

정지영상에서 저작권 보호 및 위변조 검출을 위한 하이브리드 디지털 워터마킹 기법[☆]

A Hybrid Digital Watermarking Technique for Copyright Protection and Tamper Detection on Still images

유길상*
Kil-Sang Yoo

송근실**
Geun-Sil Song

최혁***
Hyuk Choi

이원형****
Won-Hyung Lee

요약

오늘날, 디지털 이미지 도구 프로그램은 전문가가 아니더라도 누구나 개인용 컴퓨터를 이용하여 쉽고 빠르게 영상을 조작하여 사용할 수 있게 되었다. 그 결과 디지털 콘텐츠의 저작권 보호 및 변조된 영상의 무결성을 보증하는 것이 주요 문제가 되고 있다. 본 논문에서는 워터마크 정보는 물론 조작된 영상의 위치까지 검출할 수 있는 하이브리드 워터마킹 알고리즘을 설계하였다. 이를 위해 이산 웨이블릿 변환을 이용하여 영상의 저주파 대역에 PN-시퀀스를 워터마크로 삽입하였고, 원영상 없이도 검출이 가능하게 하였다. 워터마크 신호를 파괴하기 위한 다양한 공격의 실험 결과 제안한 알고리즘은 강인성을 나타내었고 추출 후 변조된 영상의 위치도 확인할 수 있었다.

Abstract

Digital image manipulation software is now readily available on personal computers. It is therefore very simple to tamper with any image and make it available to others. Therefore, copyright protection of digital contents and insurance of digital image integrity become major issues. In this paper, we propose a hybrid watermarking method to identify locations of tampered region as well as copyright. Our proposed algorithms embed the PN-sequence into low frequency sub-band of the wavelet transform domain and it doesn't need the original image in extraction procedure. The experimental results show good robustness against any signal processing with tamper detection on still image.

☞ Keyword : Digital watermarking, Tamper detection, copyright protection

1. 서론

인터넷의 급속한 확산으로 디지털 콘텐츠가 손쉽게 인터넷을 통하여 보급, 유포되고 있다. 그러나 디지털화된 데이터들은 쉽게 불법적으로 복제 및 조작할 수 있고, 원본과 복사본의 구분이

불가능하여 저작권 보호 문제가 심각하게 대두되고 있다 [1].

워터마킹은 보호하고자 하는 멀티미디어 콘텐츠에 자신의 소유권 정보에 해당하는 일련의 마크를 삽입함으로써 필요시 삽입된 정보를 추출하여 소유권을 확인할 수 있는 알고리즘을 말한다.

워터마킹의 여러 가지 기법으로는 원 영상의 LSB(Least Significant Bit)들을 변화시키는 공간영역(Spatial domain)기법 [2][3]과 JPEG 압축 알고리즘에 사용하는 DCT(Discrete Cosine Transform)기법, 그리고 DWT(Discrete Wavelet Transform)을 이용하는 주파수 영역(Frequency domain)기법 [4][5]이 있다.

* 준 회원 : 중앙대학교 영상공학과 박사과정
lucky@ms.cau.ac.kr(제1저자)

** 준 회원 : 중앙대학교 첨단영상대학원 박사
ylem74@netsgo.com(공동저자)

*** 비 회원 : 서울시립대학교 컴퓨터과학부 교수
chyuk@venus.uos.ac.kr(공동저자)

**** 종신회원 : 중앙대학교 첨단영상대학원 교수
whlee@cau.ac.kr(공동저자)

☆ 본 연구는 교육부 BK21 사업의 지원으로 수행되었음

본 논문에서는 워터마크의 강인한 검출을 위하여 정지영상을 웨이블릿 변환한 후 PN-sequence를 워터마크로 저주파 대역에 삽입한다. 이때 삽입 방식은 일반적인 스프레드 스펙트럼 방식에서처럼 변환 계수에 워터마크를 더하는 것이 아니고 이웃 계수값의 평균값과의 상대적인 위치로 워터마크를 삽입함으로써 변조된 위치를 확인할 수 있는 기능을 갖게 하였다. 또, 선형 귀환 쉬프트 레지스터(LFSR: Linear Feedback Shift Register)를 사용하여 발생된 랜덤 수열을 이용하여 삽입 위치를 정함으로써 함수에 사용된 Seed 키 값을 알고 있는 사용자만이 워터마크를 추출할 수 있도록 하여 보안성(security)을 강화하였다.

본 논문의 구성은 다음과 같다. 2장에서는 제안한 알고리즘의 구성과 삽입, 추출과정을 기술하고, 3장에서는 제안한 워터마킹 알고리즘의 위변조 확인 기능 및 압축 및 여러 가지 주파수 공격에 대한 강인성을 실험을 통하여 확인하였다. 4장에서는 실험결과를 통한 문제점과 향후과제에 대하여 기술하고 본 논문의 결론을 맺는다.

2. 제안 알고리즘

본 논문에서는 영상의 위변조 확인 기능과 저작권 확인 기능을 동시에 구현하기 위한 워터마킹 방법을 제안한다. 이를 위해 워터마크 삽입 과정에서 웨이블릿 변환을 이용하였고, 위변조 확인에 많이 사용되는 양자화 워터마킹과 일반적인 워터마킹 방법으로 많이 사용되는 대역확산 워터마킹을 함께 고려한 하이브리드 방식을 제안한다. 제안 방식을 살펴보면 다음과 같다.

2.1 워터마크 삽입 알고리즘

Wavelet변환은 수학, 전기, 의료, 통신 등 다양한 분야에서 독립적으로 발전되어 왔다. 특히, 이미지 압축에 있어서 wavelet 변환을 이용하면 높은 이미지 압축률과 고화질을 제공함이 알려져

있다. 현재 Wavelet은 JPEG2000에 사용되고 있는데 기존 JPEG보다 평균적으로 약 20%의 압축률 향상이 보고되었다. 웨이블릿은 단일 원형 함수에 기반을 둔 뛰어난 time-scale 국부성으로 데이터를 다중해상도 표현 및 분석이 가능하는데, 영상을 1단계 분할하면 네 개의 분할 영상을 얻을 수 있다. 각 부 영상들은 저주파 대역으로부터 고주파 대역으로의 계층적 정보를 가지고 있으며, 저주파 대역은 다음 단계로의 분할에 이용된다. 2단계 이산 웨이블릿 변환된 각 대역별(subband) 구조를 살펴보면 그림 1과 같다. 웨이블릿의 기저 함수는 다음 식과 같이 표현할 수 있다.

$$(W_{\psi}f)(a, b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt \quad (\text{식 1})$$

결국 DWT를 2차원적으로 적용할 때 image size가 $N \times N$ 이라면, 최종적으로 4개의 $N/2 \times N/2$ size의 분할 영상(LL, LH, HL, HH)이 분석 결과로 나온다. 여기서 LL은 저주파 성분, LH는 영상의 수평 고주파 성분을 나타내고, HL은 수직 성분, HH는 대각선 고주파 성분을 담고 있다. 제안 알고리즘에서 워터마크의 삽입은 웨이블릿 변환된 저주파 LL대역에서 이루어지며, 검출시 비밀성(Security)을 강화하기 위하여 선형 귀환 쉬프트 레지스터(LFSR: Linear Feedback Shift Register)를 이용하여 생성한 랜덤 수열을 이용하여 삽입 위치를 결정한다.

워터마크 삽입 방법은 아래 식과 같다. 식(2)와



(그림 1) 2-Level 웨이블릿 변환

식(3)에서, w_i 는 삽입하고자 하는 LL 저주파 대역의 원래 계수값이고, α 는 삽입강도(scale factor), x_i 는 bipolar형태 $x = \{-1, 1\}$ 를 가지는 PN-sequence 로써 워터마크로 사용되었다.

$$m_i = \left\lfloor w_i - \frac{w_{i-1} + w_{i+1}}{2} \right\rfloor \quad (\text{식 } 2)$$

$$w_i' = (w_{i-1} + w_{i+1})/2 \cdot (1 + \alpha \cdot x_i),$$

if $m_i < \text{Threshold}$ $w_i' = w_i$, Otherwise (식 3).

삽입 과정을 살펴보면, 먼저 저주파 대역의 선택된 위치에서 이웃 평균값을 구하고, 현재 위치의 계수 값과의 차가 임계값(threshold)을 초과하면 워터마크를 삽입하지 않고 임계값의 범위에 있으면 식(3)을 이용하여 워터마크를 삽입한다. 임계값은 워터마크된 영상의 비가시성을 만족하기 위하여 사용하는 것으로, 2단계 분해된 영상의 경우 Threshold = 40, 삽입강도 $\alpha = 0.02$ 를 사용하였다. 3단계 분해된 영상에서는 Threshold = 80으로 설정함으로써 비가시성을 만족할 수 있었다.

마지막으로 워터마크 삽입 알고리즘에 의하여 주파수 공간에 워터마크가 삽입된 영상을 다시 웨이블릿 역변환(IDWT: Inverse DWT)하게 되면 워터마크가 삽입된 영상을 최종적으로 얻게 된다. 원영상과 워터마크가 삽입된 영상은 인간 시각적으로 차이를 느낄 수 없음을 확인하였다.

2.2 워터마크 검출 알고리즘

검출 단계는 강인한 워터마크 검출을 위한 워터마크 추출과정과 변형된 위치 검출을 위한 추출과정 두부분으로 구성된다.

워터마크 검출 방식은 크게 원본을 필요로 하는 Non-blind 워터마킹과 원본을 필요로 하지 않는 블라인드 워터마킹 기법이 있다. 전자의 경우 워터마크 추출시 원본 영상을 제거하고 워터마크의 상관도를 구하므로 검출 성능이 우수한 반면, 검

출 시에 원본 영상을 사용할 수 있는 제한된 사람만이 검출이 가능하다는 문제점이 있다. 본 논문에서 제안된 워터마크 추출 알고리즘은 워터마크 검출을 위해서 원영상을 필요로 하지 않는다.

검출 과정은 다음과 같다. 먼저, 워터마킹된 영상으로부터 삽입된 워터마크를 추출하기 위하여 워터마킹된 영상을 이산 웨이블릿 변환을 수행한다. 다음으로 삽입 시에 사용한 seed 값을 이용하여 랜덤 수열을 발생시키고 이들 순서를 이용하여 워터마크가 삽입된 위치를 찾는다. 워터마크 수열을 추출하기 위해서는 아래 식을 이용한다. 즉, 현재의 계수값과 이웃 계수들의 평균값과의 차가 임계치를 초과하거나 0에 가까운 신호일 경우 워터마크가 없거나 영상이 변형되었다고 판단하며, 0.2를 초과하는 경우 1의 값을 -0.2보다 작은 경우 -1의 워터마크 값 X_i 로 결정한다.

$$S_i = w_i' - (w_{i-1} + w_{i+1})/2 \quad (\text{식 } 4)$$

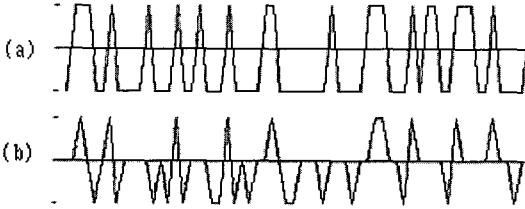
$$x_i = \text{signal}(s_i) = \begin{cases} 0, & \text{if } s_i \in \{0.2 \sim -0.2, |s_i| > \text{Threshold}\} \\ 1, & \text{if } s_i > 0.2 \\ -1, & \text{if } s_i < -0.2 \end{cases}$$

(식 5)

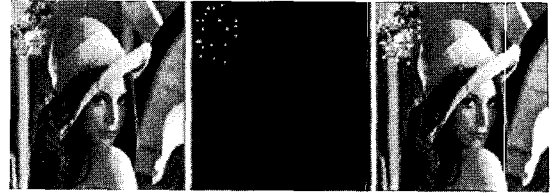
추출된 워터마크 수열은 삽입시에 사용한 레퍼런스 워터마크 수열 집합과의 상관도를 비교하여 특정 기준치를 넘는 피크가 뜨는 경우, 워터마크가 있다고 판단하며 이 경우의 PN-sequence값을 삽입된 워터마크 원순열로 판단하여 변조된 계수들을 판단한다. 그림 2에서 (a)와 (b)의 변조유무를 비교하고, 오류가 발생된 계수값은 해당하는 영상위치에 조작표시를 함으로써 조작위치를 알 수 있게 된다.

3. 실험결과 및 평가

본 논문에서 제안하는 워터마크 삽입 및 추출 방법은 MATLAB 6.1에서 구현되었다. 본 실험은



(그림 2) 두 Sequence 비교: (a) Original PN-sequence, (b) Extracted PN-sequence



(그림 3) 로고가 추가된 영상에서 조작된 위치 검출 결과

256*256*8bit gray 영상을 사용하였고, 이산 웨이블릿 변환으로 Haar-wavelet을 적용하여 영상을 주파수 분할하고, 워터마크를 삽입하였다. 제안된 방식에 의해 생성된 영상은 원영상과 시각적으로 구별되지 않았으며, 위변조 확인 기능 및 Lossy compression(손실 압축), Low/High filtering(저/고주파 필터링), Noise addition(잡음추가), Cropping, 복합공격 등의 다양한 영상처리와 변형에 대해서 워터마크를 추출할 수 있는지 실험하였다.

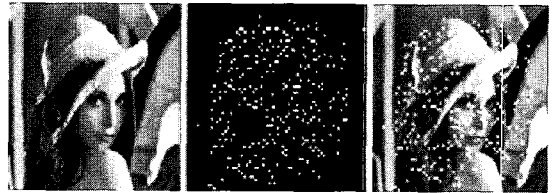
본 실험에 사용된 영상들은 디지털 워터마크 이미지로 널리 사용되고 있는 영상들으로써 [6], 대표적인 실험 결과값은 Lena 영상을 이용한 결과이다.

3.1 여러 영상에서의 비가시성 평가

워터마킹 알고리즘에 의해 삽입된 영상과 원영상과의 비가시성(Invisibility)을 평가하기 위한 척도로서 다음 식을 사용하여 PSNR(Peak Signal to Noise Ratio)을 구하였고, 원본 워터마크와 추출된 워터마크 사이의 상관도(Correlation)를 평가하기 위한 방법으로는 Normalized Correlation을 구하였다 [7]. 평가된 결과는 소수 셋째 자리에서 반올림한 값을 기록하였다.

$$PSNR = 20 * \log_{10} \frac{255 \cdot 255}{\sum (f(x) - f'(x))^2} \quad (\text{식 7})$$

$$NC = \sum_{x,y} \frac{w_{x,y} \cdot w'_{x,y}}{w_{x,y}^2} = \sum_{x,y} \frac{w \cdot w'}{w^2} \quad (\text{식 8})$$



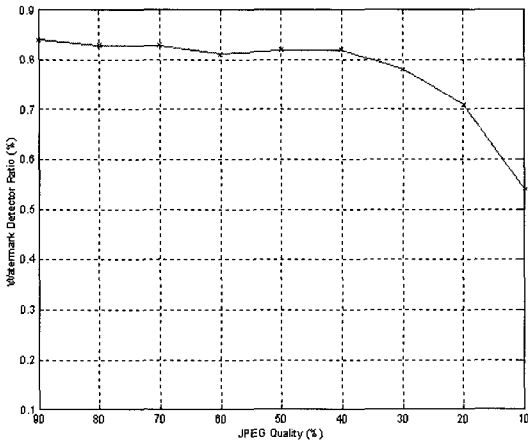
(그림 4) Pinch된 영상의 조작위치 검출 결과

원 영상과 워터마킹된 영상의 비가시성 측정결과 42.49[dB]으로 측정되었다. 워터마킹 비가시성을 만족하기 위하여 최소한 30[dB]이상이어야 한다 [8]. 서로 다른 주파수 분포를 가지는 영상에 대한 실험결과에서도 42[dB] ±1의 차이값의 범위 내에서 비슷한 결과값을 보였다.

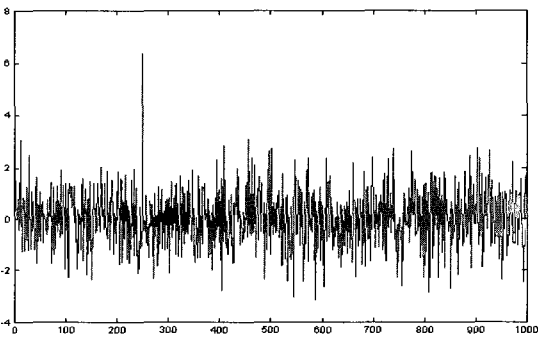
3.2 영상의 조작위치 검출

추출된 워터마크로부터 PN-sequence 랜덤집합과의 상관도를 비교하여 피크(pick)가 나오는 수열을 선택한 뒤, 두 수열에서 차이가 생기는 부분을 원영상에서 변형된 부분으로 생각할 수 있다. 즉, 그림 3에서 좌측영상은 워터마크가 삽입된 영상에 학교 로고를 추가한 영상인데, 위와 같은 방식으로 검출하면 가운데 그림과 같이 변조된 부분을 확인할 수 있었다. 오른쪽 영상은 변조된 영역을 영상에 함께 표시한 것이다.

그림 4는 Lena영상의 가운데 부분을 -20% Pinch를 가한 후 조작위치를 검출한 결과이다. 원본과 대조하기 전에는 조작여부를 눈치 챌 수 없으나 검출결과를 통하여 가운데 부분이 심하게 Pinch되어 있음을 알 수 있다. 그림 4의 워터마크 상관도 측정결과는 NC=0.58이었다.



(그림 5) JPEG 압축 실험



(그림 6) JPEG 6% Quality Compression

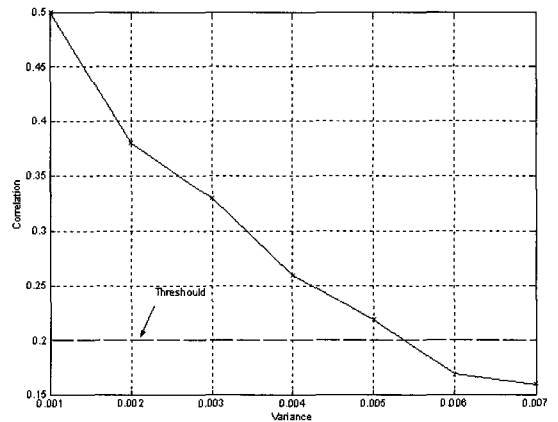
3.3 JPEG 압축 실험

그림 5는 Lena 실험영상들에 대하여 JPEG 압축 100 Quality에서 10% Quality씩 압축해 가며 압축에 대한 워터마크 추출 강인성을 실험한 결과이다. 실험에서 나타난 결과와 같이 10% Quality 압축하여도 워터마크가 완벽히 검출됨을 확인하였다. 다른 영상을 통하여 압축실험을 한 결과에도 비슷한 결과를 얻을 수 있었다. 따라서, 본 알고리즘은 특정 영상에 의존적이지 않고 영상에 따라 약간의 차이를 보이지만 전체적으로 대부분의 영상에 대하여 적합한 알고리즘임을 알 수 있다.

특히, 그림 6에서 보여 주듯이 영상을 6% Quality까지 압축 후에도 워터마크가 검출되어 JPEG 압축에 강건함을 알 수 있다.

(표 1) 저주파/고주파 필터링 공격 실험

Attacks		NC
Low pass	Blurring	0.71
	Blurring more	0.46
High pass	Sharpen	0.95
	Sharpen More	0.85
	Sharpen Edges	0.92



(그림 7) 가우시안 노이즈 추가 공격 실험

3.4 고주파/저주파 필터링 공격 실험

표 1은 워터마크된 영상에 저주파 필터를 통과시키는 Blurring 실험과 고주파 필터인 Sharpening 처리, 그리고 영상에 노이즈가 추가되었을 경우에 대한 워터마크 강인성 실험이다. 실험결과와 같이 저주파/고주파 필터 공격 및 영상잡음에 대하여도 강인성을 보였다.

3.5 노이즈 추가 실험

이 실험에서는 워터마크가 삽입된 영상에 노이즈를 첨가했을 때 검출율을 알아보기 위한 실험을 하였다. 잡음으로는 Gaussian white noise를 사용하여 평균값을 0으로 하여 분산(variance)값을 변화시켜 가며 실험하였다. 그림 7에서 보듯이 분산 = 0.006보다 큰 경우에는 워터마크 검출이 잘 되지 않음을 알 수 있다.

(표 2) StirMark Cropping 공격 실험결과

Attacks	NC
Cropping 2%	0.75
Cropping 10%	0.62
Cropping 15%	0.55
Cropping 20%	0.49
Cropping 25%	0.43
Cropping 50%	0.18
Cropping 75%	0.04

3.6 Cropping and Scaling

StirMark 3.1 [9]을 사용한 클로핑 실험결과, 영상 어느 부분을 제거하더라도 남은 25%만 있어도 완전한 워터마크를 추출할 수 있음을 알 수 있었다. (StirMark를 이용한 50% 클로핑은 영상의 가로세로 비율을 50% 잘라냄으로서 실제 남겨진 영상은 25%이다.) 영상을 StirMark 75% 클로핑하여 검출하였을 경우에는 워터마크 검출이 되지 않았지만, 전반적으로 제안한 알고리즘이 클로핑 공격에 강인함을 알 수 있었다. 이러한 이유는 워터마크 비트를 전 영상에 걸쳐 랜덤하게 삽입하였기 때문에 일부만 남아 있어도 상관도 검출이 가능하였다.

StirMark를 사용한 Scaling 0.5 ~ 2.0 실험결과에서도, 영상의 줄임이나 확장 후에도 워터마크가 검출됨을 확인하였다. 스케일링 후 NC값은 0.84로 측정되었다.

3.7 복합공격 실험

표 3은 클로핑 공격과 더불어 JPEG압축공격, 혹은 클로핑 공격된 영상에 저주파/고주파 영상처리가 가해진 영상에서의 워터마킹 검출 실험 결과이다. StirMark 공격틀로 클로핑 50%된 영상과 블러링 공격, 클로핑 50%된 영상과 고주파필터 공격, 클로핑 20%된 영상과 노이즈가 추가된 복합공격, JPEG 압축영상에 대하여 각각 25%클로

(표 3) 복합공격 실험결과

Mixed image Processing	NC
Cropping 25% + 3x3 Median filtering	0.30
Cropping 25% + Gaussian filtering 3x3	0.34
25% Cropping + Sharpening 3x3	0.22
Cropping 25% + JPEG Quality 50	0.37
Cropping 50% + JPEG Quality 60	0.17

핑과 50% 클로핑을 했을 경우에도 워터마크가 검출됨을 확인하였다.

4. 결론 및 향후 연구과제

본 논문에서 사용된 워터마크는 추출시 강인성을 위하여 원영상을 이산 웨이블릿 주파수 변환한 뒤 저주파 대역의 주변픽셀과의 상관도를 계산하여 삽입되었다. 워터마크 추출은 원영상 없이도 추출할 수 있도록 하였고, 추출 후 워터마크와 레퍼런스 수열과의 비교를 통하여 영상의 조작된 부위까지 검출할 수 있다.

실험결과, 제안한 방식은 워터마크 신호를 파괴하기 위한 다양한 공격에서 강인성을 나타내었고, 조작된 영상의 블록 위치까지 검출할 수 있음을 확인하였다.

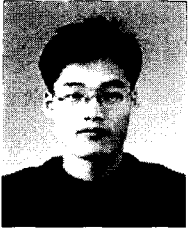
앞으로의 연구과제로는 제안한 워터마킹 기법을 기반으로 공간영역에서의 기하학적인 변형에 대한 강인성이 강화된 디지털 워터마킹 알고리즘의 연구를 계획하고 있다.

참고문헌

- [1] Chiou-Ting Hsu, Ja-Ling Wu, "Hidden digital watermarks in images", Image Processing, IEEE Transactions on, Volume: 8 Issue: 1 pp. 58~68, Jan. 1999.
- [2] Ioannis Pitas, "A method for watermark casting on digital image, Circuits and Systems for Video Technology", IEEE Transactions on, Volume: 8,

- Issue: 6, pp. 775~780, Oct. 1998.
- [3] Cox, I. J. Kilian, J. Leighton, F. T. Shamoon, T., "Secure spread spectrum watermarking for multimedia", *Image Processing, IEEE Transactions on*, Volume: 6, Issue: 12, pp. 1673~1687, Dec. 1997.
- [4] Podilchuk, C.I.; Wenjun Zeng, "Image-adaptive watermarking using visual models, Selected Areas in Communications", *IEEE Journal on*, Volume: 16, Issue: 4, pp. 525~539, May 1998.
- [5] Kundur, D.; Hatzinakos, D. "A robust digital image watermarking method using wavelet-based fusion, *Image Processing*", 1997. Proceedings., International Conference on, Volume: 1, pp. 544~547, 1997.
- [6] fabien a. p. peticolass Home, http://www.cl.cam.ac.uk/~fapp2/watermarking/image_database/index.html.
- [7] Stefan Katzenbeisser and Fabien A. P. Peticolas, "Information Hiding Techniques for Steganography and Digital Watermarking", ARTECH HOUSE INC, pp. 111~112, 2000.
- [8] Petitcolas, F.A.P, "Watermarking schemes evaluation", *IEEE Signal Processing Magazine*, Volume: 17 Issue: 5, pp. 58~64, Sep 2000.
- [9] http://ww.cl.cam.ca.uk/~fapp2/watermarking/image_watermarking/stirmark/.

● 저자 소개 ●



유길상

2003년 중앙대학교 첨단영상대학원 영상공학과 졸업(석사)
2003년~현재 : 중앙대학교 첨단영상대학원 영상공학과 박사과정 재학
관심분야 : 디지털 영상처리, 영상검색, 디지털 워터마킹
E-mail : lucky@ms.cau.ac.kr



송근실

1996년 호원대학교 전자계산학과 졸업(학사)
1998년 수원대학교 교육대학원 교육(석사)
2003년 중앙대학교 첨단영상대학원 영상공학과(박사)
관심분야 : Digital Watermarking
E-mail : ylem74@netsgo.com



최혁

1994년 서울대학교 전자공학과 졸업(학사)
1996년 서울대학교 대학원 전자공학과 졸업(석사)
2002년 서울대학교 대학원 전기컴퓨터공학부 졸업(박사)
2003년~현재 : 서울시립대학교 컴퓨터과학부 교수
관심분야 : Watermarking, 콘텐츠 보안
E-mail : chyuk@venus.ucs.ac.kr



이원형

현재 : 중앙대학교 첨단영상대학원 교수
한국컴퓨터게임학회 부회장
관심분야 : 영상처리/검색/인식, 게임
E-mail : whlee@cau.ac.ac.kr