대한 총 15가지의 추정치를 얻는다. 이렇게 추정된 σ^j , $j=1, \dots, 15$ 로 σ_1 에 대한 상대오차를 다음과 같이 얻을 수 있다.

$$\frac{|\sigma_1 - \sigma^j|}{\sigma_1} \approx E_n \quad (10)$$

이러한 상대오차의 집단을 기준으로 통계적 가설과 T검정을 통해 워터마크를 검출하고 또한 워터마크 삽입여부를 판단한다.

IV. 통계적 접근방법

1. 평균제곱오차

15가지 방법 가운데 어떤 방법으로 추정된 σ^j 이 가장 효율적으로 워터마크 제공하는가를 결정하기 위해 평균제곱오차를 이용한다. 이는 워터마크를 추정하는 σ_1 의 집단 어떤 분포를 따르는지 상관없이 평균제곱오차가 가장 작은 방법을 선택하면 된다. θ 가 실제 상대오차 값이고 $\hat{\theta}$ 이 추정된 값이라면

$$MSE = E(\hat{\theta} - \theta)^2 = \text{Bias}(\hat{\theta}) + \text{var}(\hat{\theta})$$

이다. 여기서 $\text{Bias}(\hat{\theta})$ 는 추정량 $\hat{\theta}$ 의 편의를 의미한다. 즉 MSE는 추정량과 실제값의 거리의 제곱에 대한 기대값이다. MSE에서 가장 작은 오차값이 나온 방법이 효율적인 추정 방법이다.

2. T-testing

어떤 집단의 기대값을 검정하거나 두 집단의 기대값이 같다고 볼 수 있는지를 검정하고자 할 때, T검정이 사용된다. 여기서는 두 샘플의 T검정을 이용하여 워터마크가 들어 있지 않은 집단과 워터마크를 포함하고 있는 집단의 상대 오차를 비교하여 두 집단이 다르다는 결론을 이끌어낸다. 이때 두 집단 모두 정규분포를 따른다고 가정한다.

V. 수치실험결과

9/7 필터 웨이블릿 변환을 이용하여 Lenna 영상을 LL_4 밴드까지 분해하였고, 여기서 $\rho=2.1$, $\mu=50$ 로 하여 워터마크 블록을 설정하여 강도 $E_n=0.0313$ 로 워터마크를 삽입하였다. 워터마크를 삽입한 후의 영상은 PSNR이 49.1938이었다. 모의실험을 위하여 JPEG 공격을 택하였다. 표 1은 각 JPEG 공격에 대해 추출한 E_n 의 추정치이다. 공격 연산자는 세 가지로 모델링하였다. 즉, $e=Ae$ 와 e 대신에 주어진 블록 주변의 오른쪽, 왼쪽 벡터를 이용하였다.

평균 제곱의 오차에서 나온 결과를 기준으로 가장 최적의 방법을 선택한 후에 그 방법을 T검정하였다. 대부분의 JPEG 압축에 대하여 p-value는 거의 0에 가까웠다. 그러므로 귀무가설 H_0 는 기각된다. 그러므로 모든 JPEG 압축에 대하여 거의 모든 상대오차값이 0.0313에 가깝게 추정되었다.

JPEG	추정된 E_n 값
10	0.0311
20	0.0303
30	0.0313
40	0.0310
50	0.0313
60	0.0315
70	0.0313
80	0.0312
90	0.0312

표 1. JPEG 공격에 따라 추출된 E_n 의 추정값

VI. 결론

본 논문에서는 웨이블릿 영역 위에서 특이치 분해를 이용한 워터마크 기법을 제안하였다. 엔트로피와 조건수를 기준으로 선택한 워터마크 블록을 특이치 분해하여 워터마크를 삽입하고, 견고한 워터마크 추출을 위하여 영상에 가해진 공격을 국소적으로 모델링하여 삽입시 결정된 특이치 분해를 이용하여 최적의 워터마크를 추정하였다. 특히 T점정을 이용하여 워터마크가 들어있는 영상과 워터마크가 들어 있지 않은 영상을 확실히 구분할 수 있었다. 본 논문에서 제안된 기법은 JPEG 공격에 대해 워터마크 추출을 안정적으로 수행할 수 있는 저대역 워터마크 기법이라 할 수 있다. 다양한 공격과 영상에 대한 본 워터 기법의 응용이 진행 중에 있다.

참고 문헌

- [1] Kunder, D Hatzinakos, D; "A robust digital image watermarking method using wavelet based fusion." in *Proceedings of the IEEE International Conference on Image Processing, ICIP'97, vol 1 pp.544-547, (Santa Barbara, California, USA) October 1997.*
- [2] Xu-Dong Zhang, Kwok-Yung Lo, Jian-Feng' and Desheng wang, "A robust image watermarking using spatial-frequency feature of wavelet transform," *Proceedings of ICSP2000.*
- [3] X. G. Xia, C. G. Boncelet, G. R. Arce, "A Multiresolution watermark for digital images," *Int Conf. On Image Processing*, pp. 548-551, 1997
- [4] Snahyun Joo, Yong ho Suh, Jaeho Shin, and Hizakasu Kikuchi, "A New robust watermark embedding into wavelet DC components." *ETRI journal, vol 24, Number 5. pp 401-406, October 2002*
- [5] Ali N. Akansu, Richard A. Haddad "Multiresolution Signal Decomposition, Transform, Subbands, and Wavelets" *ACADEMIC-PRESS, INC. 1997*

유 주 연(Joo-Yeon Lyu)

정회원



1992년 2월 : 한양대학교 안산캠퍼스 수학과졸업
1994년 2월 : 한양대학교 수학과석사
1995년 3월~현재 : 한양대학교 수학과 박사과정

<주관심분야> watermarking, cryptography

유 지 삼(Ji-Sang Yoo)

정회원



1985년 2월: 서울대학교 전자공학과 졸업
1987년 2월: 서울대학교 전자공학과 석사
1993년 5월: Purdue University Electrical Engineering, 박사학위 취득

1997년~현재: 광운대학교 전자공학과교수

<주관심분야> Signal & image processing, Nonlinear digital filtering, data broadcasting, watermarking,

김 동 녹(Dong-Wook Yoo)

정회원



1983년 2월:한양대학교 전자공학과 졸업
1985년 2월:한양대학교 전자공학과 석사.
1991년9월: Georgia Institute of technology, 박사학위 취득
1992년~현재:광운대학교 반도체 및 신소재공학과 교수

<주관심분야> Non-standard DWT-based Image/Video CODEC, watermarking, cryptography

김 대 경(Dong-Wook Yoo)

정회원



1983년 2월:한양대학교 수학과 졸업
1986년 2월:한양대학교 수학과 석사
1994년 : Purdue University, 박사학위취득
1995년~현재 :한양대학교 응용수학과 교수

<주관심분야> Harmonic analysis, Numerical analysis, watermarking, wavelet theory and application, numerical PDE.