비즈블 워터마크를 위한 블록 기반 내용 인증 방법

안세정, 정성환
창원대학교 컴퓨터공학과
e-mail:phillip@sarim.changwon.ac.kr

Block-wise Contents Authentication for Visible Watermark

Se-Jung Ahn, Sung-Hwan Jung
Mips Lab, Dept. of Computer Engineering, Changwon Nat'l University

요 약

본 논문은 시각적 워터마크의 삭제나 수정을 방지하기 위해 블록 기반 내용 인증(Block-wise content authentication) 방법을 제안한다. 제안한 방법은 원본 이미지와 워터마크 이미지를 사용하여 DCT 기반의 시각적 워터마크(Visible watermark)가 삽입된 이미지로 만든다. 다음, 블록 기반 서명(Block-wise signature)을 삽입하여 JPEG 과 같은 허용 가능한 이미지 조작에서는 정보가 유지되고, 워터마크의 삭제와 같은 악의적인 이미지 조작에는 인증을 허용하지 않는다. 또한 이미지의 조작된 부분을 알 수 있도록 로컬라이제이션(Localization)을 이용하여 조작된 부분의 위치를 파악할 수 있다. Lena를 포함한 여러 표준 영상을 사용하여 실험한 결과, 제안한 방법은 시각적 워터마크에 대한 다양한 공격에서 조작된 위치를 정확하게 표현할 수 있었다. 그리고 Photoshop 7.0의 Quality factor 11인 JPEG-90 압축에서 약 99%의 인증이 가능하였고, 블러링(Blurring)이나 샤프닝(Sharpening)과 같은 기타 공격에서는 각각 51%, 50%로 인증이 허용되지 않았다.

1. 서론

최근 다양한 미디어에 사용되고 있는 디지털 데이터를 보호하기 위해 디지털 컨텐츠 자체에 소유권 정보를 삽입하여 불법적인 복제나 원본 이미지 수정을 막는 디지털 워터마킹(Digital watermarking) 방법이 많이 연구되고 있다[1].

워터마킹 방법에는 강인성에 따라서 불법적인 조작에 견딜 수 있는 강인한(Robust) 워터마크 방법, 혹은 불법적인 조작이 가해진 이미지에 조작이 가해졌다는 흔적을 남기기 위한 약한(Fragile) 워터마크 방법이 있다. 그리고 가시성에 따라 워터마크가 삽입되어있다는 것을 시각적으로 알 수 있는 시각적(Visible) 워터마크 방법과 시각적 확인이 불가능한 비시각적(Invisible) 워터마크 방법이 있다[2].

그러나 최근에는 다양한 방법들을 결합하여 여러 가지 환경에 적합한 복합적 워터마킹 방법을 연구하고 있다[3-5]. 그 예로서, 불법적인 조작에는 강인하면서, 허용 가능한 조작에 대해서는 공격 시도를 나타내는 세미 프레자일(Semi-fragile) 워터마킹 방법[2]이나 또는 한번의 프린터 복사는 가능해도 추가적인 복사에는 약한 방법 등이 그 예이다.

시각적 워터마크를 위한 Mohanty의 방법[1]은 원본 이미지를 파괴하지 않으면서 시각적 워터마크를 삽입하기 위해 이미지의 평균값을 비교하여 이미지에 적응적으로 워터마크를 삽입하였다. 그러나 다양한 공격으로 시각적 워터마크를 삭제하거나 수정하면 인증(Authentication)여부를 확인할 수 없기 때문에 필수적으로 인증할 수 있는 방법이 필요하다.

본 논문은 시각적으로 워터마크가 삽입되어 있는 이미지에 대해 합법적인 조작이나 왜곡(Legitimate distortion)에는 영향을 받지 않지만, 불법적인 왜곡이나 조작에는 파괴된 부분을 시각적으로 보여주는

Localized tampering 과 Information-preserving, Lossy transformations 사이를 구별할 수 있는 Semi-fragile 워터마크 방법을 제안한다.

제안하는 방법에서는 인증을 위하여, 키에 의해 생성된 서명 비트(Signature bit)들을 각각의 DCT 기반의 블록에 삽입한다. 그리고 인증이 필요한 경우, 같은 키 값으로 생성된 서명을 워터마크 된 이미지에서 추출된 비트와 비교하여 인증한다[6].

다음 2장에서 시각적 워터마크를 삽입하는 방법을 보이고, 3장에서는 시각적 워터마크의 공격 시도된 로컬라이제이션(Localization tampering)을 표현하기 위한 방법을 설명한다. 그리고 4장에서는 실험 결과를 보이고, 5장에서 결론을 맺는다.

2. 시각적 워터마크 삽입

본 연구에서 워터마크를 원본 이미지의 왜곡을 방지하면서 시각적으로 삽입하기 위하여, 지역 이미지 특성들을 근거로 하는 삽입 파라미터(Embedding parameter)를 사용하여 이미지의 평탄한 부분이나 중요한 부분에 삽입한다[2]. 우선 원본 칼라이미지를 YCbCr 칼라 모델로 변형한 후, 8 × 8 DCT 블록으로 나눈다. 원본 이미지와 워터마크 이미지의 DCT 계수값을 다음의 수식 (1) 과 같이 계산한다.

$$c'_{ij(n)} = \alpha_n c_{ij}(n) + \beta_n w_{ij}(n) \quad n = 1, 2 \qquad (1)$$

여기에서 α_n은 블록 n의 스케일링 요소(Scaling factor)이고, β_n은 삽입 요소(Embedding factor), $c_{ij}(n)$은 원본 이미지 블록의 DCT 계수, $w_{ij}(n)$은 워터마크 이미지의 DCT 계수이다.

에지 부분은 AC 계수의 작은 편차를 가지고 있는 배경의 질감이 복잡한 블록에 비하여 많은 워터마크를 삽입할 수가 없다[1]. 그러므로 에지가 있는 블록의 α_n과 β_n은 각각 α_{max} 과 β_{min}되도록 선택하여 워터마크를 최소로 삽입한다. 반면 배경의 질감이 복잡한 블록에서는 다음의 수식(2)와 (3)을 사용한다.

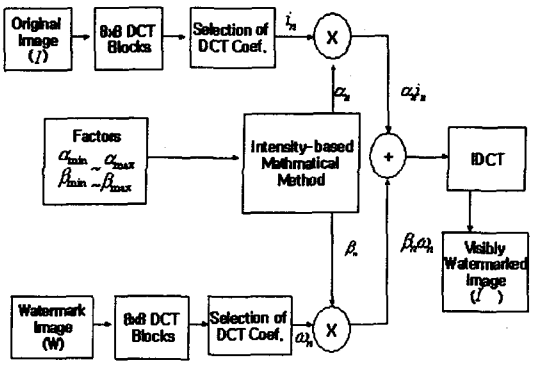
$$\alpha_n = \sigma'_n \exp.(-(\mu'_n - \mu)^2) \qquad (2)$$

$$\beta_n = (1/\sigma'_n)(1 - \exp.(-(\mu'_n - \mu)^2)) \qquad (3)$$

μ는 전체 이미지의 평균 그레이 값이고, μ_n은 원본 이미지 블록 n의 평균 그레이 값이고, σ'_n은 σ_n(DCT의 AC 계수의 편차(Variance)) 의 정규화된 로그값(Normalized logarithm)이다. 전체 이미지의 평균 그레이 값과 한 블록의 DC 계수를 비교하면 μ와 μ_n의 대소를 비교할 수 있기 때문에 대응하는 블록의 상대적 밝기(Intensity)를 알 수 있다[7].

중간-intensity 블록은 ($\mu_n \approx \mu$) 은 노이즈(Noise)에 더 민감하고, 에지가 많은 낮은-intensity 블록은 $\mu_n < \mu$이고, 에지가 적은 높은-intensity 블록은 $\mu_n > \mu$이다. 그러므로, α_n 은 $\mu_n < \mu$ 일 때 워터마크 삽입을 줄이기 위해 μ_n에 따라 증가하고, $\mu_n > \mu$일 때 워터마크 삽입을 늘이기 위해 μ_n에 따라 감소한다.

다음의 [그림 1]은 시각적 워터마크를 삽입하기 위한 과정을 블록도로 보이고 있다.



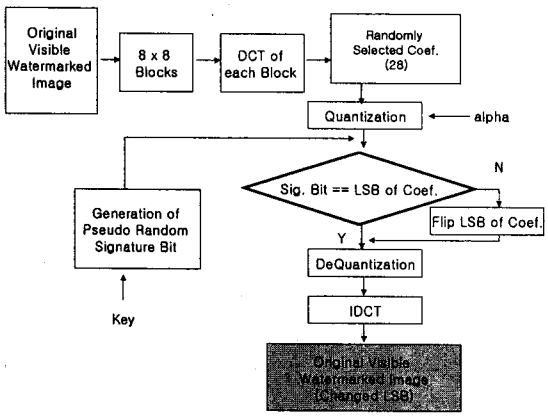
[그림 1] Visible watermark embedding method

3. 블록 기반 내용 인증

시각적 워터마크가 삽입된 이미지가 악의적인 공격으로 인해 훼손되었는지를 확인하기 위해 로컬라이제이션 방법이 사용된다[6]. 로컬라이제이션 방법에는 이미지를 블록으로 나누어 서명을 삽입하는 블록 기반 인증 방법(Block-wise authentication)과 각 픽셀에 대해 서명을 삽입하는 픽셀 기반 인증 방법(Sample-wise authentication) 두 가지가 있다. 본 논문에서는 시각적 워터마크를 삽입할 때 DCT 블록으로 나누기 때문에 서명 비트를 삽입하기 위해 역시 블록 기반 내용 인증 방법을 사용하였다.

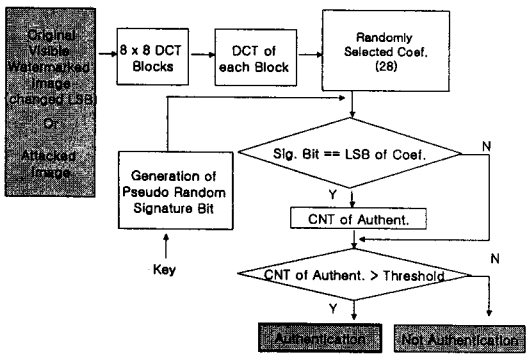
[그림 2]는 본 연구에서 시각적 워터마크가 삽입된 이미지에 대해 랜덤 서명 비트 1 또는 0을 삽입하는 방법의 블록도를 보이고 있다. 키 값을 사용하

여 랜덤 비트를 만들고, 랜덤 비트와 DCT 계수의 LSB 비트를 비교하여 다르면 뒤집기(Flip)를 하고, 같으면 뒤집지 않는 방법으로 서명 비트를 시각적 워터마크 이미지에 삽입한다.



[그림 2] Signature bit embedding for authentication

아래의 [그림 3]은 서명 비트를 삽입한 이미지나 수정, 삭제 공격을 받은 이미지를 입력받아 인증을 확인하기 위해 서명 비트를 추출하는 블락도를 보이고 있다.



[그림 3] Signature bit extracting for authentication

제안하는 방법은 인증을 위해 서명 비트를 추출할 때 원본 이미지가 존재하지 않아도 되며, 삽입할 때와 동일한 키로 랜덤 비트를 생성한다. 그런 후 랜덤 비트와 이미지의 DCT 계수의 LSB 비트를 추출, 비교하여 인증 할 수 있는 Threshold를 만족하는지 비교한다.

Semi-fragile watermark는 중간 압축(Moderate compression)일 때는 약 75% 정확성과 가벼운 압축(Light compression)일 때 약 90%의 정확성으로 수정된 영역(Altered regions)을 알 수 있다[2]. 그러므로 제안하는 방법에서는 JPEG 압축에 대한 인증 Rhreshold는 75%를 기준으로 하였고, 이미지의 수정이나 삭제에 대한 인증 Threshold는 85%를 기준으로 하였다.

4. 실험 결과 및 분석

본 논문의 실험은 256 × 256 크기의 Lena 및 Baboon 등과 같은 표준 영상을 사용하였고, 서명 삽입할 때 JPEG factor를 감안하여 양자화 알파값을 0.1로 하였다.

Photoshop 7.0의 JPEG factor에 따라 12Q, 11Q, 10Q 로 시각적 워터마크가 된 이미지를 저장 한 후, 서명 비트를 삽입하였다. 실험 결과는 다음의 <표 1> 과 같다.

<표 1> 압축 Quality factor와 image processing에 따른 authentication (단위 : %)

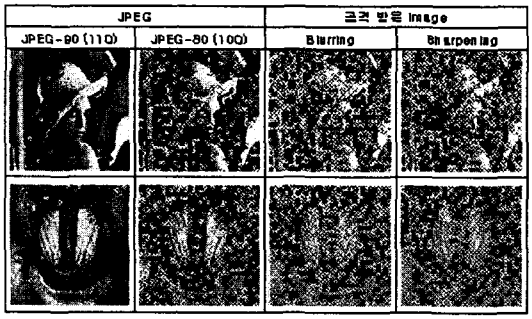
압축 / 영상	uncompression(12Q)	JPEG-90 (11Q)	JPEG-80 (10Q)	블러링	샤프닝
Lenna	99.5	99.5	79.2	51.5	50.6
Baboon	99.4	99.1	70.2	49.8	50.6
Lake	99.5	99.4	76.2	51.2	49.8
House	99.6	99.6	76.6	51.7	50.9
평균	99.5	99.4	75.5	51.1	50.5

실험 결과, JPEG 압축에서, JPEG-90 과 JPEG-80은 각각 약 99.4% 와 75.5% 로 인증 Threshold를 만족하지만, 블러링(Blurring) 이나 샤프닝(Sharpening) 과 같은 기타의 이미지 프로세싱은 인증을 허용하지 않았다. 그리고 조작된 이미지의 인증을 허용하지 않을 때, 수정되거나 삭제된 영역을 표현하여 시각적으로 보여줄 필요가 있다.

[그림 4]는 시각적 워터마크를 고의로 삭제했을 때, 삭제된 영역의 서명 비트가 일치하지 않아 조작된 블록으로 분류하여야 하는 블록을 시각적으로 보여주고 있다.

[그림 4] 시각적 워터마크의 삭제된 영역의 표현

[그림 5]는 JPEG-90 과 JPEG-80은 각각 99.4%와 75.5% 이상으로 허용 가능한 조작으로 인증이 가능하고, 블러링이나 샤프닝은 전체 이미지에 걸쳐 많은 손상으로 인증이 불가함을 눈으로 확인할 수 있었다.



[그림 5] JPEG Quality factor 와 이미지 프로세싱 공격에 따른 에러 블록의 표현

결과적으로 공격자가 시각적 워터마크의 존재를 확인하고 정교하게 워터마크를 삭제 또는 수정한다 해도 키에 의해 삽입된 서명을 변경할 수가 없기 때문에 공격당한 위치를 찾아 이미지의 인증 유무를 확인할 수 있다.

5. 결론

본 논문은 워터마크가 시각적으로 드러나는 이미지의 워터마크 삭제나 수정을 방지하기 위해 블록 기반 내용 인증(Block-wise content authentication) 방법을 제안하였다. 제안한 방법은 DCT 기반의 시각적 워터마크가 삽입된 이미지에 대해 블록 기반 서명을 삽입하여 JPEG 과 같은 허용 가능한 이미지 조작에서는 정보가 유지되고, 워터마크의 삭제와 같은 악의적인 이미지 조작에는 인증을 허용하지 않았다. 또한 이미지의 조작된 부분을 알 수 있도록 로컬라이제이션을 이용하여 조작된 부분의 위치를 파악할 수 있었다. Lena를 포함한 표준 영상을 사용하여 실험한 결과, 제안한 방법은 시각적 워터마크에 대한 다양한 공격에서 조작된 위치를 정확하게 표현할 수 있었다. 그리고 JPEG-90과 JPEG-80 압축에서 각각 약 99%, 75%의 인증이 가능하였고, 블러링이나 샤프닝과 같은 기타 공격에서는 각각 약 51%, 50%로 인증이 허용되지 않았다.

향후 연구로 다양한 컬러 모델에서도 효과적으로 워터마크를 삽입할 수 있는 방법과 이미지의 기하학적인 변형이나 왜곡에 적응적인 워터마킹 방법을 연구할 필요가 있다.

참고문헌

[1] Saraju P. Mohanty, K. R. Ramakrishnan, Mohan S. Kankanhalli, "A DCT Domain Visible Watermarking Technique for Images," IEEE International Conference on Multimedia and Expo 2, pp.1029-1032, 2000.

[2] E. T. Lin, C. I. Podilchuk, and E. J. Delp, "Detection of image alterations using semi-fragile watermarks," SPIE International Conf. on Security and Watermarking of Multimedia Contents II, 3971(14), January 2000.

[3] R. B. Wolfgang and E. J. Delp, "Fragile Watermarking Using the VW2D Watermark," Proceedings of the SPIE/IS&T International Conference on Security and Watermarking of Multimedia Contents, vol. 3657, San Jose, CA, pp. 204-213, January 25-27, 1999.

[4] P. Wong, "A watermark for image integrity and ownership verification," Final Program and Proceedings of the IS&T PICS 99, pp.374-379, Savanna, Georgia, April 1999.

[5] M. Yeung and F. Mintzer, "Invisible watermarking for image verification," Journal of Electronic Imaging, vol. 7, no. 3, pp.578-591, July 1998.

[6] C. -Y. Lin and S. -F Chang, "A Robust Image Authentication Algorithm Surviving JPEG Lossy Compression," Storage and Retrieval of Image/Video Databases, SPIE-3312, pp.296-307, 1998.

[7] 안세정, 정성환 "허용 가능한 조작에 대한 내용 적응 시그너쳐 생성 기법," 멀티미디어 학회, vol. 6, no.1, pp.255-258, May 2003.