

# 저작권자의 음성 삽입을 위한 영상 워터마킹 방법

## An Image Watermarking Method for Embedding Copyrighter's Audio Signal

고 성 식\*, 최 재 승\*\*, 김 정 화\*\*\*

(Sung-Shik Koh\*, Jae-Seung Choi\*\*, Chung-Hwa Kim\*\*\*)

\*일본 오사카시립대학 공학연구과, \*\*경북대학교 디지털기술연구소, \*\*\*조선대학교 전자공학과

(접수일자: 2004년 2월 20일; 수정일자: 2004년 8월 2일; 채택일자: 2005년 3월 7일)

디지털 미디어와 통신 네트워크의 급속한 발전으로 지적소유권 보호 기술이 절실히 요구된다. 본 논문에서 영상컨텐츠에 소유자의 음성신호를 삽입할 수 있는 새로운 영상 워터마킹 방법을 제안한다. 제안한 방법은 삽입할 워터마크로 음성신호를 이용하기 때문에 청각적으로 소유권을 주장하는데 매우 유리하다. 그리고 LBX (Linear Bit eXpansion)에 의해 인터리빙하여 음성 워터마크를 삽입하기 때문에 공격에 의해 변형되거나 특히 제거된 음성신호를 복구할 수 있는 이점을 가진다. 본 방법은 다음 세가지 기본 절차를 포함한다. 첫째, 아날로그 소유자 음성 신호를 PCM에 의해 부호화하고 새로운 디지털 워터마크를 생성한다. 둘째, 제안한 LBX에 의해 음성 워터마크를 인터리빙한다. 마지막으로 영상 콘텐츠를 이산 Haar 웨이블릿변환 (DHWT) 하고 저주파 영역을 마강공간으로 하여 인터리빙된 음성워터마크를 삽입한다. 실험 결과 본 연구에서 제안한 소유자 음성신호의 워터마크 삽입방법은 기존 워터마크 정보보다 눈에 보이지 않게 많은 정보량을 삽입할 수 있고 표준영상압축방식인 JPEG 손실압축과 특히 영상의 일부가 제거되는 영상절단과 영상회전 공격에 대해 강인성을 강건히 할 수 있었다.

**핵심용어:** 음성신호, 영상 워터마킹, LBX 인터리빙, 웨이블릿변환, 강인성

**투고분야:** 음성처리 분야 (2.1)

The rapid development of digital media and communication network urgently brings about the need of data certification technology to protect IPR (Intellectual property right). This paper proposed a new watermarking method for embedding owner's audio signal. Because this method uses an audio signal as a watermark to be embedded, it is very useful to claim the ownership aurally. And it has the advantage of restoring audio signal modified and especially removed by image removing attacks by applying our LBX(Linear Bit-eXpansion) interleaving. Three basic stages of our watermarking include: 1) Encode analogue owner's audio signal by PCM and create new digital audio watermark; 2) Interleave an audio watermark by our LBX; and 3) Embed the interleaved audio watermark in the low frequency band on DHWT (Discrete Haar Wavelet Transform) of image. The experimental results prove that this method is resistant to lossy JPEG compression as standard image compression and especially to cropping and rotation which remove a part of image.

**Keywords:** Audio signal, Image watermarking, LBX Interleaving, Wavelet transform, robustness

**ASK subject classification:** Speech Signal Processing(2.1)

### 1. 서 론

멀티미디어 저작권 환경이 디지털 물결 속에 예측할 수 없을 만큼 빠른 속도로 변화하고 있다. 인터넷 보급의 급속한 성장과 멀티미디어 데이터의 압축 및 전송 기

술의 발전으로 대용량의 데이터를 고속으로 전송과 편집 그리고 저장할 수 있게 됨에 따라 정지 영상, 비디오 동 영상, 오디오 데이터에 대해 누구나 쉽게 이용하고 복사 및 변형이 가능하게 되었다. 그래서 전자 출판이나 디지털화된 미디어에 있어서 소유권 침해에 대한 저작권 보호 문제가 중요하게 여겨지고 있다. 이에 따라 데이터 암호화기법 (encryption method)들과 디지털 워터마킹 기법 (digital watermarking method)들이 연구되어 왔

다. 워터마킹은 저작자의 서명과 같은 워터마크 정보를 디지털 콘텐츠에 기록함으로써 이후 저작권 문제가 발생했을 경우 소유권을 입증할 수 있는 방법이다.

워터마크는 삽입할 때 보이는 (visible) 워터마크와 보이지 않은 (invisible) 워터마크가 있으며 워터마크를 추출할 때 원본 콘텐츠가 필요한 논 블라인드 (non-blind) 방식과 원본 콘텐츠를 사용하지 않는 블라인드 (blind) 방식이 있다. 또한 소유권 교착상태 (overship deadlock)를 해결을 위해 내장된 워터마크는 일반적 또는 고의적인 영상처리에 의해 제거되지 않아야 하는 것 등의 조건을 만족해야 한다.

기존 디지털 워터마킹에 대한 연구는 공간 영역 기술 [1,2]과 변환영역 기술[3-6]로 연구되고 있다. 변환식을 이용하지 않는 공간영역의 데이터에서 LSB (Least Significant Bit)를 임의로 선택하여 정보를 넣는 방법, 문서 출력시에 문자간의 간격을 조정하여 정보를 넣는 방법, 희미한 기하 패턴을 첨가하는 방법 등이 있으며 변환식을 이용한 변환영역으로 DCT, DFT, Wavelets등을 이용한 방법들이 있다.

현재 국내외에서 서비스되고 있는 워터마크의 정보량은 대부분 수비트에 불과해 디지털 콘텐츠 해킹에 적극적으로 대응하기보다 소유권을 주장하는 데 활용되고 있는 정도다. 정보 추출 시 원본이 필요한 논 블라인드 방식은 활용 범위가 제한적인 반면 보안성이 뛰어나고, 블라인드 방식은 응용분야가 넓은 반면 보안성과 견고성은 상대적으로 떨어지는 단점이 있다.

본 논문에서는 영상컨텐츠의 이산 Haar 웨이블릿변환 (DHWT)영역에 소유자 음성신호 삽입할 수 있는 디지털 미디어 저작권보호 알고리즘을 제시한다. 본 방법에서 워터마크는 정보량이 많고 청각적으로 소유권을 주장할 수 있는 음성신호를 이용하였고 1차원 음성신호 데이터를 3차원 음성워터마크로 변환한 후 인터리빙을 이용하여 웨이블릿 변환영역의 저주파 계수 부분에 삽입하였다. 그리고 강인성을 견고히 하기 위해 LBX 인터리빙을 하였고 음성신호 추출 시 논 블라인드 방식을 채택하였다.

## II. 음성 워터마크와 LBX 인터리빙

기존 워터마킹 시스템은 워터마크를 시각적으로 보이지 않게 삽입하기 위해 정보량을 제한하기 때문에 워터마크 정보의 종류 역시 수 비트의 시리얼 번호나 2진로고[7] 등과 같이 간단한 데이터를 주로 이용하였다. 이장에서는 정보량이 많은 소유자의 음성신호를 음성워터마크로 변환하여 영상의 이산 웨이블릿 변환영역에 강인성 있게 삽입하는 기술을 제안한다.

### 2.1. 음성신호 정보량

영상컨텐츠에 삽입할 소유자 음성신호의 정보량은 삽입될 영상의 크기와 웨이블릿 변환 레벨에 따라 결정된다. 제안한 방법의 음성워터마크의 데이터 양은 식 (1)과 같이 계산된다.

$$\text{음성신호 정보량} = \frac{2^n \times 2^n}{4^{l-0.5}} [\text{bits}] \quad (1)$$

여기서  $2^n \times 2^n$ 는 영상의 크기이고,  $l$ 은 영상의 웨이블릿 변환 레벨 수이다. 예를 들면 오리지널 영상의 크기가  $512 \times 512$ 일 때 1레벨 웨이블릿 변환하여 저주파 영역에 인터리빙된 음성워터마크를 삽입할 때 음성신호의 정보량은  $512 \times 512 / 4^{0.5} = 131,072 [\text{bits}]$ 가 된다.

### 2.2. 음성 신호의 음성 워터마크 변환

음성신호를 영상컨텐츠에 삽입하기 위해, 음성 입력장치를 통해 소유자의 아날로그 1차원 음성신호를 입력받아서 디지털 3차원 음성워터마크로 변환하는 방법을 제시한다. 실험에 이용된 음성신호는 한글발음 "본 영상물에 대한 저작권은 고성식에게 있습니다"를 발음한 아날로그 신호이고, 그 음성신호 파형은 그림 1과 같이 나타내었다.

소유자의 아날로그 음성신호를 영상컨텐츠에 직접 삽입할 수 없기 때문에 PCM방식으로 표본화, 양자화, 부

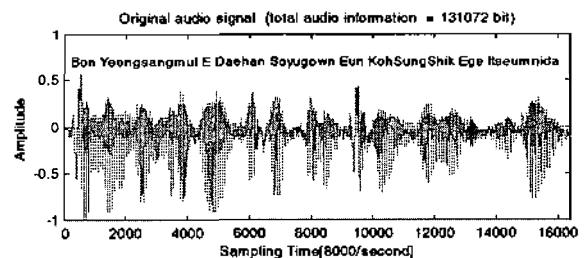


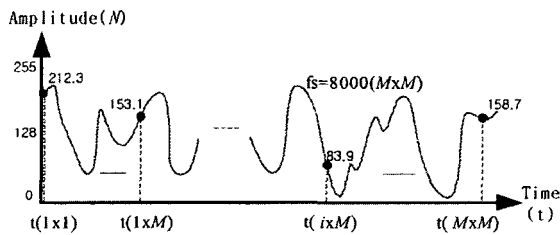
그림 1. 아날로그 음성신호 "본 영상물에 대한 저작권은 고성식에게 있습니다" 파형

Fig. 1. Wave of analog audio signal "Copyright of this image content belongs to Sung-Shik KOH"(in Korean).

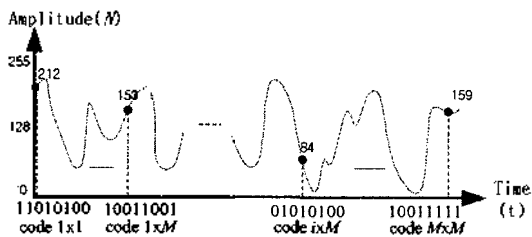
호화 변환과정을 거쳐 새로운 공간에서 음성워터마크를 생성한 후 삽입한다. 디지털 음성워터마크로 변환하는 과정은 그림 2와 같이 나타낼 수 있고 변환 절차는 다음과 같다.

- Step 1. 2-1절의 음성워터마크 정보량 계산식을 이용해 삽입 가능한 데이터양을 구한다.
- Step 2. 1차원 아날로그 음성신호를 8kHz 표본화 주파수로 16,384개의 샘플신호를 취한다.
- Step 3. 샘플당 8비트 레벨로 양자화 한다.
- Step 4. 양자화된 샘플신호 당 8비트로 부호화 하면 전체 16,384바이트가 얻어진다.
- Step 5. 부호화로 얻어진 비트열을 새로운 공간 (M×M×N)에 Raster Scan방식으로 배열한다.
- Step 6. 크기가 128×128×8 (M×M×N)인 3차원 음성워터마크를 얻는다.

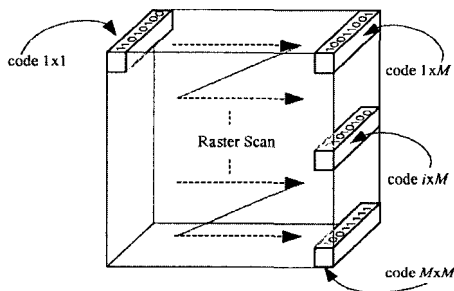
상기 절차에 의해 1차원 아날로그 음성신호는 디지털



(a) 아날로그 음성신호 샘플링과 양자화



(b) 부호화된 음성신호



(c) 음성 워터마크

그림 2. 음성워터마크 변환과정

Fig. 2. Transform sequence for audio watermark.

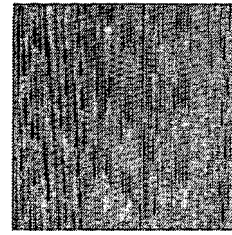


그림 3. 음성워터마크(크기128x128, 8bits/pixel)  
Fig. 3. Audio watermark.

신호 0과 1의 조합인 3차원 음성워터마크로 변환된다. 그림 3은 제안한 방법으로 구성된 음성워터마크를 나타낸다.

### 2.3. 음성 워터마크의 LBX 인터리빙

인터리빙은 통신용어로서 송신하는 반복되는 비트열을 섞어서 재배분하면 공격으로 인해 데이터 비트에 비트에러가 발생하더라도 인터리빙의 역과정 (디인터리빙)을 통하여 배열된 데이터를 원래 데이터 비트열로 배열하여 반복 패턴의 유무를 조사하여 손실된 비트열을 원래의 비트열로 복구할 수 있는 방식이다. 워터마킹 시스템에서 역시 외부 잡음이나 불법 디지털 저작물의 변형은 삽입된 음성워터마크 데이터에 비트 오류를 연속적으로 발생시킨다. 그래서 인터리빙을 하면 집중된 비트 오류가 발생하더라도 디인터리빙으로 집중된 비트 오류를 모두 분산시켜 독립된 오류로 바꿈으로써 오류 정정 부호의 효율을 높일 수 있다.

제한한 LBX 인터리빙은 선형 비트 확장을 이용해 음성워터마크를 인터리빙한다. 이는 비트열을 선형 n비트 확장하면 확장된 비트열은 일정한 패턴으로 규칙성을 가지고 반복하게 된다는 것을 이용한다. 이러한 원리로 음성워터마크를 선형 32비트 확장한 후 그림 4의 배열방법으로 재배치시켜 인터리빙한다. 음성워터마크  $W(x, y)$ 에 대한 화소의 최대값과 최소값이 각각  $W_{max}$ 와  $W_{min}$ 일 때, 선형 N비트 확장식은 다음과 같다.

$$W_N(x, y) = (2^N - 1) \left\{ \frac{W(x, y) - W_{min}}{W_{max} - W_{min}} \right\} \quad (2)$$

제한한 LBX 인터리빙 식을 얻기위해,  $W_{max}$ 을  $2^n - 1$ 로,  $W_{min}$ 을 zero로 각각 설정하면 식 (2)는 식 (3)으로 유도된다.

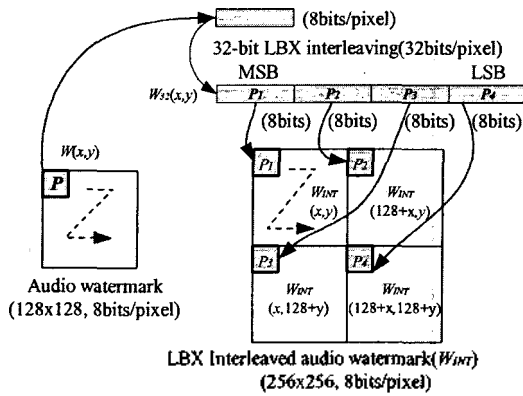


그림 4. 32비트 LBX 인터리빙 블록도  
Fig. 4. 32-bit LBX Interleaving block diagram.

$$\begin{aligned}
 W_N(x, y) &= \frac{(2^N - 1)}{(2^n - 1)} W(x, y) = \frac{(2^{nK} - 1)}{(2^n - 1)} W(x, y) \\
 &= [1 + 2^n + 2^{2n} + 2^{3n} + \dots + 2^{(K-1)n}] \cdot W(x, y) \\
 &= \left[ \sum_{i=1}^K 2^{(i-1)n} \right] \cdot W(x, y) \\
 &= R_K^n \cdot W(x, y), \text{ for } N = nK, K = 1, 2, 3, L.
 \end{aligned}
 \tag{3}$$

여기서  $R_K^n$ 는  $n$  비트 워터마크  $W(x, y)$ 에 대한 스케일 인자 (scale factor)가  $K$ 일 때의 반복인자 (repeat factor)이다. 이는  $n$  비트 워터마크 각각의 화소가  $N (=nK)$  비트로 척도변환됨을 의미한다. 만약 256 그레이 스케일 ( $n=8$ )인 음성 워터마크에 대해서 8비트열의 화소가 4번 반복 ( $K=4$ )되는 32비트의 결과를 얻고자할때, 반복인자  $R_4^8$ 는 10진수로 "16843009"이 되고 2진수로 32 비트열 ( $P_1P_2P_3P_4$ )인 "0000000100000001 ... 00000001"이 된다. 이를 이용해 LBX 인터리빙은 MSB에서 LSB까지 배열된 32비트로 확장된 음성워터마크 비트열을 그림 4와 같이 새로운 공간으로 각각 재배치시킨다. 따라서 인터리빙된 음성워터마크 ( $256 \times 256$ )는 1레벨 웨이블릿 변환의 저주파영역과 동일한 크기가 된다.

### III. 음성워터마크의 삽입과 추출

본 논문에서 제안하는 워터마킹 방법은 오리지널 영상 ( $X$ )을 DHWT하여 저주파영역을 마킹 공간으로 이용한다. 변환영역의 고주파 부분에 워터마크를 삽입하면 영상처리에 대해서 강인성이 약해지고, 저주파 부분에 음성워터마크를 삽입하면 영상의 화질 열화를 초래한다. 따라서 영상의 화질을 개선시키면서 강인성을 유지시키

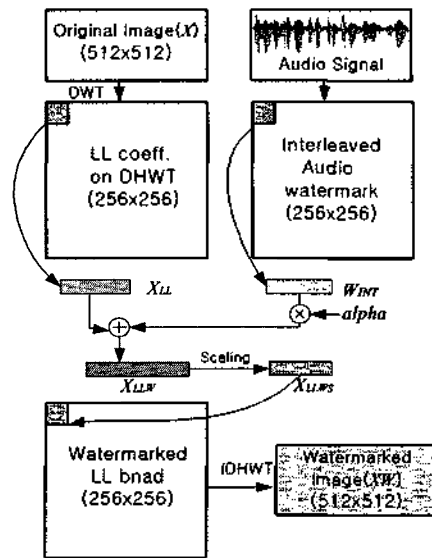


그림 5. 음성 워터마크 삽입 절차  
Fig. 5. Embedding sequency of audio watermark.

기 위해서 웨이블릿 변환영역의 저주파 영역부분 ( $X_{LL}$ )에만 LBX인터리빙한 음성워터마크 ( $W_{INT}$ )를 삽입한다. 이때 그림 5는 제안한 음성워터마크 삽입 절차를 나타낸다. 저주파 영역부분 ( $X_{LL}$ )은 가중치 ( $alpha$ )를 이용하여 식 (4)와 같이 음성워터마크와 삽입 ( $X_{LLW}$ )한 후, 식 (5)와 같이 삽입된 음성워터마크는 다시 오리지널 저주파 계수의 크기로 스케일링 ( $X_{LLWB}$ )한다.

$$X_{LLW}(x, y) = X_{LL}(x, y) + alpha \times W_{INT}(x, y) \tag{4}$$

$$X_{LLWB}(x, y) = scaling_{\min(X_{LL})}^{\max(X_{LL})} (X_{LLW}) \tag{5}$$

여기서  $alpha$ 는 워터마크의 삽입 강도를 결정하는 파라미터이고 시각적으로 화질열화를 인지할 수 없는 범위로 설정한다.

식 (4)에 의해 얻어진 웨이블릿 저주파 계수를 그대로 역웨이블릿 변환하여 워터마크가 삽입된 영상을 얻게 되면 저주파 계수의 변화로 인해 화질열화가 발생하게 된다. 따라서 워터마크된 영상의 화질을 최소화하기 위해 워터마크를 삽입한 저주파 계수를 삽입 전의 최대값과 최소값의 스케일로 식 (5)를 이용해 선형변환 하였다. 또한 워터마크를 추출할 때 역시 워터마크가 삽입된 저주파계수의 원래 크기로 선형확장변환할 때 이용된다. 이때 선형변환을 하기때문에 데이터 자체 변형은 없다. 그래서 워터마크의 변형은 없게된다. 이렇게 얻어진 음성 워터마크가 삽입된 계수 ( $X_{LLWB}$ )와 나머지 3개 주파수 영역을 이용해 역웨이블릿 변환 (IDHWT)하면 음성 워

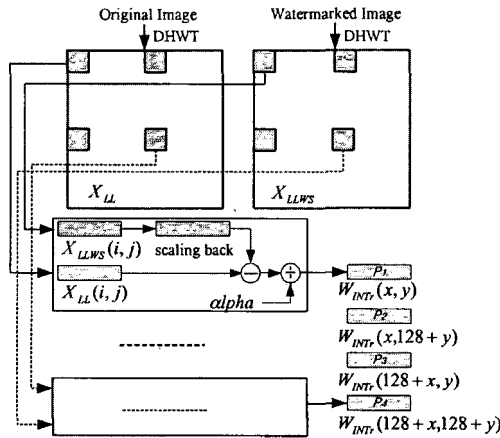


그림 6. 음성 워터마크 추출 절차  
Fig. 6. Extracting sequency of audio watermark.

터마크를 삽입한 영상 ( $XW$ )을 얻을 수 있다.

삽입된 음성워터마크의 추출은 그림 6과 같이 오리지널 영상을 필요로 한다.

추출순서는 음성워터마크를 삽입하는 절차의 역으로 처리하면 된다. 추출과정은 음성워터마크가 삽입된 영상과 오리지널 영상을 DHWT하여 저주파 영역부분의 계수를 각각 얻는다. 그리고 워터마크가 삽입된 저주파 계수는 삽입전의 스케일로 식 (6)와 같이 복원한다.

$$X_{LLW}(x, y) = scaling\ back \frac{\max(X_{LL}, \alpha)}{\min(X_{LL}, \alpha)} (X_{LLWS}) \quad (6)$$

다음으로 식 (7)을 이용해 삽입의 역 과정으로 가중치 ( $\alpha$ )를 이용해 두 영상 데이터간에 차를 구하면 인터리빙 시 재배치된 4곳으로부터 데이터 ( $P_1, P_2, P_3, P_4$ )을 얻을 수 있다.

$$W_{INT}(x, y) = \frac{X_{LLW}(x, y) - X_{LL}(x, y)}{\alpha} \quad (7)$$

추출된 데이터  $P_1, P_2, P_3, P_4$ 은 외부공격으로 인해 정보가 변형될 수 있기 때문에 최적의 워터마크 화소 선택은 제안한 역인터리빙 과정으로 그림 7과 같이 진행한다.

먼저 추출된 데이터  $P_1, P_2, P_3, P_4$ 의 평균 값 ( $m$ )을 이용해 식 (8)과 같이 평균편차 ( $D_i$ )를 구한다.

$$D_i = |P_i - m|, i = 1, 2, 3, 4 \quad (8)$$

그리고 식 (8)을 이용해 분산도를 나타내는 표준편차

( $\sigma$ )를 식 (9)와 같이 계산한다.

$$\sigma = \sqrt{\frac{1}{4} \sum_{i=1}^4 D_i^2} \quad (9)$$

정규분포의 확률곡선을 이용해 추출된 데이터  $P_1, P_2, P_3, P_4$ 의 오류 발생 유무를 결정하는 역치값 (*Threshold*)은  $m - \sigma$ 와  $m + \sigma$ 사이의 68.27%로 설정하였다. 만약 추출된 4개 데이터 ( $P_i, |i=1,2,3,4$ )의 평균 편차  $D_i$ 가 모두 역치값보다 적다면 최소 평균편차값 ( $\min(D_i)$ )을 갖는 데이터를 최적 워터마크 화소값으로 선택한다. 그러나 역치값보다 큰 편차가 존재하면 외부 공격에 의해 손실이 발생한 것으로 간주하여 역치값이상의 데이터를 제외한 나머지 데이터 ( $P_k, |k \leq i$ )로 다시 평균 편차 ( $D_k$ )를 구해서 최소 평균 편차값 ( $\min(D_k)$ )의 데이터를 최적 워터마크 화소값으로 선택한다. 그리고 그림 2의 역과정으로 복호화하면 음성신호를 추출할 수 있다.

복원된 음성신호는 고의적인 변형으로 인해 파형의 왜곡이 발생하게 된다. 따라서 최적 음성신호를 얻기 위해 후처리과정으로 FIR 저역통과 필터를 이용하였다. 그 선택은 다음과 같다.

Stopband edge frequency  $f_s$ : 2000 Hz  
 Passband edge frequency  $f_p$ : 1500 Hz  
 Sampling frequency  $f_t$ : 8000 Hz  
 Cut-off frequency :  $(f_s + f_p)/2 = 1750$  Hz  
 Passband gain :  $|G_p| = 1$  dB  
 Stopband attenuation :  $|G_s| = 50$  dB

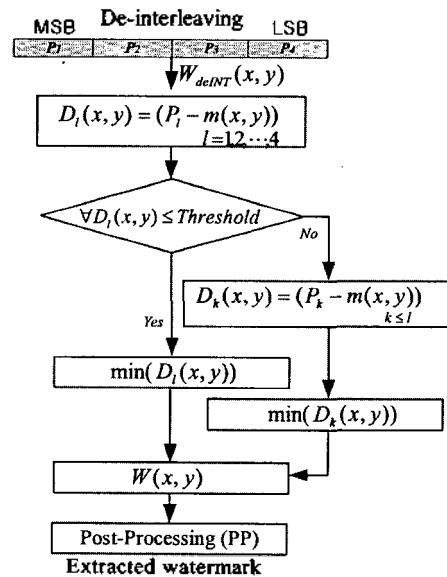


그림 7. 최적 음성 워터마크 추출을 위한 역인터리빙 과정  
Fig. 7. De-interleaving sequency for extracting optimal audio watermark.

본 연구에서 후처리과정은 필터 특성상 복원된 음성신호의 원신호의 고주파 파형보다 잡음성분의 고주파 파형이 많아지는 SNR 13dB이하에서만 적용하는 것으로 제한하였다.

### IV. 실험 및 결과고찰

본 장에서, 실험 결과를 통해 본 논문에서 제안한 워터마킹 방법의 강인성을 검증하고자 한다. 그림 8은 크기가 512×512인 "Lenna" 시험 영상이다.

본 연구에서 영상의 유사성 평가는 크기가 2"×2"이고 256 그레이레벨인 영상일 때 PSNR 식 (5)을 이용한다.

$$PSNR = 10 \log \left\{ \sum_{i=0}^{2^n-1} \sum_{j=0}^{2^n-1} \frac{255^2}{[X(i, j) - Y(x, y)]^2} \right\} \text{ [dB]} \quad (10)$$

여기서 여기서  $x(i, j)$ 는 원영상의  $i, j$  좌표의 화소값이고  $y(i, j)$ 는 평가 대상 영상의 화소값이다.

복원된 음성 품질 평가 방법은 두가지를 이용하였다. 먼저 식 (11)의 SNR기반 파형왜곡의 척도를 측정하는 객관적인 방법과 MOS (Mean Opinion Score)[8]기반 지각적 왜곡 척도를 측정하는 주관적인 방법으로 평가하였다.

$$SNR = 10 \log \left\{ \frac{\frac{1}{M} \sum_{n=1}^M S(n)^2}{MSE} \right\} \text{ [dB]}, \quad MSE = \frac{1}{M} \sum_{n=1}^M [y(n) - s(n)]^2 \quad (11)$$

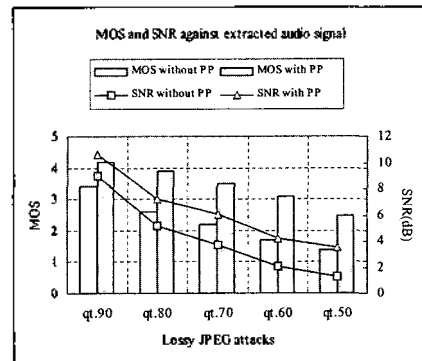
여기서  $y(n), s(n), M$ 은 각각 추출된 음성, 원음성, 샘플수 이다.



그림 8. "Lenna" 시험영상  
Fig. 8. "Lenna" test image (512×512).



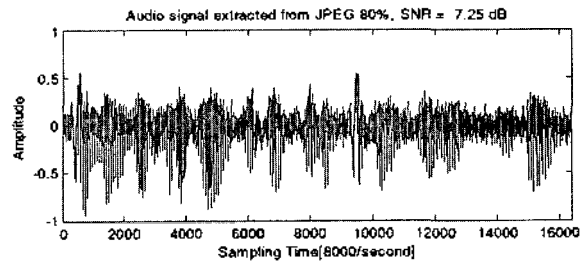
그림 9. 음성워터마크가 삽입된 영상  
Fig. 9. Image embedded with audio watermark. (alpha 0.095, PSNR 33.54dB)



(a) JPEG 공격에 대한 MOS와 SNR 비교



(b) JPEG qt. 80% 공격 영상(PSNR 31.54dB)



(c) JPEG qt. 80%에서 추출된 음성신호  
그림 10. JPEG압축공격과 복원된 음성신호  
Fig. 10. JPEG compression attack and reconstructed audio signal.

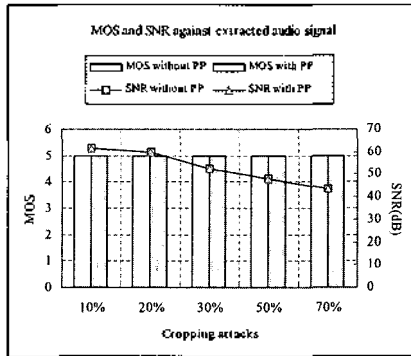
SNR은 단지 원음성과 왜곡된 음성 파형간의 대수적 차이를 나타내므로 음성의 지각적 특성변화 없이도 단순한 파형 왜곡으로도 SNR이 매우 저하되는 경우가 많기 때문에 음성의 지각적 특성을 평가하는 MOS방법을 동시에 이용하였다. 청취 모집단의 수는 남자 10명과 여자 10명으로 이루어졌고, 청취자의 반응 평가 시 5단계에 따라 5~1점의 점수를 주어 청취 시험에 참가한 20명 청취자에 의한 반응 의견의 평균치로 평가하였다.

그림 9는 본 논문에서 제안한 방법으로 가중치 alpha가 0.095일 때 그림 3의 음성워터마크를 그림 8의 시험 영상에 삽입한 영상이다. 오리지널 영상과 비교해볼 때 33.54dB의 화질을 유지하므로 시각적으로 화질 열화가 거의 없음을 알 수 있다.

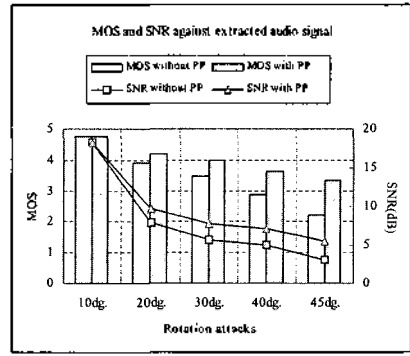
제안한 방법의 우수성을 평가하기 위해서 다양한 시험 영상을 손실 JPEG 압축, 영상절단 (Cropping), 영상회전 (Rotation) 공격으로 테스트를 하였다.

그림 10에서 그림 12는 "Lenna" 시험영상에 대한 시뮬레이션 결과를 나타낸다. 그림 10 (a)는 JPEG공격에 대한 MOS와 SNR비교 그래프이고, 그림 10 (b) 영상은 quality 80%의 공격으로 인해 31.54dB로 열화된 영상이고, 그림 10 (c)은 그림 10 (b)로부터 추출된 MOS 3.9의 음성신호이다.

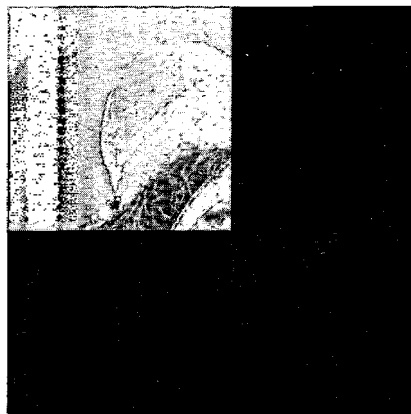
그림 11 (a)는 영상절단 공격에 대한 MOS와 SNR비교 그래프이고, 그림 11 (b) 영상은 영상절단공격으로 70%가 제거되어 7.10dB로 열화된 영상이고, 그림 11 (c)는 그림 11 (b)로부터 추출된 MOS 5의 음성신호이다.



(a) 영상절단 공격에 대한 MOS와 SNR 비교



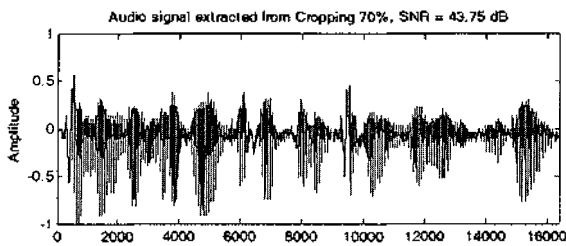
(a) 영상회전 공격에 대한 MOS와 SNR 비교



(b) 영상절단 70% 공격 영상(PSNR 7.10dB)

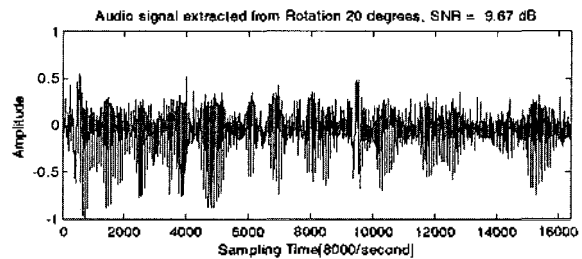


(b) 영상회전 20°공격 영상(PSNR 15.03dB)



(c) 영상절단 70%에서 추출된 음성신호  
그림 11. 영상절단 공격과 복원된 음성신호

Fig. 11. Cropping attack and reconstructed audio signal.



(c) 영상회전 20°에서 추출된 음성신호  
그림 12. 영상회전 공격과 복원된 음성신호

Fig. 12. Rotation attack and reconstructed audio signal.

그리고 그림 12 (a) 그래프는 영상회전 공격에 대한 MOS와 SNR비교 그래프이고, 그림 12 (b)는 영상회전 20도로 공격을 받았을때 일부가 절단되고 화소값이 변형되어 15.03dB로 열화된 영상이고, 그림 12 (c)은 그림 12 (b)부터 추출된 MOS 4.21dB의 음성신호이다.

시뮬레이션 결과, 복원된 음성신호는 손실 JPEG 압축 공격에서 MOS와 SNR모두 전반적으로 음질이 개선되었다. 그리고 영상절단 공격에서는 25%의 영상일부만으로도 음성신호를 모두 복원시킬수 있었고 음질 역시 최상이었다. 영상절단 공격에 대해서는 회전각도에 따라 비례해서 음성정보의 음질을 개선시킬 수 있었다.

### V. 결론

본 연구는 멀티미디어 영상 콘텐츠 소유권보호를 위해 콘텐츠 소유자의 음성신호를 이용해 저작권을 주장할 수 있는 워터마킹 방법이고, 강인성을 강건히 하기 위해 웨이블릿 영역의 저주파 영역에서 LBX 인터리빙을 이용해 음성워터마크를 영상에 삽입하였다. 제안한 방법은 외부 공격으로 인해 음성워터마크 데이터 비트열의 일부가 손실되더라도 제안한 디인터리빙과 후처리과정을 통하여 손실된 워터마크 정보를 복구할 수 있었다. 시뮬레이션 결과 표준영상압축 방식인 손실 JPEG quality에서는 전반적으로 음질이 좋은 음성신호를 복원시킬 수 있었고, 영상의 일부가 제거되는 영상절단공격에 대해서는 25%의 영상만으로도 고음질의 음성신호를 모두 복원시킬수 있었다. 그리고 영상의 일부와 화소값이 동시에 변형되는 영상회전 공격에 대해서도 음성신호를 최적으로 복원할 수 있었다.

따라서 본 논문에서 제안한 알고리즘은 공격에 대한 높은 강인성을 유지, 정보량이 많은 콘텐츠 소유자의 음성신호 이용 등에서 다양한 멀티미디어 저작권 보호에 활용될 수 있으리라 사료된다.

### 참고 문헌

1. W. Bender, D. Gruhl, and N. Morimoto, "Techniques for Data Hiding," Proc. SPIE, Storage and Retrieval for Image and Video Database III, 2420, 164-173, Feb. 1995.
2. R. G. van Schyndel, A. Z. Tirkel and C. F. Osborne, "A Digital Watermark," Proc. IEEE Int. Conf. Image Processing,

2. 86-90, Nov. 1994.
3. I. J. Cox, J. Kilian, F. T. Leighton, and T. Shammon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Trans. Image Processing, 6, 1673-1686, Dec. 1997.
4. A. Piva, M. Barni, F. Bartolini, and B. Cappellini, "DCT-based Watermark Recovering Without Resorting to the Uncorrupted Original Image," Proc. IEEE Int. Conf. Image Processing, 520-523, Oct. 1997.
5. M. Barni, F. Bartolini, V. Cappellini, A. Lippi, And A. Piva, "A DWT-based Technique for Spatio-frequency Masking of Digital Signatures," Proc. SPIE/IS&T Int. Conf. Security Watermarking Multimedia Contents, 3657, Jan. 1999.
6. D. Kundur and D. Hatzinakos, "Digital Watermarking using Multiresolution Wavelet Decomposition," Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 5, 2969-2972, May 1998.
7. 고성식, 조강석, "음성로그 삽입을 위한 디지털 영상 워터마킹에 관한 연구," 대한전자공학회 논문지, 39 (3), 21-27, Sep. 2002.
8. ITU-T, "Method for subjective determination of transmission quality", Rec. P.800, Aug 1996.

### 저자 약력

#### • 고 성 식 (Sung-Shik Koh)



1994년 2월 조선대학교 전자공학과 학사 졸업  
 1996년 2월 조선대학교 전자공학과 석사 졸업  
 2002년 2월 조선대학교 전자공학과 박사 졸업  
 2003년 3월~2004년 8월 조선대학교 전자공학과  
 겸임교수  
 2004년 9월~현재 일본 오사카시립대학 공학연구과  
 Post-doc과정  
 주관심분야: 신호처리, 컴퓨터비전, 영상워터마킹, 3차원  
 복원

#### • 최 제 승 (Jae-Seung Choi)



1989년 조선대학교 전자공학과 학사 졸업  
 1995년 일본 오사카시립대학 정보통신공학과 석사 졸업  
 1999년 일본 오사카시립대학 정보통신공학과 박사 졸업  
 2000년~2001년 일본 미쯔시바 전기산업주식회사  
 AVC사 연구원  
 2002년~현재 경북대학교 디지털기술연구소 연구원,  
 프로젝트 리더  
 주관심분야: 영상 및 음성신호처리, 잡음제거, 신경망,  
 디지털 TV 등

#### • 김 정 화 (Chung-Hwa Kim)



1979년 2월 조선대학교 전자공학과 학사 졸업  
 1981년 2월 조선대학교 전자공학과 석사 졸업  
 1991년 2월 숭실대학교 전기공학과 박사 졸업  
 1979년~현재 조선대학교 전자공학과 교수,  
 2004년 2월~현재 조선대학교 정보통신원장  
 주관심분야: 유비쿼터스, 신호처리 및 시스템, 영상처리,  
 영상 워터마킹