디지털 워터마킹 알고리즘 평가 프로파일에 관한 연구

이진홍†, 임태훈‡, 박지환†
†부경대학교 정보보호학과
‡부경대학교 메카트로닉스공학과

A Study on the Digital Watermarking Algorithm Evaluation Profiles

Jin-Heung Lee†, Tea-Hun Lim‡, Ji-Hwan Park†
† Interdisciplinary Program of Information Security, PuKyong Nat'l University
‡ Interdisciplinary Program of Mechatronics, PuKyong Nat'l University

요 약

최근, 디지털 워터마킹에 대한 연구와 더불어 알고리즘의 안전성을 평가하는 평가기술에 대하여 그 중요도가 증가되고 있다. 본 논문은 기존에 개발되어진 벤치마킹 프로그램에 대하여 그 구조와 방법에 대하여 조사하고, 여러 응용 시스템에서 다른 목적으로 이용되는 워터마킹 기술들에 대한 안전성을 평가하는 평가 프로파일에 대하여 기술한다. 본 연구 결과로 작성된 프로파일은 워터마킹 알고리즘의 안전성 평가시, 유용한 자료로 사용될 수 있다.

1. 서론

디지털 워터마킹 기술은 컨텐츠의 원 소유자를 주장할 수 있는 워터마크 정보를 삽입하여 저작권에 대한 분쟁이 발생할 경우 삽입된 워터마크 정보를 추출하여 저작권을 주장하는 기술이다. 최근 인터넷의 발달과 정보통신 기술의 발달에 힘입어 디지털 컨텐츠 활용이 크게 증가하고 있다. 컨텐츠의 활용 증가와 함께, 응용 분야 또한 확대되고 있다. 기존의 음악, 영화, 사진 등의 저작권 증명은 물론, 문서의 불법유통 방지, 신분증 위·변조 방지, 방송국 시청률 조사 등의 모니터링 시스템, 디지털 방송, DVD Player Potable Device 등 다양한 분야로 확대되고 있다.

이러한 디지털 워터마킹 기술은 다양한 특성 및 요구조건들을 만족해야 한다[1]. 강인성(robustness)은 디지털 컨텐츠의 배포 중에 발생할 수 있는 다양한 신호처리적 공격과 불법적인 사용자의 고의적인 공격에 대하여 삽입한 워터마크가 그대로 남아있어 추출이 가능해야 한다. 그리고, 비가시성(invisibility)은 워터마크의 삽입여부를 감지하지 못하고 컨텐츠의 품질을 유지해야 하는 것으로 사용자는 컨텐츠의 워터마크 삽입여부를 알 수가 없게 된다. 그리고, 이외에도 추출된 워터마크 정보에 대하여 확실한 소유권 증명이 가능하도록 하는 신뢰성(reliability), 응용 시스템 구현 시, 기본적인 플랫폼에 대해서 실시간 구현이 가능하도록 하는 효율성(efficiency)등이 요구되고 있다.

저작권 보호를 위한 워터마킹 기술은 크게 공간영역과 주파수영역에서 변경하는 방법으로 나눌 수 있다. 공간영역 방법은 영상의 픽셀 값들을 직접 변화시켜 워터마크를 삽입하는 방법이다. Schyndel은 삽입되는 워터마크를 키에 의해 발생된 난수로 원 영상의 픽셀들을 임의적으로 선택하여 LSB(Least Significant Bit)를 변형하는 방법을 제시하였다[2]. 영상의 모든 픽셀 중 평균 50%의 화소값이 변경되어 워터마크가 삽입되므로 삽입 영역이 노출되면 공격에 취약한 상태에 놓이게 된다. 또한 워터마크는 매우 적은 파워로 삽입되므로 신호처리적인 공격, 잡음 그리고 압축 등에 대하여 워터마크가 제거되는 취약성을 가진다.

주파수영역을 이용한 워터마킹 방법은 영상을 DFT (Discrete Fourier Transform), DCT(Discrete Cosine Transform), DWT(Discrete Wavelet Transform) 등의 변환에 의해 주파수 성분의 계수로 변경하고, 그 값들을 변경하여 워터마크를 삽입하는 기술이다[3]. 주파수 계수의 고주파 성분은 비가시성이 좋은 영역이지만 압축 이용시 제일 많이 압축되어 손실되는 부분으로 공격에 대한 취약성을 가진다. 따라서, 대부분의 방법에서는 중간대역에 삽입하여 이러한 공격으로부터 강인성을 유지하고 있다. Hartrung은 대역확산 원리에 기반한 워터마킹 알고리즘을 제안하였다[4]. 삽입될 워터마크의 비트를 PN 계열에 의해 확산하고 그 신호를 다시 진폭계수에 의해 확대하여 원 데이터에 삽입하게 된다. 삽입된 워터마크의 검출은 상관도를 이용한다. Podilchuk 등은 영상을 DWT를 이용하여 다해상도 영상으로 변환 뒤, JND(Just Noticeable Difference)를 삽입강도로 이용해서 워터마크를 삽입하였다[5].

본 연구는 SEDICA 지원에 의하여 수행되었음

본 논문에서는 적용 방법과 응용분야가 다양한 워터마킹 알고리즘을 평가하기 위한 평가 프로파일 생성 과정과 평가 항목에 대하여 기술하고, 그 결과 생성된 워터마킹 평가 프로파일의 유용성을 보인다. 또한, 기존에 제시된 워터마킹 알고리즘을 평가 프로파일에 적용한 결과를 도출하여 일반적인 평가 툴과 비교, 분석하였다.

2. 벤치마크 프로그램

디지털 이미지 워터마킹 알고리즘의 강인성 평가는 다른 매체에 대한 알고리즘보다 활발히 연구가 진행되고 있으며, 몇가지 벤치마킹 프로그램들이 발표되어 있다. 삽입된 워터마크를 공격하는 것은 영상을 조작하여 삽입된 워터마크 신호를 제거하거나, 원 저작자조차 검출할 수 없게 만들거나 위조된 워터마크 신호를 재 삽입하여 소유권 주장을 할 수 없게 만드는 방법이다. 대표적인 벤치마킹 프로그램으로 StirMark, Checkmark 그리고 JEWELS 등이 있다.

2.1 삽입된 워터마크의 공격 방법

삽입된 워터마크를 공격하는 방법은 다양하게 있으며 대표적으로 Kutter, Voloshnovskiy 등의 분류가 있다. Kutter는 크게 JPEG 압축, 기하학적 변환, 신호처리 공격으로 분류하였다[6]. 기하학적 공격은 영상의 회전, 크기 변환, 잘라내기 등이 포함된다. 신호처리 공격에는 필터링, 히스토그램 변화, 잡음 첨가 등이 포함된다.

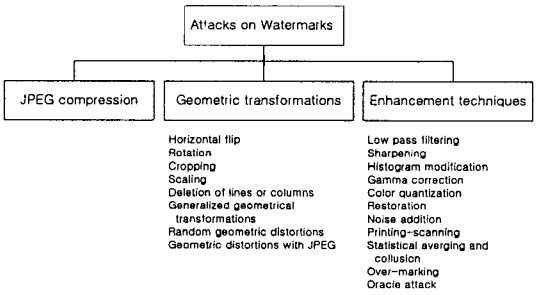

그림1. Kutter 등이 분류한 공격방법

또 다른 분류로 최근 Voloshynovskiy 등이 분류한 방법이 있다[7]. 이것은 Kutter와는 달리 암호학적으

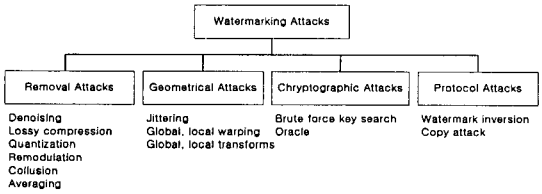

그림2. Voloshynovskiy 등이 분류한 공격방법

로 키를 찾아내는 암호학적 공격과 카피 공격 등과 같은 프로토콜 공격이 추가되었으며, 기존의 공격들은 소거공격과 기하학적 공격으로 분류 시키고 있다.

2.2 StirMark[8]

StirMark는 정지영상에 대한 디지털 워터마킹 알고리즘을 평가하기 위한 방법으로 1997년 v.1.0이 발표되었다. 그후 계속 수정되어지면서 현재 오디오, 동영상 워터마킹 알고리즘으로 확장 가능한 StirMark v.4.0이 나와있다. StirMark는 벤치마킹 방법들 중에서 가장 대표적인 것으로 강인성에 주안점을 둔 벤치마킹 방법이다.

StirMark v.4.0의 특징을 보면 워터마킹 기술은 서로 다른 목적으로 많은 응용 분야에 적용되기 때문에 벤치마킹 전에 사용 목적과 대상을 선택하도록 하고 있다. 즉, 영상의 품질 계수, 강인성, 공격의 강도 등을 적용하여 벤치마킹을 실시 한다.

StirMark는 벤치마킹을 이용한 어플리케이션이 나오지 않고 있으며, 척도로 이용되는 PSNR은 시각적인 척도로는 부적합하다. 또한, 공격 기법이 기하학적 변환에 많은 비중을 두고 있다. 그리고, 삽입·추출에 걸리는 시간을 고려하지 않고, 이미지의 사전 정보를 고려하지 않은 문제점들을 가지고 있다.

2.3 Checkmark[9]

Checkmark는 StirMark의 여러 가지 문제점을 개선하기 위해 만들어 졌다. 크게 Kutter의 공격 기준에서 Voloshynovskiy의 분류로 변경되면서 새로운 공격들이 추가 되었다. 기존의 JPEG 압축만을 고려하던 부분이 JPEG2000에서 사용되는 wavelet 압축을 포함하고 있으며, 이전에 워터마크가 삽입된 영상으로부터 워터마크를 추출하여 그 정보를 공격할 데이터에 다시 삽입하는 copy attack이 새로이 추가 되었다. 이외에도 여러 가지 신호처리적 공격들이 새롭게 포함되어 워터마크 알고리즘의 강인성 테스트를 실시한다. Checkmark는 이러한 공격 유형뿐만아니라 새로운 화질 평가 방법으로 weighted PSNR과 Watson metric을 도입하고 있다. 그리고, 출력 양식을 XML 포맷으로 나타내며, 결과 테이블은 HTML로 작성 한다. 소스는 StirMark와 달리 Matlab으로 작성되어 일반 사용자가 다루기 쉽게 되어있다.

2.4 JEWELS[10]

JEWELS는 일본 전자정보기술산업협회(JEITA)에서 워터마크 평가 지원 시스템(JEWELS)으로 개발하여, 정지영상의 저작권 보호 및 이용자의 이용 범위에 대한 가이드 라인 설정을 목적으로 디지털 워터마킹 기술의 안전성 평가를 위한 벤치마킹 프로그램이다. JEWELS는 전체적으로 11개 항목의 내성평가와 15가지의 영상 변환공격을 포함하고 있다. 각 항목들은 여러 개의 파라메터를 가지면서 다양한 공격이 가능하도록 구성되어 있다.

3. 국내 워터마킹 기술 평가 동향

디지털 워터마크 기술 평가와 관련된 국내 활동은 정지영상에 비해 오디오 컨텐츠를 대상으로 한 평가 기준이 먼저 나왔다. 2000년 5월 산업자원부 컨텐츠

저작권보호 시범 사업으로 컨텐츠 유료서비스 활성화 및 저작권 보호에 관한 인식을 확산시키기 위해 SDM(Secure Digital Multimedia) 포럼이 92개의 디지털 컨텐츠 관련 회원사와 함께 음반저작권 관리 표준화를 위해 오디오 워터마크 기술 평가를 실시하였다.

SDM 포럼의 오디오 워터마크 기술 평가 기준은 인코더로부터 72비트의 워터마크 정보(0x0~0xF의 16진수 값)를 삽입하고 디코더에서 30초마다 1회 이상의 워터마크 정보를 정확히 검출해야 한다. 또한, 허위로 워터마크 정보를 표시하는 것을 방지하기 위해 각 테스트 이후에 false watermarking detection을 체크하게 된다. 그리고, 구현성 평가를 위해 삽입 및 검출시간을 원곡의 50% 이내로 규정하고 있다. SDM의 강인성 평가 기준은 표1과 같다.

표1. SDM의 robustness 테스트 항목

Signal process	Description
Echo	Delay : 100ms, Feedback coefficient : 0.5 Delay : 200ms, Feedback coefficient : 0.25
Equalization	Typical case:10-band graphic equalizer with the following characteristics Freq. 31 62 125 250 500 1k 2k 4k 8k 16k Gain. -6 +6 -6 +6 -6 +6 -6 +6 -6 +6
Band-pass filtering	100Hz - 6kHz, 12dB/oct
Linear speed change	±10%
Time scale modification	pitch invariant time scaling : ±4%
Codecs	MPEG1 Audio Layer3(MP3):128kbps(stereo) MPEG2 AAC : 128kbps(stereo)
Noise	White noise : S/N -36dB
Cropping	30초 이상
False watermarking detection	원본에 대해서 워터마크 유무 판정

정지영상을 대상으로 하는 이미지 워터마크 기술 평가는 현재 SEDICA(Secure Digital Content Association)의 워터마킹 분과에서 디지털 워터마킹 기술의 국제적 평가 활동에 초점을 맞추어 워터마킹 기술을 연구하고 평가 시스템 개발 및 관련된 규격설정, 시나리오 및 모델 개발을 진행중에 있다.

4. 워터마킹 기술에 대한 평가 프로파일 개발

평가 프로파일은 워터마킹 기술을 적용하려는 시스템 또는 알고리즘 개발자가 시스템에 적합한 워터마킹 알고리즘의 안전성을 보장하는 기술 요구사항 등을 기술한 것이다. 저작권자나 저작권 보호 시스템 개발자는 평가 프로파일에 의해 사용 시스템에 최적인 알고리즘을 선택할 수 있다. 평가 프로파일은 평가 대상 알고리즘에 대한 기능 요구사항을 시스템 구현과는 독립적으로 표현해야 하며, 궁극적으로 알고리즘 이용자에게 적합한 알고리즘 선택의 기준을 제시하는데 있다. 평가 프로파일 작성시 다음 기준을 만족해야 한다.

- 평가의 대상(TOE)을 명료하게 확인해야 한다.
- 평가에 기본적으로 평가 대상 알고리즘에 의존해서 사용할 명세와 요구사항의 프로파일을 선택해야 한다.
- 평가기준을 선택할 때 보안 목적에 따라 일정한 보증등급을 선택, 적용해야 한다.

4.1 지각성에 대한 보증

워터마킹 기술에 대하여 제일 먼저 실시되어야 할 평가 항목 중의 하나가 지각성 평가이다. 지각성은 적용 대상에 따라 서로 다른 등급으로 평가되어야 한다. 대부분의 워터마크 기술은 일반 사용자 환경, 저작권자 환경, 그리고 엄격한 평가 실시 환경에서 워터마크 삽입의 흔적이 보이지 않도록 구성하게 된다. 따라서 이런한 환경을 고려하여 표2와 같은 네가지 등급으로 지각성을 보증한다.

표2. 지각성에 대한 보증 레벨

보증등급	내용
Perceptible Level 1	지각적으로 미세한 변화가 눈에 보이지만 감각적으로 거슬리지 않는 정도
Perceptible Level 2	일반 사용자 환경하에서 지각되지 않는 정도
Perceptible Level 3	studio 환경에서 원본과 비교했을 때 지각되지 않는 정도
Perceptible Level 4	엄격한 환경에서 다수의 평가자에 의해 평가되어 지각되지 않는다고 판단된 정도

4.2 JEWELS를 이용한 강인성에 대한 보증

디지털 워터마크 기술은 사용 환경에 따라서 일반적인 영상처리 공격에 대한 강인성 평가항목은 적절하지 않을 수 있다. 그래서 디지털 워터마크를 이용하는 사용자가 사용목적에 맞게 임의의 영상처리를 추가할 수 있는 평가구조로서 유연성을 가지도록 구성하는 것이 바람직하다.

JEWELS의 평가항목을 공격 형태에 따라 그림3과 같이 새롭게 분류한다. General attacks은 일상 생활에서 기본적으로 쉽게 일어날 수 있는 공격들을 정의하고 있으며, Geometric attacks은 영상의 기하학적 변환을 가져오는 공격들로 구분 시켰다. 그리고, Malicious attacks은 제 3자의 고의적인 변환에 의해 워터마크를 삭제하려는 시도로서의 공격 형태를 분류하였다.

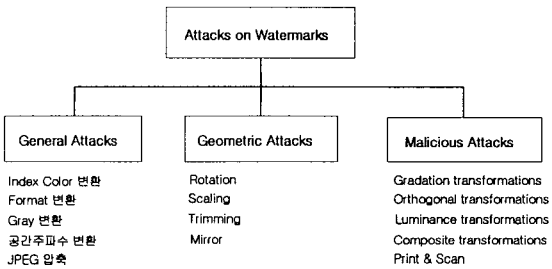

그림3. JEWELS의 공격형태 분류

표3은 분류된 공격 형태에 의해 워터마킹 알고리즘 등급에 따른 안전성 평가 요구사항을 나타내고 있다. 알고리즘의 전체 등급을 K1~K4로 네 등급으로 분류하였으며, 가장 강인한 알고리즘 평가는 K4등급 평가이다. 표3에서 '○'은 해당 공격에 대하여 워터마크가 정확히 검출되어야 함을 나타내고, '×'는 해당 공격에 대한 안전성 평가를 하지 않음을 나타낸다. '-'는 응용 분야에 따라 적절히 만족해야 함을 나타내고 있다.

표3. 공격에 대한 등급별 요구사항

Attacks	Parameter	K1	K2	K3	K4
Index color	·	○	○	○	○
Format 변환	BMP→JPEG	×	○	○	○
	JPEG→BMP	×	○	○	○
Gray 변환	·	○	○	○	○
공간주파수 변환	8	○	○	○	○
	16	○	○	○	○
	54	○	○	○	○
	40	○	○	○	○
JPEG 압축	1/5	○	○	○	○
	1/10	○	○	○	○
	1/20	×	○	○	○
	1/40	×	○	○	○

(a) General attacks에 대한 평가항목

Attacks	Parameter	K1	K2	K3	K4
90° Rotation	·	○	○	○	○
확대	4/3	×	○	○	○
	2	×	○	○	○
	3	×	×	○	○
	4	×	×	-	-
축소	3/4	×	○	○	○
	1/2	×	○	○	○
	1/3	×	×	○	○
	1/4	×	×	-	-
Trimming	3/4	×	○	○	○
	1/2	×	○	○	○
	1/3	×	×	○	○
	1/4	×	×	-	○
Mirror	수평	○	○	○	○
	수직	○	○	○	○

(b) Geometric attacks에 대한 평가항목

Attacks	Parameter	K1	K2	K3	K4
계조변환	0.5	×	○	○	○
	0.75	×	○	○	○
	1.25	×	×	○	○
	1.5	×	×	○	○
사영변환 (X,Y)	(5,1)	×	-	○	○
	(10,3)	×	×	-	○
	(20,6)	×	×	×	-
	(30,9)	×	×	×	-
회도변환	2	×	×	○	○
	4	×	×	×	-
색조변환	5	○	○	○	○
	10	×	○	○	○
합성변환	Gray	×	○	○	○
	Color	×	○	○	○
프린터/스캔	·	×	-	○	○

(c) Malicious attacks에 대한 평가항목

4.3 평가 프로파일 제시

다음은 신분증 위·변조 방지용 워터마크 기술에 대한 평가 프로파일을 구성한 것이다.

[신분증 위·변조 방지용 워터마킹 기술]
- 용도 : 신분증의 특정위치에 부착되는 증명사진의 위·변조 방지용
- 지각성에 대한 보증레벨 : Perceptible Level 1
- 강인성에 대한 보증레벨 : K2(프린터/스캔 포함)
- 테스트 이미지 : classics, smooth areas, textures & fine details
- 평균 검출 속도 : Real time

위·변조 방지를 위한 워터마크 알고리즘은 일반적인 신호처리 공격과 외부 충격 등에 의해 손상되는 공격들에 대해서 강인성을 유지해야 한다. 그러나, 신분증 이미지에 대한 확대, 축소, 계조변환, 사형변환, 휘도 변환 등과 같이 이미지가 가지는 값을 변경하는 공격은 일어나지 않는다.

5. 결론

디지털 워터마킹은 저작권 보호 이외에 다양한 응용분야에서 이용되고 있으며, 많은 알고리즘이 개발, 적용되고 있다. 또한, 새롭게 개발된 많은 알고리즘의 안전성 평가에 대한 필요성도 증가되고 있다. 본 논문에서는 JEWELS를 이용하여 워터마킹 기술에 대한 안전성을 평가하는 기준을 제시하였다. 본 평가기준은 각각의 응용분야에 따라 평가 프로파일을 다르게 설정하여 시스템의 특성에 맞게 평가하고, 새로운 공격유형 생성시 평가 프로그램에서 확장 가능하도록 구성하였다.

지금까지 워터마킹 알고리즘 평가 시스템은 국외에서 주도적으로 진행되고 있으며, 그 대상은 저작권 보호에 한정되어 있다. 따라서, 본 연구를 통하여 국내의 워터마킹 알고리즘 평가 시스템에 관한 연구를 활성화할 수 있을 것이다. 이를 위해, 국내 벤치마킹 프로그램 개발이 진행되어야 하며, 세분화된 공격형태 분류와 응용 분야에 따른 다양한 프로파일 작성이 필요하다.

[참고문헌]

[1] I. J. Cox, J. Kilian, T. Leighton, T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", IEEE Trans. on Image Processing, Vol. 6, No. 12, pp.1673-1687, 1997.

[2] R. G. van Schyndel, A. Z. Trikel and C. F. Osborne, "A Digital Watermark", Proc. IEEE Int. Conf. Image Processing, Vol. 2, pp.86-90, 1994.

[3] D. Kundur, D. Hatzinakos, "Digital Watermarking Using Multiresolution Wavelet Decomposition", International Conf. on Acoustic, Speech and Signal Processing, Vol. 5, pp.2969-2972, 1998.

[4] F. Hartung, B. Girod, "Watermarking of Uncompressed and Compressed Video", Signal Processing 66(1998), pp.283-301, 1997.

[5] C. Podilchuk, W. Zeng, "Image Adaptive Watermarking Using Visual Models", IEEE Journal on Selected Areas in Communication, Vol. 16, No. 4, pp.525-539, 1998.

[6] M. Kutter, F Petitcolas "A Fair Benchmark for Image Watermarking Systems." Proc. of SPIE: Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 226-239, 1999.

[7] M. Kutter , S. Voloshynovskiy, A. Herrigel, "The Watermark Copy Attack", Proc. of SPIE: Security and Watermarking of Multimedia Contents II, Vol. 3971, 2000.

[8] http://www.cl.cam.ac.uk/~fapp2/watermarking/stirmark/

[9] http://watermarking.unige.ch/Checkmark/

[10] http://www.jeita.or.jp