

# 인간 시각 시스템을 이용한 반복적 워터마킹 기법

김희정, \*김지홍

부산외국어대학교 컴퓨터전자공학부, \*동의대학교 컴퓨터영상공학부

## An Iterative Watermarking Method Using Human Visual System

Hee-Jung Kim, \*Ji-Hong Kim

Div. of Comp. and Elect. Eng., Pusan Univ. of Foreign Studies

\*Div. of Comp. and Visual Eng., Dong-eui Univ.

### Abstract

In this paper, a new watermarking method using an iterative scheme is proposed. The proposed method consists of two parts: watermark generation and its insertion. In the watermark generation, random signals with normalized distribution  $N(0,1)$  are produced and modified using an iterative scheme. In the watermark insertion, the resulting watermark signals are inserted into selected transform coefficients with various scaling parameters in order to achieve the invisibility and robustness against illegal attacks. The simulation results show that the proposed method has good performance for various attacks including image compression, filtering, etc.

### I. 서론

최근 인터넷 활용의 증가와 컴퓨터, 스캐너, 프린트 등의 광범위한 사용으로 인해 디지털 데이터의 획득 및 교환이 점차 용이해짐에 따라 디지털 정보의 불법적 사용 기회가 확대되고 있다. 이에 따라 정보 보호에 대한 필요성이 절실히 요구되고 있으며, 이러한 정보 보호는 영상, 음성, 문자 정보 등에 대한 저작권 또는

지적 소유권을 보호하여 불법적인 복사와 분배를 제한하는 중요한 수단이 된다.

워터마킹 기법은 심층암호(steganography)로 알려진 데이터 은닉 기술과 유사성을 갖는 새로운 형태의 디지털 정보 보호 기법이다[1]. 워터마킹 기법은 워터마크가 삽입되는 영역에 따라 공간영역 기법과 변환영역 기법으로 분류할 수 있으며, 손실 및 무손실 압축, 필터링, 절단, 회전, 신축 등 다양한 신호처리 및 기하변환에 대해 견고하여야 한다. 그러나 영상의 특성과 적용되는 신호처리 기법의 특성을 고려하지 않고 공간영역 또는 변환영역에서 처리할 경우, 워터마크의 식별 또는 제거가 비교적 쉽게 이루어질 수 있다.

본 논문에서는 인간 시각 시스템을 고려한 반복적 워터마킹 기법을 제안한다. 제안된 기법은 손실압축과 필터링 등에 대해 견고하며, 시각적으로 인지가 용이하지 않은 특징을 가진다. 이를 위해 워터마크의 생성 및 삽입 위치 설정 과정에서 POCS (Projection Onto Convex Set) 이론이 적용되며, 또한 영상을 구성하는 각 블록의 특징에 따라 세 종류의 스케일링 파라미터를 사용한다.

본 논문의 II 장에서는 제안된 방법에 대해 기술하며 III 장에서는 모의 실험을 수행하여 성능을 분석한다. IV 장의 결론으로 논문을 맺는다.

## II. 반복 기법을 이용한 변환영역 워터마킹

제안된 워터마킹 기법은 다음과 같이 이루어진다. 워터마크로 사용될 정규 분포의 랜덤 신호를 변환 블록의 중간 주파수 영역의 변환계수로 설정하여 역변환 과정을 수행한다. 역변환된 워터마크 영상에서 평탄영역 등 시각적으로 인지하기 쉬운 영역에 위치한 워터마크를 삭제한 후, 다시 변환을 수행하여 워터마크 삽입 영역의 변경된 워터마크를 초기의 워터마크로 복원한다. 이러한 과정을 반복 실행한 후 최종적으로 생성된 변환영역 워터마크를 입력영상의 변환계수에 대해 삽입하고, 이어서 역변환을 실행함으로써 워터마크된 영상을 생성한다. 입력영상의 변환계수에 대해 워터마크를 삽입하는 과정에서는 먼저 입력영상을 서로 중복되지 않는 블록으로 분할한 후, 각 블록을 에지 화소의 밀도에 따라 세 종류로 분류한다. 그리고 각 종류에 대해 서로 다른 스케일링 파라미터를 적용한다. 삽입된 워터마크의 검출을 위해서는 훼손된 워터마크된 영상과 입력 영상을 각각 변환한 후, 이들로부터 추출된 워터마크와 초기에 사용된 워터마크의 상관성을 검사하는 방식을 사용한다.

### 2.1 블록의 분류

삽입된 워터마크의 가시성은 배경(입력영상)의 명도(luminance), 공간 주파수, 무늬(texture) 등에 의해 다음과 같이 영향을 받는다[2].

- 배경이 밝을수록 삽입된 워터마크의 가시성은 낮아진다(luminance masking).
- 배경에 무늬가 강할수록 삽입된 워터마크의 가시성은 낮아진다(texture masking).
- 삽입된 워터마크의 가시성에 대한 공간주파수 분포의 영향은 CSF(contrast sensitivity function)에 의해 모델링된다(frequency masking).

본 논문에서는 무늬 마스크(texture masking) 효과를 고려하여 입력영상을 구성하는 블록들을 세 종류로 분류한다. 즉 입력영상  $f$  는 먼저 식 (1)과 같이 서로 중복되지 않는  $8 \times 8$  크기의 블록  $B_k$ ,  $k = 0, 1, \dots, K-1$  로

분할된다.

$$f = \sum_{k=0}^{K-1} Y B_k = \sum_{k=0}^{K-1} Y f_k(i, j), \quad 0 \leq i, j \leq 7 \quad (1)$$

일반적으로 영상의 블록들은 서로 다른 특징을 가지며, 이러한 특징은 삽입된 워터마크의 가시성에 영향을 미칠 수 있다. 비가시성이 저해되지 않는 범위에서 가능한 최대값의 워터마크를 삽입할 수 있도록 하기 위해 입력 영상의 블록들을  $G_1, G_2, G_3$  의 세 종류로 분류한다.  $G_1$  은 약한 무늬를 가진 블록이며,  $G_2, G_3$  은 각각 중간 정도 및 강한 무늬를 가진 블록이다. 이러한 분류는

$$B_k \in \begin{cases} G_1 & \text{if } \text{num}\{e(i, j)=1, (i, j) \in B_k\} < T_1 \\ G_2 & \text{if } T_1 \leq \text{num}\{e(i, j)=1, (i, j) \in B_k\} < T_2 \\ G_3 & \text{if } \text{num}\{e(i, j)=1, (i, j) \in B_k\} \geq T_2 \end{cases} \quad (2)$$

와 같이 에지(edge) 화소의 밀도에 따라 이루어진다. 식 (2)에서  $\text{num}\{\}$  는 괄호의 조건을 만족하는 화소들의 수를 나타내며,  $e(i, j)$  는 입력 영상  $f$  에 대한 이진 에지 배열이다. 그리고  $T_k, k=1, 2$  는 문턱치를 나타낸다.

### 2.2 POCS 이론을 이용한 워터마킹

POCS 이론은, 두 컨벡스 집합의 교집은 초기 추정치로부터 두 집합 중의 한 집합으로 투영을 한 후, 이어서 다른 집합으로의 투영을 반복하면 구할 수 있음을 나타낸다. 본 논문에서 사용되는 첫번째 컨벡스 집합 TCS(transform coefficients set)는

$$TCS = \{w \mid W(u, v) = N(0, 1), \forall (u, v) \in \mathbf{R}_M\} \quad (3)$$

로 정의된다. 여기서  $w$  는 워터마크 영상을,  $N(0, 1)$  은 평균 0 과 분산 1 의 정규분포를 갖는 랜덤 신호를 각각 의미하며,  $\mathbf{R}_M = \{(u, v) \mid 3 \leq u+v \leq 6, u, v=0, 1, 2, \dots, 7\}$  을 나타낸다. 그리고  $F_w$  는 워터마크 영상의 변환계수 블록을 의미한다. 두번째 컨벡스 집합 RSS(region of support set)는

$$RSS = \{w \mid w(i, j) = 0, \forall (i, j) \in \mathbf{E}_0\} \quad (4)$$

이며, 여기서  $\mathbf{E}_0 = \{(i, j) \mid e(i, j) = 0\}$  을 의미한다. 컨벡스 집합을 이와 같이 설정한 이유는 TCS 의 경우 DCT 변환영역의 중간 주파수영역에 워터마크로 사용될 랜덤 신호를 삽입함으로써 워터마크된 영상의 화질이 크게

훼손되지는 않으며, 동시에 JPEG 등의 손실압축에 대해 워터마크가 견고하도록 하기 위함이다. 그리고 RSS는 입력 영상에서 인지가 용이한 영역에는 워터마크가 삽입되지 않도록 함으로써 워터마킹된 영상에서 워터마크가 쉽게 노출되지 않도록 하기 위해서이다. 따라서 두개의 컨벡스 집합 TCS와 RSS를 반복적으로 투영할 경우, 공간영역에서는 에지 또는 무늬영역 등 워터마크의 삽입으로 인한 시각적 왜곡효과가 적은 부분에 위치하며, 주파수영역에서는 중간 주파수영역에 분포함으로써 다양한 손실압축 및 저역·고역 필터링 등의 공격에 대해 견고성을 갖는 워터마크를 생성할 수 있다.

이상에서 정의된 컨벡스 집합을 이용하여 워터마크를 생성하는 과정은 다음과 같다. 먼저 초기의 워터마크로 사용하기 위해 평균이 0 이고 분산이 1 인 랜덤 신호  $N(0,1)$ 을 생성한 후 이를 변환 블록 중간 주파수영역  $\mathbf{R}_M$ 의 변환계수로 설정한다. 그리고 그 외의 영역에 대한 변환 계수는 0 으로 설정한다. 이와 같이 설정된 초기 워터마크 블록을 역방향 DCT 변환한 후, 입력 영상의  $E_0$  영역에 위치한 복원된 워터마크 신호는 제거한다. 즉 에지 또는 무늬 등 잡음으로 인한 시각적 왜곡 효과가 작은 부분에 위치하는 값은 그대로 둔 채, 평탄 영역 등 잡음에 민감한 부분에 위치한 값은 0 으로 삭제한다. 이와 같은 과정은 RSS에 대한 투영을 의미하며, 본 논문에서는 영역 제로화(region zeroing) 과정이라고 한다. 이어서 순방향 DCT를 수행하여 결과로 주어진 워터마크의 변환 계수 블록에서 중간 주파수영역은 초기에 사용된 워터마크 신호를 재 입력하며 그 외의 주파수 영역에 위치한 변환계수는 그대로 둔다. 이러한 과정을 본 논문에서는 워터마크 재입력(watermark re-insertion) 과정이라고 한다. 영역 제로화 과정과 워터마크 재입력 과정을 추출된 워터마크가 수렴할 때까지 반복한다. 워터마크 신호의 수렴을 결정하는 방법은 2 가지로 분류할 수 있다. 먼저 워터마크 재입력 과정에서 변화하는 워터마크 신호의 양이 일정한 기준 이하이면 수렴하는 것으로 판단하는 방법이 있고, 또 다른 방법으로는 반복 횟수를 임의로 설정하는 방식이 있다.

### 2.3 워터마크의 삽입 및 추출

입력 영상을  $8 \times 8$  크기의 블록으로 분할한 후 각 블록에 대해 DCT 변환한 후 위의 과정에서 생성된 변환영역 워터마크를 삽입하며, 삽입된 계수  $F_w(u, v)$ 는

$$F_w(u, v) = F(u, v) \{1 + \alpha_k \cdot W(u, v)\}, \quad k = 0, 1, 2 \quad (5)$$

와 같다. 식 (5)에서  $F(u, v)$ 와  $W(u, v)$ 는 각각 입력 영상의 변환 계수와 변환 영역 워터마크를 의미하며,  $\alpha_k$ 는 워터마크 신호에 대한 가중치를 나타낸다. 그리고 워터마크가 삽입된 변환 계수  $F_w(u, v)$ 를 역방향 DCT 변환함으로써 워터마킹된 영상을 생성한다.

워터마크의 추출 과정에서는 입력 영상이 사용되며, 삽입 과정의 역순과 유사하게 진행된다. 먼저 시험 영상과 입력 영상을 각각 역변환하여 시험 영상에 포함되어 있는 훼손된 워터마크를 추출한다. 그리고 워터마크를 생성하기 위해 사용되었던 랜덤 신호와 추출된 워터마크의 상관성을 검사한다. 즉 시험 영상의 변환 계수를  $F_w^*(u, v)$ 로 나타낼 때 추출된 워터마크  $W^*(u, v)$ 는

$$W^*(u, v) = \frac{F_w^*(u, v) - F(u, v)}{\alpha_k \cdot F(u, v)}, \quad (u, v) \in \mathbf{R}_M \quad (6)$$

이며, 추출된 워터마크와 초기 워터마크의 상관성을 검사하기 위해 정규화된 상관함수(normalized cross correlation)

$$NC(W, W^*) = \frac{\sum_{(u, v) \in \mathbf{R}_M} W(u, v) W^*(u, v)}{\sum_{(u, v) \in \mathbf{R}_M} [W(u, v)]^2} \quad (7)$$

를 사용한다. 그리고 만일  $NC(W, W^*) > T$ 이면 시험 영상에 워터마크가 존재하는 것으로 판단한다.

## III. 모의 실험

모의실험을 위해 정규 분포  $N(0,1)$ 을 갖는 2000 개의 랜덤 신호 열을 생성하였으며, 이 중 1000 번째 열이 워터마크로서 사용하였다. 모의실험 결과들을 그림 1 과 2 에 나타내었다. 그림 1 은 시험 영상 Lena 와 이에 대한 워터마킹된 영상을 나타낸다.

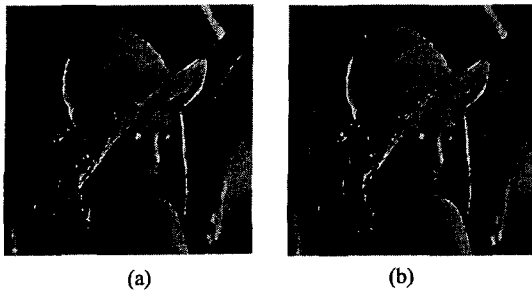


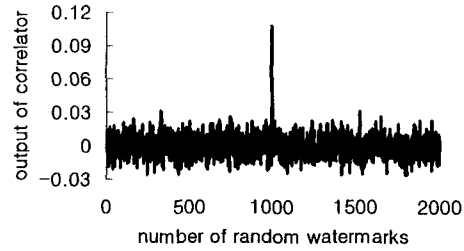
그림 1. (a) 시험 영상 (b) 워터마킹된 영상

그림 1 (b)에서 볼 때 워터마크가 에지와 무늬 등의 영역에 삽입되었고 특히 에지 화소의 밀도에 따라 가변적으로 스케일링 파라미터가 적용되었기 때문에 워터마크가 시각적으로 쉽게 인지되지 않음을 볼 수 있다. 그리고 워터마킹된 영상에 대한 워터마크 검출기의 응답은 그림 2와 같다. 그림 2 (a)는 JPEG 방식을 이용하여 손실 압축된 워터마킹된 영상에 대한 워터마크 검출기의 응답 특성이다. 이 그림에서 볼 때 워터마킹된 영상이 심하게 훼손된 경우에도 워터마크는 견고하게 검출됨을 볼 수 있다. 그림 2 (b)는 그림 1 (b)의 워터마킹된 영상에 대해 3x3 크기의 지역통과 필터를 적용했을 때의 검출기의 응답을 나타낸다.

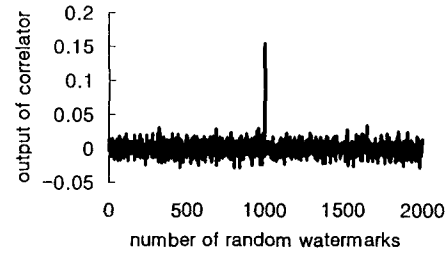
이러한 실험 결과들에서 볼 때 제안된 워터마킹 방법은 압축, 필터링 등의 다양한 영상 훼손에 대해 견고함을 알 수 있다.

#### IV. 결론

본 논문에서는 워터마크가 시각적으로 쉽게 인지되지 않고 다양한 공격에 견고한 새로운 변환영역 워터마킹 기법을 제안하였다. 제안된 기법에서는 변환 블록의 중간 주파수영역에 워터마크를 삽입함으로써 다양한 영상 훼손에 대해 견고성을 가지며, 동시에 공간영역에서는 에지 또는 무늬 부분에 워터마크가 집중되도록 함으로써 워터마크의 인지가 거의 불가능하다. 모의 실험을 통해, 제안된 방법은 손실압축, 지역통과 필터링 등의 다양한 공격에 대해 견고함을 볼 수 있었다.



(a)



(b)

그림 2. 워터마크 검출기의 응답 특성 (a) JPEG 압축된 워터마킹된 영상에 대한 응답 (b) 지역통과 필터링된 워터마킹된 영상에 대한 응답

#### 참고 문헌

- [1] Mitchell D. Swanson, Mei Kobayashi, and Ahmed H. Tewfik, "Multimedia Data-Embedding and Watermarking Techniques," *Proc. of IEEE*, Vol. 86, No. 6, pp.1064-1087, June 1998
- [2] J. Huang, Y. Q. Shi, and Y. Shi, "Embedding Image Watermarks in DC Components," *IEEE Trans. on Cir. and Sys. for Vid. Tech.*, Vol. 10, pp.974-979, Sep. 2000
- [3] Ingemar J.Cox, Joe Kilian, Thomson Leighton, and Talal Sharnoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. on Image Proc.*, Vol. 6, No. 12, Dec. 1997
- [4] Jiwu Huang and Yun Q. Shi, "Adaptive Image Watermarking Scheme Based on Visual Masking," *Electronics Letters*, Vol. 34, No. 8, pp.748-750, Apr. 1998