내적과 양자화를 이용한 영상의 워터마킹

이승욱, 호요성
광주과학기술원 정보통신공학과
광주광역시 북구 오룡동 1 번지

Image watermarking using Projection and Quantization

Seung-Wook Lee and Yo-Sung Ho
Kwangju Institute of Science and Technology (K-JIST)
1 Oryong-dong, Puk-gu, Kwang-ju, Korea
E-mail: swlee@gogh.kjist.ac.kr, hoyo@kjist.ac.kr

요약

디지털 워터마킹이란 디지털 컨텐츠의 저작권을 보호하기 위해 보이지 않는 임의의 데이터를 미디어에 삽입하는 방법이다. 본 논문에서는 주파수 영역으로 변환된 원 데이터를 임의의 방향으로 정의된 데이터와 내적(Inner Product)하여 이를 워터마크 정보에 따라 변화시키는 방법을 제안한다. 이 방법은 비밀키가 워터마크 데이터와 연관된 것이 아니기 때문에 어떠한 종류의 데이터도 삽입할 수 있다는 장점이 있다. 사용되는 비밀키는 주파수 영역으로 변환된 원 데이터와 내적되는 임의의 데이터를 만드는데 사용된다. 또한 워터마크의 견고성을 높이기 위해, 주파수 영역으로 변환된 원 데이터와 내적된 임의의 데이터는 잡음처럼 되므로 이를 인간 시각 특성을 사용하여 모델링하였다. 제안된 방법은 원 영상 없이 워터마크를 검출할 수 있으며, 워터마크의 견고성 실험을 위해 JPEG, Cropping, Resizing, Gaussian 잡음 등을 적용하였다.

1. 서론

최근 다양한 멀티미디어 데이터 전송 및 저장 장치, 그리고 저작 도구의 발달로 인해 네트워크를 통한 멀티미디어 서비스가 급격하게 늘어나고 있다. 디지털 미디어는 아날로그 미디어와는 달리 복사하기도 쉽고, 복사본과 원본 사이의 구별이 어려워 저작권 문제가 크게 대두된다. 이를 해결하기 위한 방법으로 암호화 방법(Encryption)이 많이 연구되었다. 데이터 암호화 방법은 사용자의 인증 절차를 거친 후 암호화된 데이터를 복원하는 방법인데, 일단 암호화된 데이터가 해독되면 더 이상 데이터를 보호할 수 없다는 단점이 있다.

이에 비해, 디지털 워터마킹 기술은 텍스트, 영상, 오디오, 비디오 등의 디지털 저작물에 눈이나 귀로서는 식별이 불가능한 임의의 마크를 삽입하여 저작권자나 소유자의 허락 없이 저작물을 복사하거나 배포하는 것을 방지하는 기술이다. 만약 사용자들이 워터마크된 디지털 정보를 불법으로 복제하거나 소유주의 허락없이 사용했을 경우에는 소유자가 자신의 워터마크를 추출함으로써 자신의 소유권을 주장할 수 있다.

기존의 워터마크 방법은 크게 공간 영역(Spatial Domain) 삽입과 주파수 영역(Frequency Domain) 삽입으로 나뉘어진다. 공간 영역 삽입은 원 데이터의 시각적으로 덜 중요한 LSB(Least Significant Bit)에 워터마크 데이터를 삽입하는 방법이다. 이는 구현하기 쉽다는 장점은 있으나, 압축이나 Cropping 등의 공격에 약하다는 단점이 있다. 주파수 영역 삽입은 일반적으로 대역확산 통신(Spread Spectrum Communication)[1]에 기반을 두고 있다. 원 데이터가 주파수 영역으로 변환되고, 시각적으로 중요한 부분인 저주파 영역에 워터마크가 삽입된다. 이때 삽입되는 워터마크는 어떤 특정한 비밀키(Private Key)에 의해서 발생되는 의사 가우시안 노이즈(Pseudo Gaussian Random Sequence)가 된다. 이때 DCT, Wavelet, Cepstrum 등의 변환이 사용된다[2]. 워터마크를 추출할 때는 똑 같은 비밀키를 사용하여 워터마크를 추출하고, 특별한 방법으로 정의된 상관도(Correlation)[1]를 계산하여 워터마크의 존재 여부를 확인한다.

2. 워터마크 삽입

본 논문에서는 원 영상을 8x8 화소의 블록 단위로 2 차원 DCT 를 수행하여 각 블록당 4 비트의 워터마크를 삽입한다. 워터마크는 임의의 영상신호를 0 과 1 로 바꾼 비트열이 된다. 즉, 본 논문에서 사용되는 워터마크는 어떤 멀티미디어 데이터도 될 수 있다. 그림 1 은 워터마크 삽입의 전체적인 블록도를 보여준다.

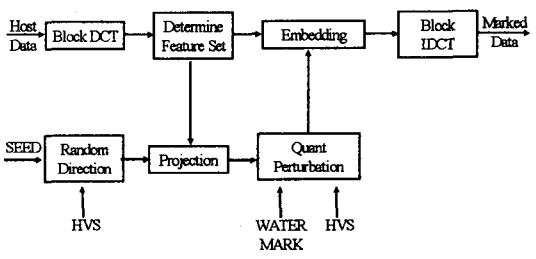


그림 1. 워터마크 삽입의 블록도

2.1 Feature Set 의 결정

주파수 영역에서 워터마크가 삽입되는 부분을 Feature Set 이라고 정의하는데, 어떤 부분을 이용하느냐에 따라서 워터마크의 강인성과 비 가시성 사이의 Trade-Off 가 생긴다. 저주파 성분을 Feature Set 으로 하면 강인성은 얻을 수 있으나, 시각적으로 중요한 부분에 워터마크를 삽입하였기에 영상의 열화가 눈에 보인다는 단점이 있다. 반대로, 고주파 부분을 Feature Set 으로 정의하면 영상의 열화는 줄일 수 있지만, 워터마크의 강인성은 얻을 수 없다는 단점이 있다. 제안하는 알고리즘에서는 여러 가지의 공격에 강인하게 하기 위해 저주파 부분을 사용한다. 그리고 블록당 네 비트가 삽입되는데, 네 개의 Feature Set 은 다음과 같이 정의된다.

$$F_{1b}=(c_{10b},c_{20b})\quad F_{2b}=(c_{01b},c_{02b},c_{03b})$$
$$F_{3b}=(c_{11b},c_{12b})\quad F_{4b}=(c_{21b},c_{30b}) \tag{1}$$

여기서 F_{ib} 는 b 번째 블록의 i 번째 Feature Set 이고, c_{ijb} 는 b 번째 블록의 (i,j)위치의 DCT 계수이다.

2.2 Random Direction Table 의 결정

Feature Set 의 데이터는 임의의 방향 데이터와 내적(Inner Product)되는데, 이 방향 데이터는 사용자 키로부터 결정된다. 먼저 사용자 키로부터 네 가지 seed_vector 를 만든다. 이를 이용하여 네 가지 테이블을 만드는데, 이를 Random Direction Table (RDT)이라한다. 이 테이블은 워터마크를 삽입할 때 잡음이 되는 부분이므로, 이를 다음과 같이 정의되는 MTF[3]에 반비례하게 정의한다.

$$H(u,v)=a\left(b+c\frac{\sqrt{u^2+v^2}}{2N}f_s\right)\cdot\exp\left(-c\frac{\sqrt{u^2+v^2}}{2N}f_s\right)^d \tag{2}$$

RDT 인 Z_1, Z_2, Z_3는 각각 $\lfloor N_{row}/8\rfloor$ x $\lfloor N_{col}/8\rfloor$ x 2 의 크기이고, Z_2는 $\lfloor N_{row}/8\rfloor$ x $\lfloor N_{col}/8\rfloor$ x 3 의 크기이다. N_{row}와 N_{col}은 각각 영상의 크기를 나타낸다. 각 테이블의 인자는 다음과 같이 정의된다.

$$z_{ij\_row}=seed\_vector_i(j)/H(u,v) \tag{3}$$
$$u,v:\ \text{Feature Set의 인덱스}$$

여기서 z_{ij_row}는 i 테이블의 j 번째 행을 의미한다. 그리고 원 영상없이 워터마크를 추출하기 위하여 RDT 의 크기를 1 로 정규화한다.

2.3 내적과 양자화

Feature Set 의 데이터와 먼저 계산된 RDT 를 내적하여 p_{ib} 를 계산한다.

$$p_{ib}=F_{ib}\cdot z_{ij\_row}\quad \text{for } i=1,2,3,4 \tag{4}$$

다음으로 내적 p_{ib}는 워터마크의 삽입을 위해 레벨 T 를 기준으로 $p_{quant}=kT$ (k는 정수)와 같이 양자화된다. 즉 양자화된 값에 어떤 특정한 값을 더하거나 빼어도 그 값은 어떤 일정한 형태를 가지게 되므로 어떤 값(워터마크)이 첨가되었는지를 쉽게 알아낼 수 있다.

2.4 워터마크의 삽입

워터마크의 삽입을 위해 양자화된 내적값을 워터마크에 따라 다음과 같이 변화시킨다.

$$p'=\begin{cases}p_{quant}+0.25T & \text{워터마크=1}\\ p_{quant}-0.25T & \text{워터마크=0}\end{cases} \tag{5}$$

이렇게 하면 워터마크 정보가 삽입된 p'은 어떤 기준점을 중심으로 양쪽에 존재한다. 따라서 삽입된 정보를 추출하기 위해서는 기준점보다 "크냐? 작냐?"만 계산하면 된다.

이렇게 워터마크가 삽입된 내적값을 이용하여 Feature Set 의 데이터에 직접 삽입하고 블록 단위로 IDCT 하여 워터마크가 삽입된 영상을 얻는다. 이때 블록 단위의 워터마크 삽입은 다음과 같이 이루어 진다.

$$\bar{v}'=\bar{v}+(p'-p)\bar{z} \tag{6}$$

여기서 $\bar{v}$는 원래의 데이터이고, $\bar{v}'$은 워터마크가 삽입된 데이터이며, $(p'-p)\bar{z}$는 잡음이다. 그림 2 는 이를 벡터로 표현한 것이다.

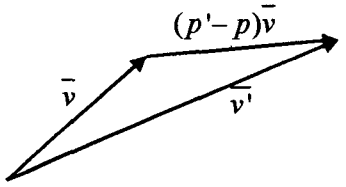


그림 2. 워터마크의 삽입

3. 워터마크 추출

삽입된 워터마크를 추출하는 과정은 삽입하는 과정과 유사하다. 먼저 워터마크가 삽입된 부분을 정의하고 사용자에게 주어진 비밀키를 이용하여 RDT 를 만든다. 이를 이용하여 워터마크가 삽입되어 있는 p'를 구해 워터마크를 검출한다.

3.1 p'의 계산과 워터마크 추출

앞에서 언급한 것과 같이, p'에 워터마크가 삽입되어 있으므로, p'의 값이 주어지면 워터마크를 추출할 수 있다. $p'=p_{quant}\pm0.25T=(k\pm0.25)T$ 와 같이 정의되므로, 다음과 같은 과정을 거친 후 삽입된 워터마크를 추출할 수 있다.

$$\text{Extracted bit}=\begin{cases}1 \text{ if } \frac{p'}{T}-Round\left(\frac{p'}{T}\right)>0\\ 0 \text{ if } \frac{p'}{T}-Round\left(\frac{p'}{T}\right)<0\end{cases} \tag{7}$$

Blind 워터마크를 위해 p'은 원 영상 없이 계산

되어야 한다. 워터마크가 삽입된 부분과 RDT 각 인자들 사이의 내적을 구하면 $\bar{v}' \bullet \bar{z} = [\bar{v} + (p' - p)\bar{z}] \bullet \bar{z}$이 되는데, 이 값은 $p+(p'-p)$이 되고, 최종적으로 p'이 된다. 그 이유는 RDT를 구할 때 $\bar{z} \bullet \bar{z} = 1$ 이 되도록 정규화하였기 때문이다.

3.2 워터마크 추출 임계값의 결정

본 논문에서는 이미지를 워터마크로 삽입했기 때문에, 워터마크의 존재유무를 판단하는 기준으로 추출된 워터마크 자체를 사용할 수도 있고, 추출된 워터마크와 삽입된 워터마크의 상관도를 계산하여 이용할 수도 있다. 이를 위해 먼저 상관도를 다음과 같이 정의한다.

$$X(similarity) = \sum_{i=1}^{M} s_i \qquad (8)$$

여기서 s_i는 i번째 추출된 워터마크와 i번째 삽입된 워터마크의 유사도를 나타낸다. 즉, 두 비트가 같으면 1, 그렇지 않으면 0으로 정의된다. 즉, 상관도는 정확하게 추출된 비트의 개수와 동일하다. 이는 [2]의 상관도 정의와 비슷하다. 이번 실험에서는 임의로 추출된 비트열과 주어진 워터마크의 비트열 사이의 상관도를 확률적으로 모델링하여 임계값을 결정한다. 삽입된 워터마크가 1이 될 확률과 추출된 워터마크가 1이 될 확률을 각각 $Pr(w_i=1)=p$, $Pr(d_i=1)=r$이라고 정의하자. 임의로 추출된 비트열과 삽입된 워터마크의 비트열은 서로 독립이라고 가정할 수 있으므로 $Pr(s_i=1) = pr+(1-p)(1-r)$이 된다. 따라서 확률변수 s_i는 0과 1의 값을 가지는 Bernoulli 분포가 된다. 그러면 상관도 X는 iid (Independent Identically Distributed) Bernoulli 분포의 합이고 이는 Binomial 분포이다[4]. 따라서 상관도가 k가 될 확률은 다음과 같다.

$$\Pr(X_B = k) = \binom{M}{k}(1+2pr-p-r)^k(p+r-2pr)^{M-k} \qquad (9)$$

평균 $E[X_B]$는 $M(1+2pr-p-r)$이 되는데, 이를 m으로 정의한다. 그리고 분산 $VAR[X_B]$는 $M(1+2pr-p-r)(p+r-2pr)$이 되는데, 이를 σ^2로 정의한다. 여기서 M이 충분히 크다고 가정하면 Central Limit Approximation[4]을 적용할 수 있다. 그러므로 상관도가 k가 될 확률은 다음과 같이 근사화 된다.

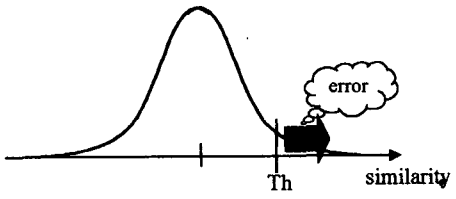
$$\Pr(X_G = k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(k-m)^2}{2\sigma^2}\right) \qquad (10)$$

상관도를 이용한 워터마크의 존재 여부를 판단하는 임계값을 Th라고 하면, 본 논문에서는 상관도가 Th보다 크면 워터마크가 있다고 결정한다. 따라서, 만약 임의로 추출된 데이터 비트열과 삽입된 워터마크 비트열의 상관도가 임계값 Th보다 크면 워터마크가 있다고 결정한다. 이런 기준을 적용하면 워터마크를 삽입하지 않았는데 워터마크가 존재한다고 결정하는 오류(False Alarm Probability)가 존재한다. 이 오류는 임의의 비트열과 워터마크의 비트열의 상관도가 우연히 어떤 임계치보다 큰 경우이다. 좀더 정확한 임계값의 결정을 위해서는 False Alarm 확률과, 워터마크가 있는데 없다고 결정하는 False Negative 확률을 모두 계산해야 한다. False Negative 확률을 계산 하려면 워터마크된 영상에 가해지는 여러 가지 고의적인 공격의 강도를 알아야 된다. 이는 너무 복잡하므로 모델링 하기가 쉽지않다. 그래서 본 논문에서는 False Negative 확률만을 계산한다. 이러한 추출 오류는 다음과 같이 계산할 수 있다.

$$\Pr(error) = \int_{Th}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(k-m)^2}{2\sigma^2}\right) dk \qquad (11)$$

이를 계산하면 다음과 같다.

$$\Pr(error) = \int_{\frac{Th-m}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = Q\left(\frac{Th-m}{\sigma}\right) \qquad (12)$$



본 실험에서는 오류 확률을 3.17×10^{-5} [$Q(4)$]으로 하여 임계값은 $Th=4\sigma+m$이 된다. 그림 3은 위의 수식을 표현한 것이다.

그림 3. 상관도의 확률분포

4. 실험 및 결과

본 실험에서는 원 영상으로 512x512 크기의 회색도 영상을 사용하였다. 전체 영상을 8x8 화소의 블록으로 나누어, 각 블록마다 4 비트씩 삽입하여 총 삽입되는 비트수는 16384 비트가 된다. 워터마크 영상으로는 32x64 크기의 8 비트 회색도 영상을 삽입하였다. 상관도의 임계치를 계산하기 위해 $Pr(d_i=1)=0.5$로 가정하여 $Th=8848$을 구하였다.

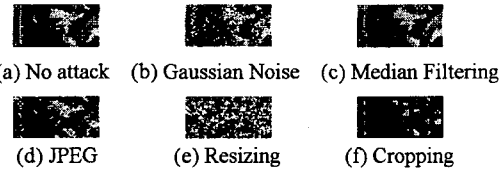
그림 4는 실험에 사용된 원 영상과 워터마크 영상을 보여준다. 그림 5는 여러 가지 공격을 가했을 때 추출된 워터마크 영상을 보여준다. 그림 5(a)는 공격을 가하지 않았을 때의 결과이다. 추출된 워터마크와 삽입된 워터마크의 상관도는 16384로 삽입된 워터마크는 정확하게 추출되었다. 그림 5(b)는 가우시안 잡음 삽입하였을 때의 결과인데 상관도는 14124이다. 추출된 워터마크를 살펴보면 비트 단위로 오류가 발생하는데 이는 마치 Salt-Pepper 잡음처럼 작용한다. 이는 미디언 필터를 사용하여 어느 정도 복원할 수 있다. 그림 5(c)는 복원된 영상을 보여준다. 실험에 의하면 가우시안 잡음을 25.8 dB가 될 때까지 삽입하여도 상관도를 이용하여 워터마크의 존재여부를 확인할 수 있었다. 그림 5(d)는 JEPG으로 11.43배 압축한 후 추출된 워터마크를 보여주고, 상관도는 14819이다. 이 또한 미디언 필터링 하여 향상시킬 수 있다.

그림 5(e)는 워터마크된 영상의 크기를 반으로 줄

인 후 원래대로 크게한 영상에서 추출된 워터마크를 보여준다. 그림에서 보듯이 7285 비트의 오류가 생겨 추출된 워터마크 영상은 아무 의미를 가지지 않는다. 이러한 결과는 삽입된 워터마크의 75%정도가 사라졌기 때문에 일어나는 현상이다. 이러한 경우에도 상관도는 9126 으로 계산되어진 임계값보다 크기에 워터마크의 존재를 확인할 수 있다. 그림 5(f)는 워터마크된 영상의 50%를 잘라낸 후의 추출 결과이다. 이 경우에도 그림 5(e)와 비슷하게 비트 오류는 크지만 상관도는 임계치보다 크기 때문에 워터마크가 있다고 할 수 있다.

(a) 원 영상 (b) 워터마크 영상

그림 4. 원 영상과 워터마크 영상



(a) No attack (b) Gaussian Noise (c) Median Filtering

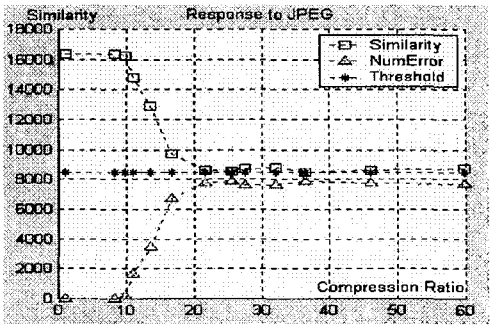
(d) JPEG (e) Resizing (f) Cropping

그림 5. 추출된 워터마크

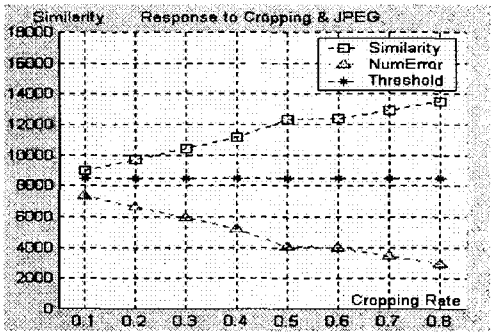
그림 6 은 실험 결과를 그림으로 정리한 것이다. 그림 6(a)는 여러 JPEG 압축율에 대해 추출된 워터마크의 상관도를 나타내고, 그림 6(b)는 Cropping 과 12 배의 JPEG 압축을 가하였을 때의 상관도를 나타낸다.

5. 결론

본 논문에서는 인간의 시각 특성을 이용한 새로운 워터마킹 방법을 제안하였다. 제안된 알고리즘은 워터마크 데이터와 사용자의 비밀키가 서로 연관이 없기 때문에 어떠한 종류의 데이터도 삽입할 수 있다. 워터마크의 존재 여부를 알 수 있는 객관적인 값인 상관도의 임계치도 확률적인 접근으로 제안하였다. 워터마크 추출실험에서는 여러 가지 고의적인 공격에 대해서 강인함을 보였고, Cropping 과 JPEG 을 동시에 적용하였을 경우에도 추출된 워터마크의 상관도가 임계값보다 높음을 확인하였다.



(a) JPEG 압축에 대한 상관도



(b) Cropping 과 JPEG 압축에 대한 상관도

그림 6. JPEG 과 Cropping 에 대한 실험결과

감사의 글

본 연구는 광주과학기술원(K-JIST) 초고속광네트워크연구센터(UFON)를 통한 한국과학재단 우수연구센터(ERC)와 교육부 두뇌한국 21(BK21) 정보기술사업단의 지원에 의한 것입니다.

참고 문헌

[1] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1673-1687, Dec. 1997.

[2] F. Hartung and M. Kutter, "Multimedia Watermarking Technique," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079-1107, Jul. 1999.

[3] B. Chiprasert and K. R. Rao, "Human Visual Weighted Progressive Image Transmission," *IEEE Transaction on Communications*, vol. 38, no. 7, pp. 1040-1044, Jul. 1990.

[4] A. Leon-Garcia, *Probability and Random Process for Electrical Engineering*, Addison-Wesley Publishing Company, 1994.

[5] M. D. Swanson, Bin Zhu and A. H. Tewfik, "Data Hiding for Video-in-Video," *International Conference on Image Processing*, vol. 2, pp. 676-679, 1997.