EWM 통계량을 이용한 스테가노그래픽 자료 감지 기법†

(Detecting Steganographic Contents Using EWM Statistics)

지 선 수*
(Seon-su Ji)

요 약 가장 일반적이고 효과적으로 사용하는 혼합된 정보은닉 기술인 스테가노그래피에서 의사소통의 존재를 숨기면서 송수신하는 자료은닉 기술과 관련된 통계적 기법을 적용하는 연구가 필요하다. 즉, 인터넷상에 존재하는 임의의 원본이미지에 비밀(은닉) 메시지가 포함된 변조된 혼합이미지를 가장 효과적으로 관리하고 찾아내는 감지 기법의 연구가 필요하다. 이 논문에서 원본이미지에 비밀자료를 숨기기 위한 스테가노그래피에 RGB, DCT 및 EWM 통계기법을 이용하여 은닉자료를 감지하고 그 위치를 찾는 기법을 확인한다. 그리고 카이스퀘어 검정법을 이용하는 기존의 방법과 비교한다.

핵심주제어 : 스테간분석, 스테가노그래피, 이산코사인변환, 자료은닉, 카이스퀘어 검정, EWM

Abstract For a message hiding technique to be effectual, it needs to have availability, confidentiality and integrity. Steganography is the science of hiding one message within other types of digital contents. In this case, attempt to defeat steganalysis by restoring the statistics of the composite image to resemble that of the cover, these discrepancies expose the fact that hidden communication is happening. In this paper, I present a steganography scheme capable of concealing a piece of secrete information in a host image and base on the technique's DCT, RGB, statistical restoration.

Key Words : Chi-square Test, Data Hiding, DCT, EWM, Steganalysis, Steganography

1. 서 론

암호개념은 인류역사가 시작되면서부터 이용되어 온 것으로 현대사회에서 객체간의 비밀 보장의 필요성이 증대되면서 암호기법에 대한 연구가 급속히 확산되어지고 있다. 최근의 암호기법은 인터넷과 연결된 컴퓨터 도구의 중요한 분야로 인식되어 활발하게 연구되어지고 있다. 정보보호의 필요성은 수천 년 전부터 매우 중요한 개념으로 인식되어 왔다. 경쟁자의 비밀정보를 가로채어 자신에게 유리한 정보를 획득, 변조하고자하는 노력과 상대방에게 비밀정보가 노출되지 않도록 하는 연구는 과거, 현재 그리고 미래에 끊임없이 계속될 것이다.

정보를 보호하는 완벽한 송수신 수단은 현재 존재하지 않으며, 인터넷 송수신 상의 정보는 항상 제 3자에게 위변조, 탈취될 수 있다는 것을 가정하여야 한다. 이러한 현실적 상황에서 암호화된 정보를 획득하더라도 원래의 의미를

† 이 논문은 2007년도 강릉대학교 학술연구조성비 지원에 의하여 수행되었음.

* 강릉대학교 컴퓨터정보공학부 교수

해독하기 위해 암호화된 정보를 분석하는데 시간과 비용이 은닉정보의 가치에 비해 많이 발생하도록 하는 것이 암·복호화 시스템을 개발하는 가장 중요한 핵심이다. 최근에는 정보의 불법유출 및 수정이 국가적, 사회적, 개인적 문제화 되고 있다. 따라서 물리적 대책, 인적자원에 대한 관리대책, 기술적 대책, 법적·제도적 대책 중에서 정보 시스템이 요구하는 정보의 보안 수준에 따라 효율적이고 계층적인 보안 대책을 제공할 수 있는 방법 중에 가장 경제적이고 안정성이 뛰어난 암호화된 정보전달 기법을 이용하는 것이다.[17][18]

고대 암호 기법은 전쟁과 외교 분야에 이용되었지만 현대 정보사회에서는 암호화 기법이 상업적으로 이용되고 있다. 즉, 전자상거래의 핵심요소인 전자화폐, 결재, 투표, 인터넷뱅킹 구현 등 광범위한 범위에서 암호 기법들이 적용되어 우리의 일상생활에 깊숙하게 활용되고 있다. 멀티미디어 산업이 폭발적으로 팽창하고 네트워크 기술이 발전하면서 문서, 사진, 음성, 동영상 자료와 같은 다양한 콘텐츠들이 디지털화되어 웹 브라우저 상에서 제공되는 자료의 원본과 똑같은 형태로 자신의 컴퓨터에 효율적으로 저장하여 활용할 수 있게 되었다. 그러나 이와 같은 편리성만큼의 역기능이 발생할 수 있다. 즉, 콘텐츠 사용자가 정보를 이용하는데 필요한 승인, 보상 및 접근제한 문제, 제 3자에 의한 가로채기와 위변조, 콘텐츠 소유권에 대한 권리주장에 이르기까지 다양한 문제가 발생되는 부작용을 낳게 되었다.[6] 그러므로 네트워크상에서 송수신되는 콘텐츠에 대해 암호화를 통한 정보보호의 필요성이 대두되었다.

정보은닉 기술에서 스테가노그래피는 의사소통의 존재를 숨기면서 통신하는 기술이다. 즉, 어떤 송신자와 수신자 사이에 통신이 존재한다는 것을 숨기는 것이다. 스테가노그래피는 고대 그리스로부터 유래된 것으로 '덮어진 쓰기(covered writing)'라는 의미를 갖고 있다. 즉, 임의의 정보 안에 다른 정보를 숨기는 방법을 의미하며 이러한 관점에서 카피라이트 마킹과 매우 유사한 특징을 갖고 있다. 스테가노그래피는 그래픽의 화소들 사이에 메시지를 숨기는 기술로서 메시지를 숨기는 것과 더불어 그 메시지의 전송 여부 자체를 알지 못하게 하는 것이 주목적이다. 최근 스테가노그래피 기법은 자료를 JPG, BMP와 같은 이미지 파일 형태로 암호화하여 WAV 확장자를 갖는 음성파일 안에 은닉하는 방법을 적용한다.[5][6][7][11]

인간의 시각이나 청각을 통할 경우 암호화된 메시지를 포함했을 때와 포함되지 않았을 때와의 차이가 없으며, 전송할 경우에 일반 파일과 동일하게 보이기 때문에 정보노출의 위험이 최소화된다는 장점을 가지고 있다. 그러나 스테가노그래피 기법의 치명적 약점 중의 하나는 다량의 데이터를 은닉하는 경우에 제약이 따른다는 것이다. 일반적으로 원본이미지 파일의 전체 비율 중 숨기고자 하는 정보의 비율이 15%가 넘을 때 조작된 이미지 파일은 관찰자의 시각만으로도 확인이 가능할 수 있게 된다. 따라서 일반적으로 원본이미지에 비해 은닉자료의 크기가 5~10% 내에서 정보은닉 기법을 활용한다.[1][15] 또 다른 하나는 웹사이트에 비밀메시지가 포함된 혼합이미지를 올려놓을 때 미세하지만 RGB 값이 변경될 수 있다는 것이다. 그럼에도 불구하고 만약 정보를 은닉한 이미지 파일을 획득하더라도 콘텐츠를 확인하기 위해서 암호 운영자가 설정한 비밀번호를 알고 있어야만 가능하기 때문에 현실적으로 은닉정보를 확인하기란 매우 어렵다.

정보은닉 기술에서 스테가노그래피는 통신의 존재를 숨기면서 통신하는 기술과 관련되어 통계적 기법을 적용하는 연구가 필요하다. 즉, 인터넷상에 존재하는 임의의 원본이미지(original/cover secrete image/data)에 비밀 메시지(image/data)가 삽입된 변조된 이미지(stego image/data)를 가장 효과적으로 찾아내는 다양한 감지 기법의 연구가 필요하다. 따라서 제안하고자 하는 연구는 다음과 같다. 일반적으로 chi-square 통계기법을 이용하여 숨겨진 이미지의 존재유무를 찾아내지만 원본이미지에 비해 숨기는 자료의 크기가 매우 작을 때는 은닉자료의 존재를 감지 할 수 없다. 제안하는 연구에서는 지수가중이동(exponentially weighted moving : EWM) 통계량을 적용하여 원본이미

지에 비해 은닉자료의 크기가 매우 작을 경우 효과적으로 암호화된 자료를 감지하고 그 위치를 찾아내는 기법을 제안한다.

2장에서 정보은닉과 관련된 스테가노그래피에 대한 관련 연구에 관하여 조사하고, 3장에서는 EWM 통계량을 이용한 통계적 모델을 제시한다. 4장에서의 모의실험 결과를 가지고, 5장에서 결론을 제시한다.

2. 관련연구

FKI[18]는 2005년 기준으로 산업기술 해외유출 방지활동을 통한 경제적 효과가 32조 9천억원에 이르는 등 기술유출 방지 및 정보보안은 국가 및 국내기업 생존의 관건이며, 기술 유출 방지 및 보안을 위해서는 사용자의 보안의식 제고와 함께 보안역량의 강화가 시급하다고 주장하였다. 따라서 컴퓨터와 관련된 모든 분야에서 암호화 및 정보은닉기술은 절대적으로 중요한 역할을 하게 되었다. 암호화의 주된 목적은 뒤섞기 정보에 의해 신뢰성, 비밀성과 안정성을 제공하는 것이다. 고전적인 스테가노그래피 체계는 암호화시스템을 은밀하게 유지하려는 반면에 현대 스테가노그래피는 은밀한 정보 즉, 은밀한 키(열쇠)가 알려지지 않는 한 감지할 수 없도록 위장을 시도한다. 일반적으로 정보은닉 과정은 덮개매체에 있는 중복(잔여, redundant) 비트를 구분하는 것으로 시작한다. 중복비트는 덮개매체의 완전한 파괴 없이 변경될 수 있는 비트이다. 끼워 넣는 과정은 은밀한 메시지에서 자료를 보전하기 위해 중복비트의 부분 집합을 선택한다. stego매체는 메시지 비트를 가지고 중복비트를 대치하는 것에 의해 만들어 진다.[4][5]

2.1 시각 암호법

1990년대 중반 일부학자들에 의해 제안된 방법으로 비밀 메시지를 들춰내기 위해 개별적인 이미지가 고려되고, 대응되는 이미지는 키로 이용되는 1회용 암호의 그래픽 형태로 해석될 수 있다. 또한 흑백 화소 이미지로 유효한 은밀한 정보의 부호 매김과 배분에 적당하다. Naor와 Shamir는 시각적 암호를 위해 정보를 숨기는 새로운 방법을 제시하였는데 두 개의 깨끗한 이미지를 이용하여 비밀메시지를 숨기는 기법을 사용하였다. 즉, 정보를 숨기기 위해 시각적 암호와 스테가노그래피의 조합인 혼합기술을 이용할 것을 제안하였다.[9]

2.2 이미지 스테가노그래피

이미지 스테가노그래피는 많은 학자들에 의해 폭넓게 연구되어졌다. 정보를 이미지에 숨기는데 이용된 여러 가지 방법을 이용하였다. 즉, 첫 번째로 원본이미지에서 각 픽셀의 최하위비트(least significant bit : LSB)에 숨겨지는 자료를 위치시키는 방법, 두 번째로 원본이미지에 비밀메시지를 끼워 넣기 위해 각 픽셀의 적당한 유효비트를 가지고 덮개매체에 수정을 최소화하는 가능성을 보이는 방법, 세 번째로 정보를 숨기기 위한 디지털 매체의 변환영역을 이용하는 방법 -DCT(discrete cosine transform) 혹은 DWT(discrete wavelet transform)와 같은 함수가 폭넓게 이용되었다- 등이 사용되었다. Zöllner 등은 정보이론상 결정적이지 않은 선택을 채택하여 안전한 스테가노그래피의 문제를 해결하기 위하여 접근법을 제시한다. 그들의 모형에서 원래 매체는 상대에게 알려졌으나 전처리 단계는 덮개매체로 랜덤성(randomness)을 소개하였다. 상대가 변형된 덮개매체를 얻을 수 없는 경우에 경쟁자는 원래매체와 stego매체 차이를 관찰해서 은닉되어진 메시지에 관하여 정보를 추론할 수 없다. 즉, 안전한 스테가노그래피를 위해 숨기려는 메시지를 은닉하는데 이용한 비밀키는 상대에게 알려지지 않으며 상대는 덮개매체를 모른다는 필요조건을 제시하였다. 이것은 덮개매체를 디지털 카메라 또는 그림을 스캔닝하는 것에 의해 덮개매체를 만드는 것이 충분함을 보여주었다.[15] Provos 등은 검출의 가능성을 줄이는 방법으로 변경할 비트에 정보를 끼워 넣는 과정이 선택되는 것을 허용하는 오류수정 부호를 채택하여 그럴듯한 법적 부인권을 제공하기 위하여 동일한 덮개매체에 있는 다량

의 자료집합을 은닉할 수 있다는 것을 보였다. 또한 정보를 변환하기 위해 LSB를 통한 정보은닉 도구인 Outguess에서 초기에 히스토그램 보완을 시도하였다.[10][11][12]

Potar과 Chang은 개인적인 비밀성과 신뢰성을 향상시키기 위해 암호작성에 위장요인을 제공하는 것을 제안하였으며 다음과 같은 4단계의 해법을 제시하여 구현하였다. 즉, 변장암호를 위한 이미지 선택단계, 문자자료를 숨기기 위한 이미지 영역의 선택단계, 공간영역 안에 수정된 이미지에 의한 정보은닉단계, 복호화를 이용한 숨겨진 문자를 복원하는 단계 등을 통하여 이미지 암호를 가지고 문자암호를 추출하는 기법을 제시하였다. Solanki와 Sullivan 등은 통계적 스테가노 분석을 피할 수 있는 은닉하는 기법의 설계를 제안하고, 양자화 인덱스 변조(quantization index modulation : QIM)의 타당성을 보였으며, 통계적 복원은 DCT-히스토그램과 봉쇄성 방법 모두의 스테간 분석 수행능력에 영향을 줄 수 있음을 보였다.[14][15] Dang과 Kota는 주파수영역으로 공간영역에서 픽셀 값에 대응하기위해 이산 퓨리에변환 함수를 이용하였으며 이러한 접근형태가 숨기려는 메시지의 능력에는 제한이 있으나 좀 더 좋은 견고성을 제시할 수 있음을 보였다. 즉, 숨겨진 메시지의 크기를 줄이는 압축과 암호화와 스테가노그래피를 성공적으로 혼합시키는 것에 의해 이미지에 비밀자료를 숨기는 것을 허용하는 도구를 C#.NET 라이브러리 등을 이용하여 개발하였다.[3]

Provos와 Honeyman 등의 학자들에 의해 일반적인 스테가노그래픽 도구에 대응하여 시각 및 통계적인 공격 방법에 대해 설명하였다. 그들은 일반적인 스테가노그래픽 기술이 덮개매체에 있는 통계적인 특성을 변경시킬 수 있다는 것을 제시하였다. 예를 들어 그들은 JPEG 이미지에 있는 포함된 자료의 특정한 프로그램을 평가하였다. 이러한 프로그램에 의해 은닉된 숨겨진 정보를 감지하기 위하여 그들은 숨겨진 정보를 나르는 이미지의 색상분포를 추정하고, 관찰분포에 대응하는 것과 비교하는 χ^2-검사기법을 이용하였다.[10][11][12]

2.3 오디오 스테가노그래피

오디오 스테가노그래피는 이미지 파일의 최하위 비트 변경과정과 유사하다. 오디오 파일의 몇몇 비트의 LSB를 변경하여 기록하며, 사소한 변화는 최대량이 청각에 의해 분화될 수 없는 소리에서 일어난다. 현재 가장 폭넓게 연구되어지는 분야로서 오디오 신호에 외부자료를 끼워 넣는 방법으로 낮은 비트암호화 방법, 단계코딩, 확산스펙트럼방법과 청각마스킹을 이용하여 기존의 암호법 기술과 결합하여 사용된다. Cvejic와 Sepp는 저전력 음색 및 이것의 강력함을 잡음에 삽입하고 끼워 넣어진 음성표면을 획득하여 오디오 신호에 있는 자료를 혼합하는 것을 제시하였으며, 부가적인 잡음 및 무작위 획득 등을 이용하여 비밀스러운 의사소통을 위한 스테가노그래피가 강력하고 안전하다는 것을 보였다.[2]

미래의 스테가노그래피는 음성과 이미지를 혼합 사용하여 오디오와 시각적인 비밀 정보를 보호하는 방향으로 흐르고 있다. 인터넷과 전자상거래의 폭발적 성장과 개방성 추구는 의사소통을 위한 통신에서 향후 부당한 전자적 공격에 취약한 구조가 되는 원인이 될 것이며 네트워크 자산에 대한 공격의 증가되는 양과 복잡화는 자료파괴, 신뢰성 상실, 생산성 하락 등으로 연결될 수 있다. 그러므로 네트워크 자원의 신뢰성, 무결성, 인증, 가용성 등을 보장하기 위한 정보보호 조치는 필수적인 요소이며 인터넷 운영의 주된 핵심체가 될 수밖에 없다.

3. 통계적 모델

스테가노그래피의 구성요소는 원본(커버)메시지와 삽입메시지로 구분된다. 커버메시지는 제3자에게 아무런 의심 없이 전달되는 실제적인 의미가 없는 메시지 즉, 의도하고자 하는 사용자에 의해 비밀메시지를 숨길 수 있는 데이터이다. 삽입메시지는 두 집단 간에 비밀스럽게 전달되는 실질적인 메시지이다. 또한 커버메시지에 삽입메시지를 숨기는 과정에서 메시지 검출

을 제한시키기 위해 키를 설정하는데 이것을 '스테고키'라고 한다. 이렇게 커버메시지와 삽입 메시지가 혼합되어 수신자에게 실질적으로 전달되는 데이터를 '스테고데이터'라고 한다. 즉, 스테가노그래피는 암호화된 메시지를 동영상, 이미지, 음악파일 등 다른 데이터 안에 혼합하여 제3자가 내부에 숨겨진 데이터 자체를 알아차리지 못하도록 은닉하여 전달시키는 방법이다.[5][6][7][11]

인터넷상에서 이미지를 송신할 때 압축하여 웹에 올려놓으며 이미지 압축과정에서 이미지의 정보를 한 곳에 집중시키는 능력이 뛰어나고 또한, 계산이 단순하며 성능이 좋은 이산코사인변환 기법을 사용한다.

커버이미지에 비밀메시지가 삽입된다면 삽입된 부위에 RGB 값의 변화가 있을 수밖에 없다. 각각의 이웃한 비트에서 계산한 RGB 값의 차이를 이용하여 미세한 외부변화를 정확하게 감지할 필요가 있다. 즉, 이미지를 64(8×8) 혹은 256(16×16) 비트씩 하나의 블록으로 설정하여 이웃한 비트의 RGB 값을 비교하여 위·변조된 이미지를 찾아낼 수 있다.

n_{01}	n_{02}	n_{03}	n_{04}	n_{05}	n_{06}	n_{07}	n_{08}
n_{09}	n_{10}	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{16}
n_{17}	n_{18}	n_{19}	n_{20}	n_{21}	n_{22}	n_{23}	n_{24}
n_{25}	n_{26}	n_{27}	n_{28}	n_{29}	n_{30}	n_{31}	n_{32}
n_{33}	n_{34}	n_{35}	n_{36}	n_{37}	n_{38}	n_{39}	n_{40}
n_{41}	n_{42}	n_{43}	n_{44}	n_{45}	n_{46}	n_{47}	n_{48}
n_{49}	n_{50}	n_{51}	n_{52}	n_{53}	n_{54}	n_{55}	n_{56}
n_{57}	n_{58}	n_{59}	n_{60}	n_{61}	n_{62}	n_{63}	n_{64}

일반적으로 공간영역과 주파수 영역을 상호 변화시키는 매핑기술의 대표적인 DCT는 임의의 데이터 배열을 코사인 함수의 합으로 표현할 수 있다는 성질을 이용한 것이다. 일반적으로 계산의 복잡성 때문에 2차원 N×N DCT 계수(coefficient)는 다음의 공식에 의해 구하여 사용할 수 있다.[17][20]

$$C(u,v)=\alpha(u)\alpha(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}f(i,j)\cdot \cos\left[\frac{\pi}{N}(i+\frac{1}{2})u\right]\cos\left[\frac{\pi}{N}(j+\frac{1}{2})v\right] \quad (1)$$

여기에서 u, v는 2차원 N×N 블록에서 계수의 위치를 나타낸다. u, v = 0, 1, 2, ..., N-1에 대해 $\alpha(u)$와 $\alpha(v)$는 다음 식에 의해 결정된다.

$$\alpha(u)=\begin{cases}\frac{1}{\sqrt{2}}, & u,\ v=0\\ 1, & \text{이외의 경우}\end{cases} \quad (2)$$

$f(i,j)$는 DCT 변환이 이루어지기전의 블록내에서 0부터 255사이의 픽셀(밝기)값을 의미한다. i, j는 블록내에서 픽셀의 위치를 가르킨다.

지금까지의 대부분 연구에서 변조된 이미지를 감지하기위해 이웃한 RGB 값을 기반으로 한 카이스퀘어 검정방법을 이용하였다. 이웃한 값의 평균을 다음과 같이 계산할 수 있으며,

$$y_i^* = \frac{n_{2i}+n_{2i-1}}{2}$$
$$y_i = n_{2i}, \quad (3)$$

이것을 이용하여 카이스퀘어 통계량을 계산한 후 기각역과 비교하여 비밀자료의 삽입 유무를 판단한다. 즉, $\chi^2 > \chi^2_{(df,\ \alpha)}$이면 이미지에 비밀메시지가 포함되어 있다고 판단한다.

$$\chi^2 = \sum_{i=1}^{df+1}\frac{(y_i-y_i^*)^2}{y_i^*} \quad (4)$$

여기에서 df는 자유도, α는 유의수준(Type I error)을 나타낸다.

일반적으로 카이스퀘어 검정법을 이용하여 은닉자료의 존재유무를 정확하게 판단할 수 있다. 그러나 미세한 크기의 존재유무를 확인하기에는 부족한 면이 있으며 은닉자료의 숨겨진 위치를 찾아낼 수 없는 단점이 있다. 따라서 자료의 미세한 변화에 예민하게 적용하며, 이웃한 자료와

의 관계 등을 고려할 때 이웃한 두 값의 차를 고려한 지수가중이동(EWM) 통계량을 이용하는 것이 합리적이다.

$$EWM_i = \lambda \cdot d_i + (1-\lambda) \cdot EWM_{i-1} \qquad (5)$$

여기에서 $d_i = (n_{2i} - n_{2i-1}) + \epsilon_i$을 고려하며, ϵ_i는 평균이 0이고, 분산이 σ^2인 정규분포를 따르는 백색잡음으로 간주한다.

이웃한 비트의 RGB 값의 급격한 변화에 따른 기각역은 다음과 같이 쓸 수 있다.

$$\mu_o \pm K\sqrt{\frac{\lambda}{2-\lambda}} \cdot \sigma_o \qquad (6)$$

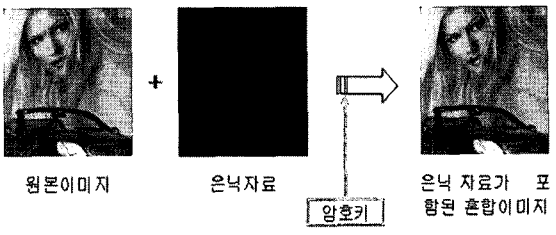
여기에서 K는 임의의 상수이며, μ_o, σ_o는 관리값(control value)이다. $0 < \lambda \leq 1$이며 λ는 평활상수(smoothing constant)이다. λ를 결정하는 이론적인 방법은 없으며 경험적으로 $0.10 < \lambda < 0.35$을 사용한다.

미세한 외부변화를 감지하기 위해 이미지를 64비트의 블록단위로 (5)식과 (6)식을 이용하여 이웃한 비트의 RGB 값의 차이를 표시하여 기각영역을 벗어났는지 혹은 영역 안에 있는지를 확인하여 은닉메시지의 존재유무를 판단한다. 즉, 커버이미지에 삽입자료가 포함되어 있는지의 존재유무를 확인할 수 있으며 또한, 그 위치를 찾아낼 수 있다.

4. 적용기법

일반적으로 커버이미지에 비해 삽입메시지의 크기가 3% 이상이 되어야 은닉메시지를 쉽게 감지할 수 있다. 또한 이미지의 경우에는 커버이미지에 비해 삽입이미지의 크기가 16% 이상이 되어야 한다. 여기에서는 커버이미지에 비해 삽입메시지가 크기가 1% 이하의 자료가 숨겨지는 경우에 메시지의 존재유무를 감지하는 기법을 제시하며 기존의 방법과 효율성을 비교한다. 여기에서는 이산코사인변환 과정에서 64비트(8×8), 1블록 단위를 고려한 상태에서 이웃한 비트의 RGB 값 등을 구하는 과정을 JAVA 언어를 이용하여 구현하였다.

여기에서는 512×512(32.533Byte) 크기의 원본이미지에 68Byte 크기의 메시지를 삽입한 경우를 생각한다.

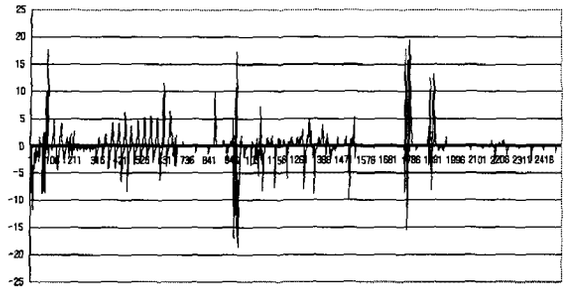


(그림 1) 원본이미지에 은닉자료가 포함되는 과정

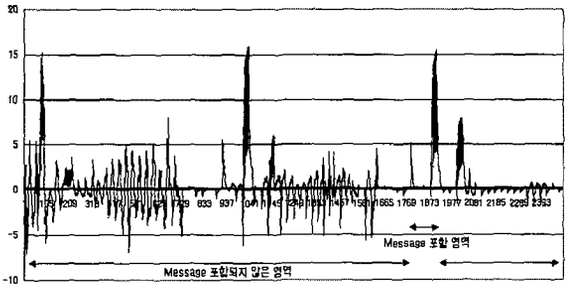
원본이미지에 비해 은닉자료가 0.20%인 경우 이웃한 RGB 값을 기반으로 한 카이스퀘어 통계량, χ^2=80.81 < $\chi^2_{\alpha=0.05}$=83.29로서 카이스퀘어 검정기법을 적용하면 은닉메시지가 존재한다는 것을 감지하지 못한다.

(5)식의 EWM 통계량을 이용하여 원본이미지의 이웃한 비트의 RGB 값의 차이를 <그림2>와 같이 나타낼 수 있다. (5)식의 EWM 통계량을 이용하여 원본이미지에 삽입메시지가 포함될 경우의 이웃한 RGB 값의 차이를 <그림3>과 같이 보일 수 있다.

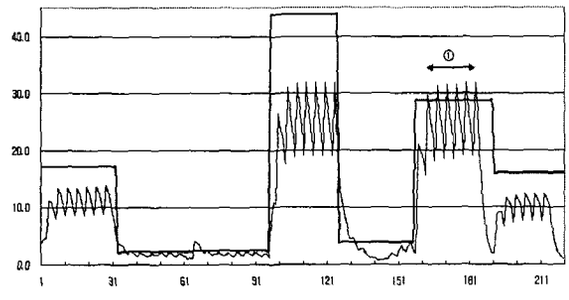
평활계수(λ)를 0.30으로 하고, 64 비트의 블록단위를 고려할 때 EWM 통계량과 기각역을 <그림4>와 같이 나타낼 수 있다. <그림2>와 비교하여 특정영역에서 RGB 값의 변화가 존재함을 확인할 수 있다. 즉, 그림의 ① 부분에서의 이상신호를 참고로 하여 은닉자료가 존재한다는 것을 확인할 수 있다.



(그림 2) 원본이미지의 이웃한 RGB값 차이 (EWM 통계량을 이용)



(그림 3) (원본이미지+삽입메시지)의 이웃한 RGB값 차이(EWM 통계량을 이용)

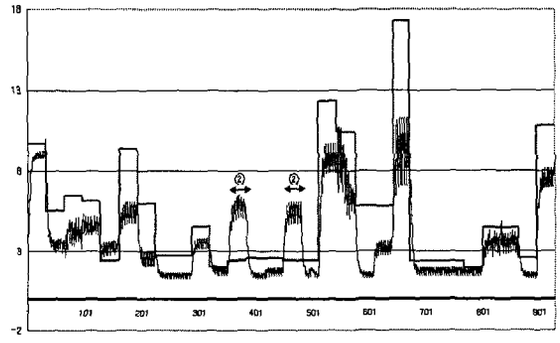


(그림 4) (원본이미지+삽입메시지), λ=0.3, 64 비트의 블록단위를 고려할 때 EWM 통계량과 기각역

이웃한 RGB 값을 기반으로 한 EWM 통계량을 이용하는 경우 평활계수를 0.10, 0.15, 0.20, 0.25, 0.30, 0.35 등으로 변경함에 따라 기각역을 벗어나는 상태를 조사한 결과 평활계수가 0.30 이 넘어설 때 은닉자료가 존재한다는 것을 감지할 수 있다.

λ	χ^2기법	EWM
0.10	×	×
0.15	×	×
0.20	×	×
0.25	×	×
0.30	×	○
0.35	×	○

참고로 600×450(16,304Byte) 크기의 원본이미지에 3,287Byte(20%) 크기의 이미지를 삽입한 경우에 EWM 통계량을 이용할 경우 은닉이미지의 유무를 <그림5>와 같이 ② 부분에서 쉽게 확인할 수 있다. χ^2=208.95 > $\chi^2_{\alpha=0.05}$로서 쉽게 은닉자료가 존재함을 확인할 수 있다.



(그림 5) (원본이미지+삽입이미지), λ=0.3, 64 비트의 블록단위를 고려할 때 EWM 통계량과 기각역

원본이미지에 비해 삽입자료가 매우 작은 경우에 기존의 χ^2 통계량을 이용한 검정법으로 은닉자료를 감지할 수 없다. 그러나 평활계수를 크게 하여 이웃한 RGB 값을 기반으로 한 EWM 통계량을 이용하면 은닉된 자료를 감지할 수 있으며 그 위치를 확인할 수 있다.

5. 결 론

유비쿼터스 환경에서 컴퓨터와 관련된 모든 작업 특히, 인터넷상에서 송수신되는 콘텐츠의 보호와 이용자 간에 송수신되는 메시지의 안정

성이 매우 중요한 역할을 하게 되었다. 정보보호를 위한 암호의 대중화는 예상외의 역기능 즉, 범죄에 이용되는 부작용을 낳게 되었다. 그러므로 정보보호와 더불어 은닉된 정보를 찾아내는 정밀한 감지기법 연구 또한, 중요한 요소가 되었다. 이웃한 비트의 RGB 값의 차이의 변화를 감지하는데 χ^2 통계량을 이용하는 것보다 지수가중이동 통계량 이용하는 것이 평활계수에 따라 유리할 수 있음을 확인하였다. 같은 조건에서 마코브체인[13]을 이용하여 혼합이미지로부터 은닉자료의 유무를 감지하는 기법은 향후 좀 더 연구가 진행되어야 한 부분이다.

참 고 문 헌

[1] R. Chandramouli(2003), "A Mathematical Framework for Active Steganalysis", Multimedia Systems 9, pp. 303 - 311.

[2] N. Cvejic and T. Sepp(2005), "Increasing Robustness of LSB Audio Steganography by Reduced Distortion LSB Coding", Journal of Universal Computer Science, Vol. 11, No. 1, pp. 56-65.

[3] Xuan-Hien Dang and K. C. S. Kota(2006), "Case Study : An Implementation of a Secure. Steganographic System", Security and Management '06 : Las Vegas, Nevada, USA, pp. 84-90.

[4] H. Farid(2002), "Detecting Hidden Messages Using Higher-Order Statistical Models", International Conference on Image Processing(ICIP), Rochester, NY, USA.

[5] H. Farid(2001), "Detecting Steganographic Messages in Digital Images", Technical Report TR2001-412, Dartmouth College.

[6] J. Fridrich, M. Goljan and D. Soukal(2003), "Higher-order statistical steganalysis of palette images", Proceedings of the SPIE, Vol. 5020, pp. 178-190.

[7] G. Jipper(2003), "Investigator's Guide to Steganography", CRC.

[8] I. S. Moskowitz, G. E. Longdon and L. Chang(2000), "A New Paradigm Hidden in Steganography", Proceedings, New Security Paradigms Workshop on Sept. 2000, Ballycotton, Co. Cork, Ireland.

[9] M. Naor and A. Shamir(1995), "Visual Cryptography", EUROCRYPT '94, A. De Santis Ed., Vol. 950 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, pp. 1-12.

[10] N. Provos and P. Honeyman(2003), "Hide and Seek:An Introduction to Steganography", IEEE Security&Privacy, pp. 32-44.

[11] N. Provos and P. Honeyman(2002), "Detecting steganographic content on the internet", ISOC NDSS'02, San Diego, CA.

[12] N. Provos(2001), "Defending Against Statistical Steganalysis", 10th USENIX Security Symposium. Washington, DC.

[13] K. Sullivan, U. Madhow, S. Chandrasekaran and B. S. Manjunath (2006), "Steganalysis for Markov Cover Data With Applications to Images", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, Vol. 1, No. 2, pp. 275-287.

[14] K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath and S. Chandrasekaran (2005), "Statistical Restoration for Robust and Secure Steganography", Proc. IEEE International Conference on Image Processing, Genova, Italy.

[15] N. Szabo(1999), "The Abuse of Statistics in Cryptography", Nick Szabo's Papers and Concise Tutorials.

[16] Syed Ali Khayam(2003), "The Discrete Cosine Transform(DCT): Theory and Application", Department of Electrical &Computer Engineering, Michigan State University.

[17] 김현곤, 원동호(2000), "지적재산권 보호를

위한 정보은닉 기술 및 표준화 연구", NCA IV-RER-00010, 한국전산원.
[18] 전국경제인연합회(FKI)(2005), "산업기술 유출방지 및 보호지원에 관한 법률(안)의 제정에 관한 의견", 정책건의자료.
[19] "The Discrete Cosine Transform(DCT)", [Online] Available http://www.cs.cf.ac.uk/Dave/Multimedia/node231.html
[20] Stephen Manley, "A Java implementation of the Discrete cosine Transformation", [Online] Available http://www.nyx.net/~smanley/dct/DCT.html
[21] "Steganography Software", [Online] Available http://home.earthlink.net/~emilbrandt/stego/softwarewindows.html

지 선 수(Seon-Su Ji)

- 정회원
- 1984년 충남대학교 계산통계학과(학사)
- 1986년 중앙대학교 응용통계학과(석사)
- 1993년 중앙대학교 응용통계학과(박사)
- 2006년 명지대학교 컴퓨터공학과(박사수료)
- 원주대학 컴퓨터정보관리과 교수
- (현) 강릉대학교 컴퓨터정보공학부 교수
- 관심분야 : 혼잡제어, 정보보안(암호학), 이미지 프로세싱