

## 2단계 참조와 국부 상관성을 이용한 워터마킹 기법

조익환 , \*김태희 , \*\*정동석  
인하대학교 전자공학과

# Watermarking Method Using 2-Step Reference and Local Correlation

Ik-Hwan Cho , \*Tae-Hee Kim , \*\*Dong-Seok Jeong

Dept. of Electronic Engineering, Inha University

E-mail : cho8629@dreamwiz.com , \*g1982559@inhavision.inha.ac.kr , \*\*dsjeong@inha.ac.kr

### 요 약

본 논문에서는 2단계 참조를 통하여 워본 워터마크에 보다 근접한 워터마크를 추정해내는 워터마킹 방법을 제안한다. 이 방법은 워터마크 추출시 워본 영상을 필요로 하지 않으며 워터마크는 이진 신호가 된다. 여기서 참조 워터마크는 워터마크의 왜곡 정도를 알려주는데 이것을 가지고 좀 더 정확한 워터마크를 추정해 낼 수 있게 된다.

본 논문에서는 또한 추출된 워터마크와 원본 워터마크를 비교할 때 좀 더 정확한 비교를 위해서 국부 상관성을 이용한 방법을 제안한다. 실험 결과 공격의 특성에 구애받지 않고 일정한 성능을 보여줄을 확인 할 수 있었다.

### I. 서론

인터넷 및 통신수단의 발전에 힘입어 우리는 엄청난 양의 정보를 매일 매시간 접하고 있다. 이러한 정보들은 얼마 전까지만 해도 대부분이 아날로그 방식이었지만 현재는 거의 대부분이 디지털 형태로 배포되고 있는 실정이다. 이러한 디지털 형태의 정보들은 우리가 배포하고 관리하기가 간편하고 체계적이어서 그 사용빈도는 갈수록 높아질 전망이다. 그러나 이러한 디지털 정보는 배포 및 관리의 편리성뿐만 아니라 쉽게 복제할 수 있다는 성질을 갖는다. 이는 또 하나의 편리함을 제공할 수 있다고도 말할 수 있지만 다른 한편으로는 무제한적으로 저작권이 침해될 수 있다는 것을

의미하기도 한다. 실제로 현재 인터넷에 있는 영상을 무단으로 사용하거나 음악 파일을 무단으로 복제하여 사용하는 경우도 있다. 이러한 디지털 정보의 저작권 보호를 위한 수단으로서 디지털 워터마킹(Digital Watermarking)기술이 있다. 워터마킹 기술은 디지털 정보 내에 인간이 구별하거나 파괴할 수 없도록 저작권 정보를 삽입하여 만일 분쟁이 발생하였을 경우 디지털 정보에서 워터마킹 정보만을 추출하여 원래 워터마크와 비교하여 저작권 문제를 해결할 수 있도록 해준다. 디지털 정보에 워터마크를 삽입하여 저작권을 보호하려는 워터마킹 기술은 I.J. Cox<sup>[1]</sup>의 Spread Spectrum 이론에서 보듯이 채널 상에서 정보를 주고받는 통신과 유사하다. 일단 워터마킹에서의 삽입되는 워터마크를 통신상에 전달되는 정보라 하면 워터마크가 삽입되는 디지털 정보는 통신상의 채널이 된다. 이렇게 워터마크가 삽입된 디지털 정보가 외부의 어떠한 공격(Attack)을 받아 왜곡이 일어나는 것은 통신상에서 잡음이 첨가되어 전달되는 정보에 왜곡이 생긴 현상에 비유할 수 있다. 이렇듯 워터마킹은 채널 상에서 잡음이 첨가된 정보를 정확히 수신하는 것으로 볼 수 있다. 그러므로 채널의 수신 측에서 정보를 수신할 때 정보의 어느 부분에 잡음이 첨가되었는지 알 수 있다면 좀 더 정확한 정보를 추정해 낼 수 있을 것이다. 따라서 참조 워터마크를 분석하여 워터마크의 어느 부분이 공격을 받았는지 정확한 워터마크를 추정할 수 있다. 이러한 참조 워터마크 개념은 Deepa Kundur<sup>[2],[3]</sup>가 주장한 이론이며 우리는 이 이론을 발전시켜 참조 워터마크를 2단계로 이용하여 성능을 높이도록 하였다.

## II. 워터마크 삽입

본 논문에서는 웨이블릿 변환(Wavelet transform)과 양자화 삽입(Quantization embedding)방법을 이용하여 주 워터마크(Main Watermark)와 참조 워터마크(Reference Watermark)를 영상에 삽입하고자 한다. 여기서 워터마크를 삽입할 때 고유의 웨이블릿 변환 계수와 양자화에 사용될 양자화 단계값  $Q$ 가 키(Key)값으로 사용되고 참조 워터마크는 일반에게 공개된다. 여기서 워터마크가 삽입될 영상을  $f$  라 정의한다.

### 2.1 웨이블릿 변환

먼저 워터마크를 삽입할 영상을  $L$  단계로 웨이블릿 변환을 취한다. 웨이블릿 변환을 취하면 각 단계마다 LH,HL,HH 3개의 영상이 생성되고 마지막으로 저주파 영역(LL)의 영상이 1개 존재하게 된다. 따라서 저주파 영역의 영상을 제외하면  $3L$  개의 고주파 영역이 생성된다. 여기서 우리는  $f_{v,i}(m,n)$ 을  $i$ 번째 단계의 3개의 LH,HL,HH 영상 중  $v$ 번째 영상에서  $(m,n)$ 위치의 픽셀값으로 정의하기로 한다.

### 2.2 워터마크 영역(Watermark Region) 형성

본 논문에서 사용할 워터마크는 0과 1로 구성된 이진 신호로 구성된다. 먼저 삽입할 주 워터마크와 참조 워터마크를 하나의 워터마크 영역으로 만든다. 주 워터마크의 길이를  $N_w$ 라 하고 참조 워터마크의 길이를  $N_r$ 라 하고 두 워터마크는 무작위 수열로 생성시킨다. 일반적으로  $N_w$ 와  $N_r$ 은 같은 값으로 한다. 주 워터마크와 참조 워터마크를 그림1과 같이 한 비트마다 서로 번갈아 가며 배열하여 마치 바둑판 같은 워터마크 영역을 형성한다.

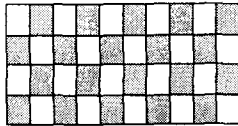


그림 1. 워터마크 영역

이와 같이 워터마크 영역을 형성하는 이유는 주 워터마크가 받게 될 왜곡과 같은 영향이 참조 워터마크에도 미칠 수 있도록 하기 위해서이다. 하나의 주 워터마크는 주변에 참조 워터마크를 가지게 되어 주 워터마크에 왜곡이 일어나면 참조 워터마크에도 왜곡이 일어날 가능성이 높아지게 된다. 그리하여 후에 워터마크 추출 시 참조 워터마크가 왜곡된 정도를 분석하여 그 정보를 주 워터마크 추출 시에 참조하게 된다.

### 2.3 양자화 워터마크 삽입(Quantization Watermark Embedding)

{0,1}의 이진 신호인 워터마크는 양자화 삽입 방법을 사용하여 웨이블릿 변환 상에서 웨이블릿 계수에 삽입

한다. 먼저  $f_{v_1,i}(m,n)$ ,  $f_{v_2,i}(m,n)$ ,  $f_{v_3,i}(m,n)$ 을 오름차순으로 정렬하고 정렬된 모습은 다음과 같이 표현 할 수 있다.

$$f_{v_1,i}(m,n) \leq f_{v_2,i}(m,n) \leq f_{v_3,i}(m,n) \quad (1)$$

여기서  $v_1, v_2, v_3 \in 1, 2, 3$  이고  $v_1 \neq v_2, v_2 \neq v_3, v_3 \neq v_1$ 이다.

삽입될 워터마크의 한 비트는 위에서 열거한 3개의 값 중 중간 값을 변화시켜 삽입하게 된다. 한 개의 워터마크 비트를 삽입하기 위해  $f_{v_1,i}(m,n)$ 과  $f_{v_3,i}(m,n)$  사이를 미리 정해진  $Q$ 값으로 양자화한다. 그림2에 양자화 워터마크 삽입과정을 도시하였다.



그림 2. 양자화 워터마크 삽입과정

여기서 양자화된 각 단계의 값  $\delta$ 는 다음과 같이 구할 수 있다.

$$\delta = \frac{f_{v_3,i}(m,n) - f_{v_1,i}(m,n)}{2Q - 1} \quad (2)$$

만일 삽입하고자 하는 워터마크 비트 값이 0이면 중간 값인  $f_{v_2,i}(m,n)$ 값을 그 값에서 가까운 0의 양자화 값으로 바꾼다. 또 워터마크 비트 값이 1이라면 그 값에서 가까운 1의 양자화 값으로 바꾼다. 이렇게 중간 값을 양자화 값으로 바꿈으로써 한 개의 워터마크 비트 삽입이 이루어지게 된다. 이와 같은 방법으로 워터마크 영역 전체를 삽입하고 또 이 워터마크 영역을 저주파 영역만을 제외한 영상의 전 영역에 연속적으로 삽입하게 된다. 저주파 영역에 워터마크를 삽입하게 되면 워터마크에 의한 영상의 왜곡이 커지기 때문에 제외한다. 위와 같이 워터마크를 삽입하여 역 웨이블릿 변환을 취하면 워터마크가 삽입된 영상을 얻을 수 있다.

## III. 워터마크 추출

본 논문은 워터마크 추출 시에 원본 영상을 필요로 하지 않기 때문에 주어진 영상으로부터 키 값만을 가지고 워터마크를 추출 할 수 있다. 여기서 워터마크를 추출할 영상을  $g$  라 정의한다.

### 3.1 웨이블릿 변환

먼저 삽입과정과 마찬가지로  $L$ 단계로 웨이블릿 변환을 수행한다. 이 때 사용하는 웨이블릿 계수는 삽입 시에 사용했던 계수를 사용한다.

### 3.2 양자화 워터마크 추출(Quantization Watermark Extraction)

삽입 시와 마찬가지로  $g_{v_1,l}(m,n), g_{v_2,l}(m,n), g_{v_3,l}(m,n)$ 를  
오름차순으로 배열한다.

$$g_{v_1,l}(m,n) \leq g_{v_2,l}(m,n) \leq g_{v_3,l}(m,n) \quad (3)$$

여기서  $v_1, v_2, v_3 \in 1, 2, 3$  이고  $v_1 \neq v_2, v_2 \neq v_3, v_3 \neq v_1$ 이다.  
위터마크 삽입 시에 사용했던 Q값을 이용하여  
 $g_{v_1,l}(m,n)$ 과  $g_{v_3,l}(m,n)$ 사이를 양자화 한 다음 중간 값인  
 $g_{v_2,l}(m,n)$ 이 0의 양자화 값과 1의 양자화 값 중 0의 양  
자화 값에 가까우면 0으로, 1의 양자화 값에 가까우면  
1로 위터마크 비트를 추출한다. 이러한 방식으로 삽입  
시에 사용했던 위터마크 영역별로 위터마크를 추출해  
낸다.

### 3.3 1단계 위터마크 참조를 이용한 추정

참조 위터마크를 이용함에 있어서 우선 위터마크 과  
정을 이진 대칭 채널(Binary Symmetric Channel, BS  
C)로 가정한다. 위터마크 자체가 이진 신호인 것도 마  
찬가지 원리이다. BSC에서 비트 오류 확률을  $P_E$ 라고  
하면 정상적으로 신호가 전송될 확률은  $(1-P_E)$ 가 된다.  
영상에 위터마크 영역이 M번 삽입되었다면 원래의 주  
위터마크와 참조위터마크는  $w_k(i), r_k(i)$ 로, 추출된 주  
위터마크와 참조 위터마크는  $\hat{w}_k(i), \hat{r}_k(i)$ 로 표현가능하  
고 여기서  $k=1,2,3,\dots,M, i=1,2,3,\dots,N_w$  이다. 참조 위터  
마크는 공개되어 있으므로 두 개의 참조 위터마크를  
이용하여 하나의 위터마크 영역에 대한  $P_{Ek}$ 를 다음과  
같이 구할 수 있다.

$$P_{Ek} = \frac{1}{N_r} \sum_{i=1}^{N_r} r_k(i) \oplus \hat{r}_k(i) \quad (4)$$

이렇게 참조 위터마크를 통하여 구한 비트 오류율이  
주 위터마크에도 동일하게 적용이 된다고 가정하면 각  
각의 위터마크 영역에 대하여  $P_{Ek}$ 를 통해서 왜곡의 정  
도를 알 수 있다. 따라서 왜곡이 적은 부분에서 추출  
한 주 위터마크에 더 높은 가중치를 두어 최종적으로  
주 위터마크를 추정해 낼 수 있다. 적용될 가중치는  
다음과 같다.

$$\alpha_k = \frac{\log\left(\frac{1-P_{Ek}}{P_{Ek}}\right)}{\sum_{j=1}^M \log\left(\frac{1-P_{Ej}}{P_{Ej}}\right)} \quad (5)$$

최종적으로 추정해낸 위터마크는 다음과 같다.

$$\hat{w}(i) = \text{round} \left[ \sum_{k=1}^M \alpha_k \hat{w}_k(i) \right] \quad (6)$$

$\hat{w}(i)$ 가 0.5보다 클 경우 1로 0.5보다 작을 경우 0으로  
추정해 낸다.

### 3.4 제안된 2단계 위터마크 참조를 이용한 추정

1단계 위터마크 참조에서는 위터마크 영역의 각각의

비트 오류율을 구하여 서로 다른 가중치를 부여하여  
주 위터마크를 추정해 내었다. 그러나 비트 오류는 각  
각의 위터마크영역 내에서도 서로 다르게 일어난다.  
예를 들어 위터마크 영역 내의 특정 비트에서 오류가  
일어나더라도 다른 위터마크 영역의 해당 비트에서는  
오류가 발생하지 않을 수도 있다. 이는 통신채널 상에  
서 정보가 전송될 때 시간에 따라 다르게 잡음이 첨가  
될 수 있다는 것을 의미한다. 우리는 기존에 1단계에  
서 구한 각 위터마크 영역의 가중치 값과 더불어 각  
위터마크 비트별로 구한 가중치를 추가한다.

각 위터마크 영역 내에서의 비트 오류율을 측정하기 위  
하여 우리는 위터마크 영역 내에서 각각의 주 위터마크  
를 둘러싸고 있는 4개의 참조 위터마크를 이용한다.

하나의 위터마크 영역 내에서 한 위터마크 비트의 오  
류율을 구하는 식은 다음과 같다.

$$P_{Ek}(i) = \frac{1}{4} \sum_{j=1}^4 r_{kj} \oplus \hat{r}_{kj} \quad (7)$$

주변의 참조 위터마크의 비트 오류율을 중앙의 주 위  
터마크에 동등하게 적용하면 각 위터마크 비트별로 다  
른 비트 오류율을 갖게되고 이것을 가지고 또 하나의  
가중치  $\beta_k(i)$ 를 구할 수 있다.

$$\beta_k(i) = \frac{(1-P_{Ek}(i))}{\sum_{j=1}^M (1-P_{Ej}(i))} \quad (8)$$

이 값을 가지고 구한 최종적인 위터마크 비트 값은 다  
음과 같다.

$$\hat{w}(i) = \text{round} \left[ \sum_{k=1}^M \beta_k(i) \hat{w}_k(i) \right] \quad (9)$$

이렇게 2단계를 거쳐 구한 최종 위터마크 값을 살펴보  
면 0과 1사이의 값이 되는데 0.5보다 클 경우 1로, 0.5  
보다 작을 경우 0으로 판단하게 된다. 값이 0.5근처라  
면 정확한 비트 값을 판단할 수 있는 확률이 떨어지게  
되는 것을 나타낸다. 따라서 2가지 위터마크 비트값  
중 0.5와의 차이가 큰 값을 선택하여 그 값을 가지고  
최종적인 위터마크 비트 값을 결정하게 된다.

### 3.5 국부 상관성을 이용한 유사도 판별

최종적으로 추출된 위터마크와 원본 위터마크의 유  
사도 판별을 통하여 진위 여부를 판단하게 된다. 일반  
적으로 이진 신호의 유사도 판별은 원본 위터마크와  
추출된 위터마크를 서로 비교하여 전체 비트 수에 대  
한 일치되는 비트 수의 비율을 가지고 판단한다. 그러  
나 이러한 방법은 각 비트 사이의 관계를 고려하지 않  
았기 때문에 정확한 유사도 판별을 하기 어렵다. 만일  
비교할 위터마크 신호로 각 비트가 모두 1인 신호를  
사용하더라도 원본 위터마크가 무작위(Random)로 만  
들어낸 이진 신호이기 때문에 50%에 가까운 일치도를  
보이게 된다. 따라서 우리는 각 비트 사이의 연속성을

고려하여 비트 하나의 일치여부를 판단하는 것이 아니라 2개의 비트가 연속해서 일치하는가를 판단하여 비트간의 연속성을 고려하도록 하는 변형된 유사도 판별식을 이용하였고 그 식은 다음과 같다.

$$P = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{1}{w(i) \oplus \hat{w}(i) \cdot w(i+1) \oplus \hat{w}(i+1)} \quad (10)$$

### VI. 실험결과

실험을 위한 영상은 512x512의 크기를 갖는 그레이 영상을 사용하였고 주 워터마크와 참조 워터마크 모두 1024개의 이진 무작위 수열을 사용하였다. 2개의 워터마크로 이루어진 워터마크 영역의 크기는 64x32로 고정시켰다. 실험에서 사용된 웨이블릿 변환은 Antonini<sup>[4]</sup>가 제시한 계수를 이용하였다. 시행한 웨이블릿 단계는 3단계로 하였다.

실험에서는 위의 조건을 따르는 워터마크를 생성하여 영상에 삽입하고 이 영상을 여러 가지 방법의 공격을 가한 다음 다시 워터마크를 추출하여 워터마크와의 유사도를 조사하였다. 이 결과를 Deepa Kundur의 방법을 사용한 결과와 비교하였다. 본 실험에서는 Mean Filtering을 취하는 것과 Uniform Noise를 첨가하는 것을 공격으로 가정하였다.

그림3, 그림4와 표1에서 보듯이 2단계의 참조 워터마크를 사용했을 경우 약 4~8%의 성능 향상을 보이는 것을 알 수 있다. 또한 추가로 국부 상관성을 이용하여 유사도를 판별하였을 경우 기존의 방법보다 약 10%이상의 성능향상을 보였다. 이는 본 논문에서 제시한 방법이 워터마크의 연속적인 일치율을 높여 좀 더 정확한 워터마크를 검출할 수 있도록 해준다는 것을 말해준다.

— Old Method      - - - Prop. Method  
 --- Old Method+Local Corr      ··· Prop. Method+Local Co

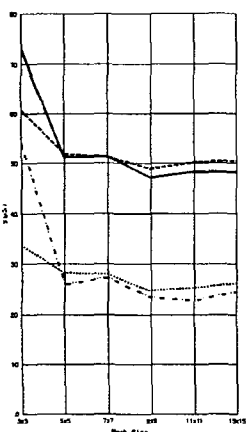


그림 3. Mean Filtering

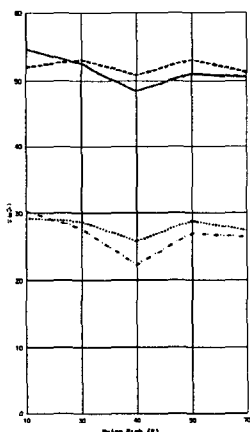


그림 4. Additive Noise

표 1. 공격 유형에 따른 유사도 (%)

	Old Method	Prop. Method	Old Method+Local Corr.	Prop. Method+Local Corr.
<b>mask size</b>	<b>Mean Filtering</b>			
3x3	72.94	60.74	54.05	33.72
5x5	51.36	52.05	26.00	28.25
7x7	51.46	51.49	27.37	28.01
9x9	47.16	48.92	23.36	24.73
11x11	48.33	50.19	22.6789	25.31
15x15	48.24	50.48	24.437	26.19
<b>noise prob. (%)</b>	<b>Additive Noise</b>			
10	54.5898	51.9593	30.30	29.22
30	52.4414	53.0273	27.63	28.73
40	48.4375	50.8789	22.28	25.80
50	51.0742	53.0273	26.97	28.83
70	50.5859	51.2695	26.39	27.37

### V. 결론

본 논문에서는 기존의 참조 워터마크 방법을 발전시켜 2단계로 참조 워터마크를 이용하고 유사도 판별시 국부 상관성을 고려하여 좀 더 정확한 워터마크를 추정해 내는 방법을 제안하였다. 하나의 참조 워터마크를 2단계로 이용하여 성능향상을 꾀할 수 있었고 공격의 특성에 구애받지 않고 일정한 성능을 보여주었다. 그러나 왜곡이 심하지 않을 경우에는 오히려 낮은 결과치를 보여주었고 일반적으로 무작위 신호를 워터마크로 사용하는 것보다 유사도가 떨어진다는 것이 단점으로 드러났다. 본 논문에서 사용한 워터마크 삽입 방법은 왜곡이 양자화 단계값(6)이상으로 커질 경우 불확실한 결과를 보여준다. 따라서 왜곡의 양을 정확히 알 수 있는 워터마크 삽입 방법과 함께 왜곡의 정도를 정확히 부여하는 가중치 식을 개발한다면 좀 더 나은 성능을 기대할 수 있을 것이다.

### VI. 참고 문헌

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

[2] Deepa Kundur and Dimitrios Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition", *In International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, volume 5, pp. 2969--2972, Seattle, Washington, USA, May 1998.

[3] Deepa Kundur, "Multiresolution Digital Watermarking : Algorithm and Implications for Multimedia Signals", *Graduate Department of Electrical and Computer Engineering, University of Toronto*, 1999.

[4] M. Antonini, "Image coding using wavelet transform", *IEEE Trans. Image Processing*, vol.1, no.2, pp.205-220, April 1992.