論文2002-39TE-3-4

음성로고 삽입을 위한 디지털 영상 워터마킹에 관한 연구

(A Study on Digital Image Watermarking for Embedding Audio Logo)

趙 康 石 *, 高 成 植 **

(Gang-Seok Cho and Sung-Shik Koh)

요 약

디지털 워터마킹 기술은 멀티미디어 데이터에 적용하여 소유물에 대한 불법 복제나 소유자의 저작권 보호 문제를 해결할 수 있는 방법 중의 하나이다. 그렇지만 디지털 영상, 비디오, 그리고 오디오와 같은 멀티미디어 보호 문제를 해결하기 위해서는 여전히 소유권 주장 방법에 대해서 극복해야할 문제점을 가지고 있다. 본 논문에서는 저작물 소유자의 음성신호(Audio signal)를 음성로고(Audio Logo)의 워터마크로 변환하고, 원 영상 내 픽셀 농도 값의 비선형적 특성를 이용하여 음성로고를 공간 영역에서 삽입한다. 그리고 추출된 음성로고를 음성신호로 변환하여 청각적으로 소유자의 음성신호를 스피커로 출력하여 소유권을 주장하는 방법을 제안한다. 실험 결과를 통해, 본 논문에서 제시하는 알고리즘이 일반적인 영상 처리나, 특히 손실 JPEG 등과 같은 다양한 공격에 대해서 강인성을 유지시킬 수 있어 음성로고를 이용해 저작물 보호를 확인할 수 있음을 검증하였다.

Abstract

The digital watermarking methods have been proposed as a solution for solving the illegal copying and proof of ownership problems in the context of multimedia data. But it is still difficult to have been overcame the problem of the protection of property to multimedia data, such as digital images, digital video, and digital audio. This paper describes a watermarking algorithm that embeds non-linearly audio logo watermark data which is converted from audio signal of the ownership in the components of pixel intensities in an original image and that insists of ownership by hearing the audio signal transformed from the extracted audio logo through the speaker. Experimental results show that our algorithm using audio logo proposed in this paper is robust against attacks such as particularly lossy JPEG image compression.

* 平生會員, 朝鮮理工大學 光電子情報學部
(Dev. of Photonics · Electronics & Information, Chosun College of Science & Technology)
** 正會員, 朝鮮理工大學 光電子情報學部
(Dev. of Photonics · Electronics & Information, Chosun College of Science & Technology)
※ 이 논문은 2001년도 조선이공대학 학술연구비 지원을 받아 연구되었음
接受日字:2002年5月3日, 수정완료일:2002年7月26日

Ⅰ. 서 론

미디어에 대한 디지털화 추세는 전송과 편집 그리고 저장을 위한 요구로 계속 발전해 가고 있다. 영상과 음성 그리고 문서 등이 디지털화되면서부터 누구나 쉽게 이용하고 복사할 수 있게 되었다. 그래서 전자 출판이나 디지털화된 미디어에 있어서 소유권 침해에 대한 저작권 보호 문제가 중요하게 여겨지고 있다. 이 때문

에, 영상 처리 분야에서 디지털 영상, 오디오, 문서 등의 워터마킹 알고리즘이 새롭게 등장하고 급속히 발전해 가고 있다. 그 중 디지털 영상 워터마크는 영상 소유자를 확인하기 위해 사용되고 영상의 불법 복제 및 배포를 차단시키는데 사용된다. 저작권 보호를 위해, 워터마크는 은밀하게 삽입되어야 하고 원 영상과 워터마크가 삽입된 영상 사이에서 시각적 차이가 없어야 한다. 또한, 강인한 워터마킹 기법은 영상처리와 손실 영상 압축과 같은 여러 가지 영상처리의 공격에도 손상되지 않아야 한다.

기존 디지털 영상 워터마킹 알고리즘은 주파수 영역 기술[1-4]과 공간 영역 기술[5-8]로 연구되고 있다. 주파수 영역 기술은 어떤 주파수 영역의 계수 값을 워터마크 정보에 따라 변형하는 반면에, 공간 영역 기술은 영상 내 선택된 픽셀의 농도 값 또는 칼라 값을 워터마크 정보에 따라 직접 변형한다. 주파수 영역에서 워터마킹 기술들은 DCT[1-2], 푸리에 변환[3], 그리고 웨이브렛 변환[4] 등이 있다. 공간 영역에서 워터마킹 기술들은 가장 간단히 최소 유효 비트(LSBs)에 워터마크를 삽입하는 방법, 비밀키를 의사 랜덤 발생기[6]로 사용하는 방법, 적응적으로 워터마크를 삽입하는 방법[7], 그리고 프랙탈 이론을 바탕으로 한 스테가노그래피 방법[8] 등이 있다.

이러한 기존 워터마킹 방법들은 삽입할 워터마크 데이터를 영상 정보가 적은 이진 영상이나 간단한 시리얼 번호를 이용하였고, 특히 공간 영역에서 워터마크 데이터가 삽입된 영상은 다양한 영상 처리와 같은 공격을 받으면 쉽게 손상되었다.

따라서, 본 논문에서는 워터마크 데이터로 멀티미디어 정보 소유자의 일차원 음성신호 데이터를 이용하여 이차원 음성로고로 변환한 후 비선형적 워터마킹 기술[9]을 적용하여 삽입한다. 그리고 추출된 음성로고를 음성신호로 변환하여 분쟁 시 청각적으로 소유권을 주장할 수 있는 방법을 제안한다. 또한 공간 영역에서 손실 JPEG 영상 압축과 같은 일반적인 영상 처리에 대하여 강인성을 유지시킬 수 있음을 증명 할 것이다.

Ⅱ. 영상에 음성로고 삽입

이 장에서는 영상의 소유권을 주장하기 위해 소유자의 음성신호를 음성로고로 변환하여 공간영역에서 음성로고 워터마크를 삽입하는 기술을 제안한다.

1. 음성로고로 변환하는 과정

저작권 보호를 위해 본 논문에서 이용된 샘플 음성신호는 한글발음 "조-강-석"을 발음한 일차원 신호이고, 그 Audio 신호 파형은 그림 1과 같다.

디지털 영상에 음성신호의 워터마크를 이용하기 위해서 일차원 음성신호를 이차원 음성로고로 그림 2와 같이 새롭게 제안된 방법으로 변환한다.

변환 방법은 샘플 음성신호의 아날로그 데이터를 디지털 신호로 변환하는 부호화 방식을 이용하여 이진 음성로고로 변환한다. 변환 과정은 다음 단계를 거친다.

Step 1. 음성신호를 6kHz 표본화 주파수로 2048개의 샘플 수를 이용한다.

Step 2. 256(8bit)/sample 레벨로 양자화 한다.

Step 3. 양자화된 신호를 Raster Scan방식으로 배열한다.

Step 4. Raster Scan 벡터 신호를 랜덤하게 스크램블링(Scrambling)한다.

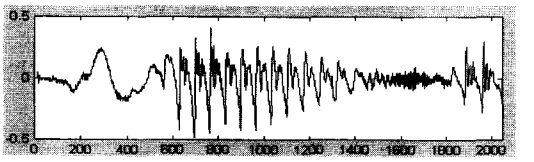


그림 1. 일차원 음성신호 "조-강-석" 파형
Fig. 1. 1D Audio Signal Waves "Cho Gang Seok".

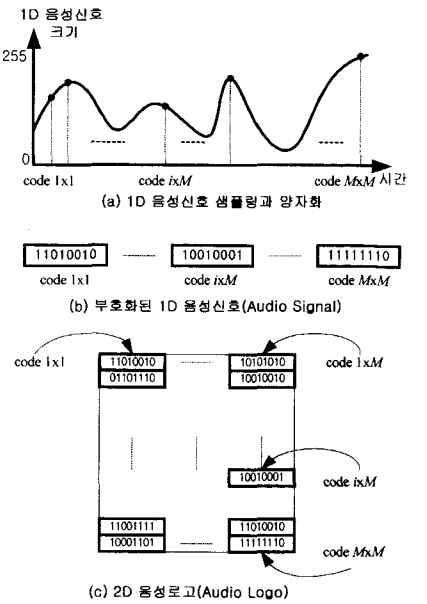
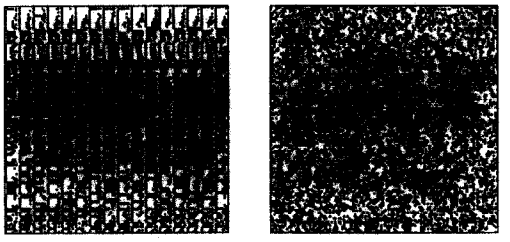


그림 2. 이차원 음성로고 구성 절차
Fig. 2. Sequence for consisting of 2D audio logo.



(a) 음성로고 (b) 스크램블링된 음성로고

그림 3. "조-강-석" 신호의 이차원 음성로고
Fig. 3. 2D Audio Logo of "Cho Gang Seok".

Step 5. 랜덤 절차로 구성된 일차원 벡터 신호를 새로운 크기 128×128(M×M)의 이차원 음성로고 배열로 변환한다.

상기 절차에 의해 음성신호는 이산 신호 0과 1의 조합인 이차원 이진 음성로고로 변환된다. 구성된 이진 음성로고(크기,128×128)와 스크램블링된 음성로고(크기,128×128)는 그림 3과 같다.

2. 비선형 특성을 이용한 음성로고 삽입

일반적인 공격에 대해 강인성을 유지시킬 수 있는 음성로고를 삽입하기 위해, 음성로고는 원 영상의 보다 중요한 영역에 삽입되어야 한다. 이러한 요구에 만족하기 위해, 본 논문에서는 Koh[9]가 제안한 방법의 원리인, 영상을 구성하는 블록들의 영상 정보는 복잡할수록 블록 내 픽셀 농도 값들의 변화가 비선형적 특성을 많이 가진다는 원리를 이용해 공간 영역에 음성로고를 눈에 보이지 않으면서 중요한 위치에 삽입시킨다. 또한, 워터마크로 이용된 음성신호 정보의 비밀성을 유지시키기 위해 음성로고를 랜덤하게 스크램블링 절차를 수행한다. 제안된 음성로고의 워터마킹 기술의 블록 다이어그램은 그림 4와 같다.

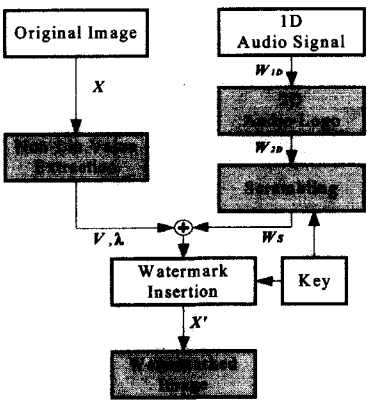


그림 4. 음성로고 삽입 시스템의 블록 다이어그램
Fig. 4. Block Diagram for Embedding Audio Logo.

원 영상 X와 음성로고 W_{2D}는 다음 식과 같이 나타낸다.

$$X = \{x(i,j),\ 0 \le i,j \le N\} \quad (1)$$

$$W_{2D} = \{w(i,j),\ 0 \le i,j \le M\} \quad (2)$$

여기서 N은 원영상의 크기이고 M은 워터마크의 크기이다. $x(i,j) \in \{0, 1, 2, \ldots, 2^l-1\}$는 원영상 픽셀들의 농도값이고 $w(i,j) \in \{0, 1\}$는 워터마크 픽셀들의 농도값이다. l은 각각의 픽셀에 할당된 비트수이다. 삽입 전에 음성로고의 픽셀은 시스템의 비밀키와 유사한 seed값을 사용하여 랜덤하게 다음 식(3)과 같이 스크램블링한다.

$$W_s = \{\text{scramble}(w(i,j),k),\ 0 \le i,j \le M\} \quad (3)$$

원 영상은 M×M 블록으로 분해되고, 그 블록 내의 픽셀 농도 특성의 차이에 따라서 비선형적으로 변형시켜 음성로고를 삽입시킨다.

$B_{M,M}$을 X에서 선택된 한 블록이라고 하자. $p_{\min}$과 $p_{\max}$, 그리고 p_{ave}는 블록 내 픽셀 농도 최대값과 농도 최소값, 그리고 농도 평균값이라고 하자. 그리고 음성로고를 픽셀 농도 값에 비선형적으로 삽입하기 위해 사용하는 파라미터 λ_{ij}와 V_{ij}는 다음 식과 같이 정의한다.

$$\lambda_{ij} = \{\max(D_{th}, (p_{\max} - p_{\min})),\ 0 \le i,j \le N/M\} \quad (4)$$

$$V_{ij} = \{\text{abs}(p_{ij} - p_{ave}),\ 0 \le i,j \le N/M\} \quad (5)$$

여기에서 p_{ij}는 $B_{M,M}$블록에서의 농도를 의미하고, D_{th}는 비선형 특성을 결정하는 역치 값이다. 스크램블링된 음성로고 W_s는 이진 영상이고 그 픽셀값 b_{ij_s}은 이진 데이터이다.

삽입하는 과정은 다음 법칙에 따라서 $B_{M,M}$ 블록의 픽셀 값을 비선형적으로 변형한다.

When $b_{ij_s} = 1$(음성로고의 화소값)

$p'_{ij} = p_{\max}$, if $p > p_H$

$p'_{ij} = p_{ave}$, if $p_L < p \le p_{ave}$

$p'_{ij} = p_{ij} + V_{ij}$, otherwise and if $p_{\max} - p_{\min} > D_{th}$

$p'_{ij} = p_{ij} + \lambda_{ij}$, otherwise and if $p_{max} - p_{min} \le D_{th}$

When $b_{ij} = 0$(음성로고의 화소값)

$p'_{ij} = p_{min}$, if $p < p_L$

$p'_{ij} = p_{ave}$, if $p_{ave} < p \le p_H$

$p'_{ij} = p_{ij} - V_{ij}$, otherwise and if $p_{max} - p_{min} > D_{th}$

$p'_{ij} = p_{ij} - \lambda_{ij}$, otherwise and if $p_{max} - p_{min} \le D_{th}$

여기서 p'_{ij}은 비선형적으로 변형된 픽셀 농도 값이다. 픽셀 농도 변화의 비선형 특성이 크면 픽셀 농도는 보다 많이 변형된다. 반면에 비선형 특성이 작으면 픽셀 농도 값은 조금 변형된다. 그러므로 블록 $B'_{M,M}$ 는 음성로고로 워터마크가 삽입된 영상 X'의 한 블록이 된다. 따라서 제안된 알고리즘은 한 블록 내 픽셀들을 음성로고에 따라 비선형적으로 변형할 수 있다.

Ⅲ. 음성로고 추출

제안한 방법으로 삽입된 음성로고의 추출을 위해서는 원 영상이 필요하다. 추출 순서는 삽입을 위한 음성로고 변환 과정(Step1~Step5)의 역순서와 같다. 음성로고 추출에 대한 전체적인 블록 다이어그램은 그림 5에서 설명된다.

추출된 음성로고 비트 값 b'_W는 원 영상과 음성로고가 삽입된 영상 사이의 블록 $B_{M,M}$과 $B'_{M,M}$의 차이로부터 얻을 수 있다. 만약 삽입된 음성로고 픽셀 값 b_W가 1 이었다면, $B'_{M,M}$의 픽셀 농도 합은 $B_{M,M}$ 보다 더 클 것이다. 반면에, 삽입된 음성로고 픽셀 값 b_W 이 0이었다면, $B'_{M,M}$의 전체 픽셀 농도 합은 $B'_{M,M}$의 전체 픽셀 농도 합보다 더 작을 것이다. 즉 음성로고 비트 값 b'_W은 다음 규칙으로 추출할 수 있다.

$b'_W = 1$, if $(B_{M,M} - B'_{M,M}) \le 0$

$b'_W = 0$, if $(B_{M,M} - B'_{M,M}) > 0$

얻어진 b'_W을 역으로 스크램블링(Step5, 4)하면 이진 음성로고를 얻을 수 있고, 음성로고를 일차원 음성신호를 얻기 위해서는 삽입 변환 과정을 역으로 실행(Step3, 2, 1)하면 된다.

따라서 멀티미디어 저작물의 소유자는 재구성된 일차원 음성신호로 저작권을 주장할 수 있다.

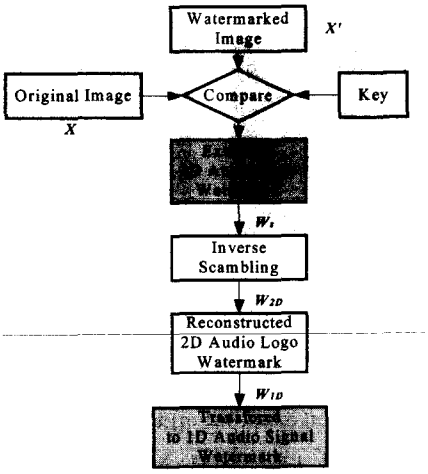


그림 5. 음성로고 추출 시스템의 블록 다이어그램
Fig. 5. Block Diagram for Extracting Audio Logo.

Ⅳ. 실험 결과

이 장에서, 실험 결과는 제안한 방법으로 삽입된 음성로고 워터마크의 공격에 대한 강인성을 확인한다. 그림 6은 크기 512x512의 원 영상이다.

음성로고의 다양한 공격에 대한 강인성을 확인하기 위해 다음 식(6) *DAR*(Detecting Accuracy Rate)[10]을 이용한다.

$$DAR = \frac{Number\ of\ bits\ correctly\ Dectected}{Number\ of\ bits\ Embedded} \times 100[\%] \quad (6)$$

그림 7은 제안한 방법에 의해 구성된 이차원 음성로고(그림3(b))를 원 영상에 삽입한 *PSNR* 30.32dB 영상이다.

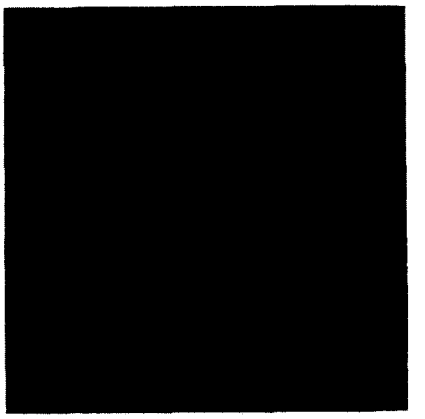


그림 6. 원 영상 512x512 Boats
Fig. 6. 512x512 Boats Image.

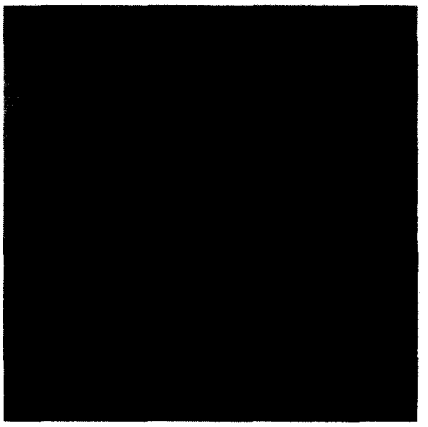


그림 7. 음성로고가 삽입된 영상(*PSNR* 30.32dB)
Fig. 7. Image embedded with Audio Logo.

표 1. 다양한 공격 후 추출된 음성신호
Table 1. Extracted Audio Signal after Attacks.

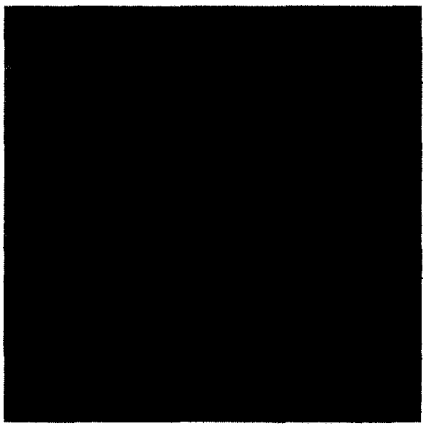
	Extracted Audio Signals							
	Boats (32.32dB)		Barbara (28.21dB)		Lenna (31.29dB)		Baboon (25.19dB)	
	PSNR (dB)	DAR (%)	PSNR (dB)	DAR (%)	PSNR (dB)	DAR (%)	PSNR (dB)	DAR (%)
Median	14.93	97.67	15.79	97.41	16.82	97.88	11.57	94.12
Rescaling	inf	100	inf	100	inf	100	inf	100
JPEG90	inf	100	inf	100	inf	100	inf	100
JPEG80	inf	100	inf	100	39.05	99.99	inf	100
JPEG70	45.05	99.99	60.12	99.99	32.67	99.93	57.11	99.99
JPEG60	31.50	99.90	29.70	99.88	24.77	99.74	37.95	99.96
JPEG50	24.08	99.68	26.00	99.74	22.63	99.68	29.73	99.89
JPEG40	20.12	99.19	23.13	99.24	17.88	99.41	22.47	99.58

표 1의 데이터는 제안한 방법으로 삽입된 음성로고가 기본적인 영상 처리와 손실 JPEG 영상 압축에 대해 강인성을 유지시킬 수 있음을 나타내고 있다. 표 1로부터, 우리는 추출된 음성로고가 Median 영상 처리에 기인하거나 손실 JPEG 영상 압축으로 인한 화질 손실이 발생하더라도 원 영상 소유자의 음성신호를 추출하여 저작권 주장을 위한 소유자의 신원을 확인할 수 있음을 알 수가 있다.

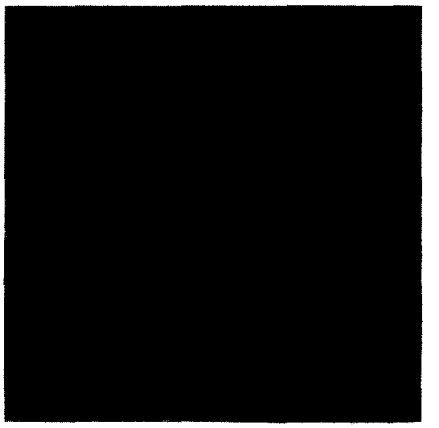
그림 8(a)는 Median Filter(3x3) 처리를 한 *PSNR* 29.41dB 영상을 나타내고, 그림 8(b)는 JPEG 60% 손실 압축된 *PSNR* 29.93dB 영상이다.

그림 9는 다양한 공격에 대한 워터마크의 *DAR*의 그래프를 나타내고 있다.

그림 10은 다양한 영상 처리 후 추출된 소유자의 음성신호를 나타낸다.



(a) Median 필터링된 영상(*PSNR* 29.41dB)



(b) JPEG Quality 60% 영상(*PSNR* 29.93dB)

그림 8. 공격에 대한 영상 화질 손실
Fig. 8. Lossy of image quality by attacks.

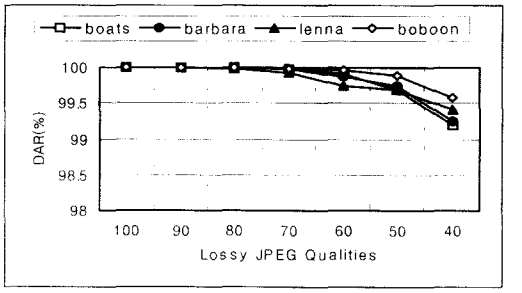
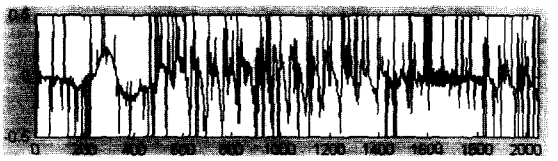
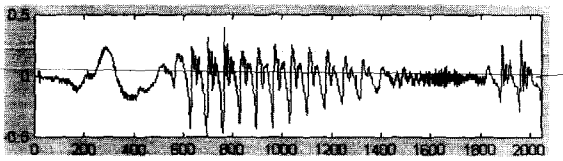


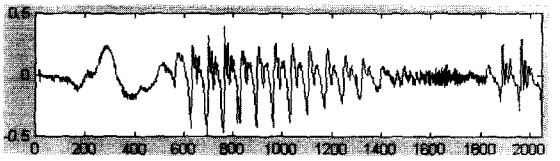
그림 9. JPEG 공격으로부터 추출된 음성신호
Fig. 9. *DAR* of Audio Signals extracted from lossy JPEG attacks.



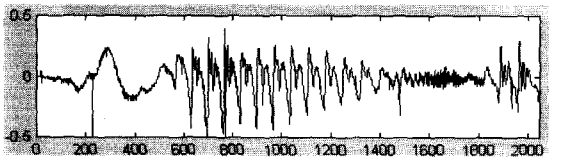
(a) Median filter 처리 후 추출된 음성신호
(*PSNR* 14.93dB, *DAR* 97.67%)



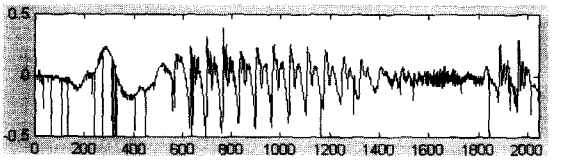
(b) JPEG 80% 처리 후 추출된 음성신호
(*PSNR* inf dB, *DAR* 100%)



(c) JPEG 70% 처리 후 추출된 음성신호
(*PSNR* 45.05dB, *DAR* 99.99%)



(d) JPEG 60% 처리 후 추출된 음성신호
(*PSNR* 31.50dB, *DAR* 99.90%)



(e) JPEG 50% 처리 후 추출된 음성신호
(*PSNR* 24.08dB, *DAR* 99.68%)

그림 10. 영상 처리된 영상으로부터 추출된 음성신호
(원 영상 : Boats image)

Fig. 10. Audio Signal Extracted Against Image Processing of the watermarked Image.

V. 결 론

본 논문에서는 공간 영역에서 픽셀 정보의 비선형 특성을 이용하여 음성로고 워터마크를 삽입하는 디지털 영상 워터마킹 방법을 제안하였다. 제안된 방법은 일차원 음성신호를 이차원 음성로고 워터마크로 변환한 후, 비선형적으로 분포되어 있는 영상 정보의 픽셀 값을 이용하여 음성로고를 삽입하였다. 음성신호를 영상 신호 차원인 2차원 음성로고로 변환하는 과정을 거치므로 과정이 다소 복잡하지만 저작물 소유자의 음성으로 직접 소유권을 확인할 수 있는 장점을 가진다.

실험적 결과를 통해 본 논문에서 제안한 워터마킹 방법은 특히 손실 JPEG 압축과 같은 공격에 대하여 강인성을 유지시킬 수 있어 소유자의 음성으로 저작권을 명확히 주장할 수 있음을 증명할 수 있었다.

참 고 문 헌

[1] Chiou-Ting Hsu, Ja-Ling Wu, "DCT-based watermarking for video", IEEE Transaction on Consumer Electronics, Vol. 44, No. 1, Feb 1998, pp. 206~216.

[2] Marc Schneider and Shih-Fu Chang, "A robust content based digital signature for image authentication", Proceedings of IEEE International Conference on Image Procesing, 1996, pp. 227~230.

[3] J. J. K. O'Ruanaih, W J. Dowling, and F M. Boland, "Phase watermarking of digital images", Proceedings of IEEE International Conference on Image Processing, Vol. 3, 1996, pp. 239~242.

[4] J. Ohnishi and K. Matsui, "Embedding a seal into a picture under orthogonal wavelet transform", in Proceeding of Multimedia, 1996, pp. 514~521.

[5] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osbone, "A Digital watermark", Proceedings of IEEE International Conference on Image Processing, Vol. 1, 1994, pp. 86~90.

[6] K. Matsui and K. Tanaka, "Video-Steganography : How to Embed a Signature in a Picture", in Proceedings of IMA Intellectual Property, Jan. 1994, Vol. 1, NO. 1, pp. 187~206.

[7] I. Pitas, "A method of signature casting on digital imaes", Proceedings of IEEE International Conference on Image Processing, Vol. 3.

1996, pp. 215~218.

[8] P. Davern and M. Scott, "Fractal based image steganography", in Proceedings of First International Workshop on Information Hiding, 1997, pp. 279~294.

[9] Sung Shik Koh and Chung Hwa Kim, "Watermarking using Non-linear Property in Spatial Domain" Proceedings of the ICIM'01, pp. 71~74, 2001.

[10] Changsheng Xu, Jiankang Wu & David Dagan Feng, "Contnet-Based Digital Watermarking for Compressed Audio", in RIAO'2000 Proceedings, April 2000.

저 자 소 개

趙 康 石(平生會員) 第38卷 TE編 第2號 參照

高 成 植(正會員) 第38卷 TE編 第2號 參照