情 報 保 護 學 會 誌
第 12 券 第 1 號, 2002. 2

스테가노그라피의 이론적 배경과 검출기법

김 형 중*

Ⅰ. 서 론

메시지를 은밀하게 전송하는 방법은 여러 분야에서 수 천년동안 강구된 중요한 기술 가운데 하나였다. 영업비밀, 군사정보, 외교정책 등을 은밀하게 교환해야 하는 일은 이전에도 그랬던 것처럼 앞으로도 여전히 중요하다. 현대적인 정보통신 환경에서 디지털 멀티미디어와 인터넷의 보편화로 은밀하게 정보를 전송하는 방법에 대한 연구도 새로운 전기를 맞고 있다. 그래서 주목을 받게 된 것이 스테가노그라피 (Steganography) 및 스테거낼리시스 (Steganalysis) 기술이다.

스테가노그라피에 대해 특별히 더 많은 관심을 갖게 된 것은 2001년 9월 11일 미국무역센터에 대한 테러 때문이었다. 아날로그 스테가노그라피 연구의 역사는 수 천년이 넘었지만 디지털 스테가노그라피 역사는 그리 길지 않다. 그렇다고 지난 10여년간 디지털 스테가노그라피 기술 연구가 없었던 것은 아니다. 그런데도 오사마 빈 라덴이 이 기술을 써서 지령을 전파했을 것이라는 기사를 미국 언론 USA Today가 내보내면서 세인의 주목을 받게 되었다. 아무튼 이 일을 계기로 스테가노그라피에 대한 경계심이 한껏 고조되었다[1, 7, 9, 11]. 여기서 짚고 넘어가야 할 것은 이 보도 이후 테러나 마약거래 등 범죄에서도 스테가노그라피가 보편적으로 쓰일 가능성이 높아졌다는 점이다. 범죄에서도 이런 기술이 적용된다면 경찰이나 검찰의 정보수집능력은 심각한 타격을 입게 될 것이 자명하다.

인터넷 시대가 도래하면서 사이버 테러에 대한 대비책을 강화하는 것은 바람직한 일이다. 그런데 사이버 테러를 해킹이나 바이러스에 국한시켜서는 곤란하다. 해킹이나 바이러스에 대한 피해는 바로 드러나지만 스테가노그라피는 은밀성이 생명이기 때문에 이로 인한 파급효과는 영원히 묻힐 가능성이 많다. 그래서 이미 미국이나 유럽에서는 스테가노그라피 분야의 기술개발을 위해 많은 노력을 기울이고 있다.

전에 NSA에서 근무한 적이 있는 Bill Hancok은 자신이 스테가노그라피와 관련된 여섯 종류의 일에 관여하고 있다고 밝혔다. 그 가운데 하나는 프랑스 항공회사의 비행기 설계도 절취사건과 관련된 것이며 나머지는 비밀로 분류되어 밝힐 수 없다고 했다. 또 AFRL (Air Force Research Laboratory), 즉 미국 공군연구소와 함께 스테가노그라피 연구를 수행하고 있는 Wetstone Technologies는 인터넷 경매 사이트 e-Bay에서 이상한 현상을 발견했다고 보고했다. 경매중인 제품의 사진에서 며칠마다 픽셀의 값이 달라지는 것을 관찰했다는 것이다. 물론 그것으로부터 어떤 것도 입증하지는 못하지만 아무튼 이상한 일이라고 말했다.

스테가노그라피는 그 응용범위가 생각보다 넓다. 독일 기업 Demcon의 Steganos Security Suite라는 소프트웨어 패키지가 10만 카피 이상 팔렸음에 주목할 필요가 있다. 인터넷에는 무료 소프트웨어가 많이 올라와 있음에도 불구하고 상업용 패키지를 구매한다는 것은 그 용도가 무엇이었든 스테가노그라피가 급속히 확산되고 있음을 입증하는 사례라고 할 수 있다. 평범한 개인도 자신의 비밀 메시지를 은밀하게 보관하고 싶을 때 이 기술을 적용한다. 기업에서의 중요한 정보를 은밀히 보관하고 싶을 때 이런 기술을 적용할 수 있다. 그런데 이런 기술이 주로 이중으로 기재된 회계장부의 은닉 등 일종의 범죄행위에 이용될 수 있다. 한편에서는 검열을 피하기 위해 스테가노그라피를 이용하기도 한다. 보통 영상에 포르노그라피 영상을 숨겨 보내기도 한다. 따라서 스테가노그라피는 앞으로 매우 중요한 기술이자 심각한 문제를 일으킬 기술이 될 수도 있다.

스테가노그라피는 메시지를 은밀하게 전송하는 기

* 강원대학교 제어계측공학과 교수 (khj@kangwon.ac.kr)

술을 말한다. 스테가노그라피에서 가장 중요한 것은 메시지가 들어있음을 전혀 눈치채지 못하게 하는 것이다. 그래서 세인의 주목을 받지 않게 흔한 사진이나 오디오 클립에 메시지를 숨긴다. 스테가노그라피를 이용하는 당사자들은 제대로만 전송할 수 있다면 비밀 메시지 교환에는 어려움이 없다. 그렇지만 스테가노그라피에 대한 관심이 증폭되면서 비밀 메시지를 탐지하려 하거나 비밀 메시지 자체를 훼손시키려는 보이지 않는 전쟁이 치열하게 전개되고 있다.

그림 1이 스테가노그라피를 상징적으로 보여준다. 예를 들어 배우 최지우의 사진에 펜타곤의 사진을 숨겨 그림 1처럼 전송한다고 가정하자. 이런 사진 속에 비밀 메시지가 들어있을 것이라고 의심하는 사람은 거의 없다. 그런데 역으로 그런 점을 스테가노그라피 전문가들이 악용한다. 그리고 실제로 호스트 영상인 배우 사진에 메시지 영상인 펜타곤을 집어넣어 스테고 영상을 만드는 것은 그리 어렵지 않다. 문제는 숨기는 방법을 새로 고안해내면 알아채기가 쉽지 않다는 것이다.

〔그림 1〕 스테고 영상의 예 (배우 사진에 펜타곤을 숨김)

그래서 먼저 비밀 메시지를 멀티미디어 데이터에 숨기고 찾아내는 기술, 즉 은닉기술과 검출기술에는 어떤 것이 있는지 확인할 필요가 있다. 스테가노그라피는 주로 은밀하게 정보를 숨기는 데 사용되는 기술이다. 이에 비해 스테거낼리시스는 비밀 메시지 은닉 여부를 탐지하고, 그로부터 메시지를 추출하거나 메시지를 훼손시키는 것을 목표로 삼는다. 메시지를 숨길 때는 평문 그 자체를 숨길 수도 있고 평문을 비문으로 바꿔 숨길 수도 있다. 그래서 비밀 메시지를 찾아내기 위해서는 스테거낼리시스와 크립터낼리시스 기술을 함께 적용하는 것이 바람직하다.

이 논문은 먼저 스테가노그라피 기술에 대해 먼저 조망한다. 이어 스테거낼리시스 기술에 대해 살펴본다. 비밀 메시지를 숨길 멀티미디어 데이터에는 오디오, 영상, 비디오, 텍스트 등 다양하지만 여기서는 편의상 이미지에 대해서만 살펴보기로 한다. 그리고 영상의 LSB를 변경하는 스테고 기법에 대한 분석 방법에 국한해서 기술한다. 그렇지만 현재까지 제안된 방법들[3-6, 13-14]은 많은 가설에 근거하고 있고, 그 가설이 아직 확증된 것은 아니기 때문에 한계를 지니고 있다. 특히 스테고 전문가들은 일단 비밀이 노출된 방법을 다시 사용하는 일이 결코 없다는 점에 주목한다면 이 스테고 분석 기법은 다른 연구의 단초를 제공할 뿐이라는 사실을 확인하는데 그쳐야 한다.

2. 스테가노그라피 기술

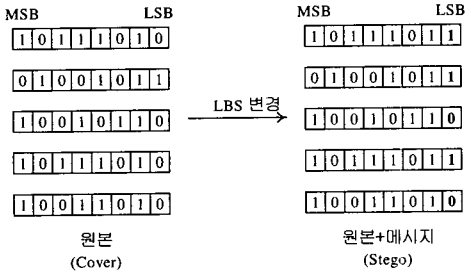
메시지를 은밀하게 영상에 숨기는 방법은 영상영역 숨기기와 변환영역 숨기기로 분류할 수 있다.

2.1. 영상영역 숨기기

초기에 사용된 방법으로 지금은 잘 쓰지 않는다. 이 방법은 영상영역에서 영상의 LBS (Least-Significant Bit) 값을 변환시키면서 정보를 숨긴다. LSB 그 자체가 의미하는 바와 같이 이 비트는 거의 무시해도 될만한 비트라고 생각하는 사람들이 많다. 물론 LSB 값을 바꿔도 영상에서는 거의 눈으로 구별할 수 없다. 일반 영상은 한 픽셀이 8비트로 구성되므로 픽셀의 LSB를 변경해서 그림 2와 같이 정보를 숨길 수 있다. 그림에서 원본의 LSB 값이 고쳐져 굵게 표시되어 있다. 굵게 표시된 것들이 숨겨진 메시다.

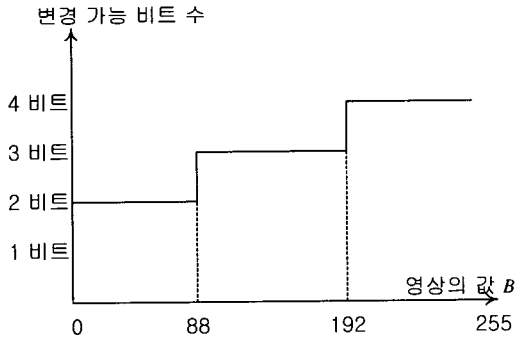
이 방법은 구현하기 쉽다는 장점을 지니고 있다. 그런데 이 방법은 동시에 여러 약점을 지니고 있다. 우선 용량의 한계를 들 수 있다. 8비트 영상일 경우 원본 영상 용량의 최대 1/8에 해당하는 메시지를 숨길 수 있다. 그 양이 적지는 않지만 다른 방법은 더 많은 정보를 숨길 수 있다. 그렇지만 더 고약한 약점은 이 방법이 공격에 매우 약하다는 것이다. 중간에서 스테고 영상을 가로채 LSB를 마음대로 고친 후 재전송해버리면 그런 사실을 모르는 당사자들은 영상으로부터 어떤 메시지도 복구할 수 없게 된다. 그저 아무 생각 없이 LSB 값을 고치는 것으로도 숨겨진 메시지를 완전히 날려버릴 수 있다는 것이 LSB의 변경에 의한 스테가노그라피의 가장 심각한 한계이다. 패킷 필터링 기술을 쓰면 의심스러운 영상을 잠시 붙들어둘 수 있고 그 사이 LSB를 변경

시켜 감쪽같이 재송할 수 있다. 이런 이유로 인해 LSB변경은 믿고 쓸 수 있는 기술이라 할 수 없다. 그래서 마찬가지 이유 때문에 워터마킹에서도 LSB 변경은 거의 쓰지 않는다.



〔그림 2〕 LSB 변경의 원리

물론 더 많은 양의 정보를 숨겨야 할 때도 있다. 그런데 정보를 더 많이 숨기면 영상의 질을 더 많이 훼손하게 된다. 따라서 영상의 질적 차이를 느끼지 못하게 하면서 정보를 숨길 수 있어야 한다. 이때 사용되는 유용한 도구가 JND (Just Noticeable Difference) 개념이다[2].



〔그림 3〕 JND에 의한 추가 정보 삽입 용량

JND란 영상에 정보를 삽입할 때 영상의 질적 차이가 눈에 띄게 달라지기 시작하는 한계 순간에서의 용량이라고 할 수 있다. 모든 픽셀의 LSB를 고치는 대신 JND를 이용할 경우 픽셀에 따라 한 비트 이상 여러 비트의 값을 변경할 수 있다. 영상의 값이 "200"일 때 1을 더해 "201"로 만드는 것과 영상의 값이 "2"인데 "1"을 더해 "3"으로 만드는 것은 다 "1"을 더하지만 영상에 미치는 영향의 서로 다르다. JND에서는 영상의 픽셀 값이 크면, 예를 들어 "200"의 경우 4비트까지 변화시켜도 영상의 질적 차이가 드러나지 않는다 (그림 3 참조). JND를 이용하면 LSB만 이용하는 것에 비해 숨길 수 있는 용량은 훨씬 증가하고 스테고 영상의 질도 높게 유지할 수 있다. 예를 들어 호스트 영상의 용량에 비해 30% 이상 45%까지 은닉영상을 숨겨도 PSNR은 33 dB 이상 40 dB까지 유지할 수 있다.

본 논문에서 다루는 스테가노그라피 기술은 JND를 사용하지 않고 단순히 LSB에만 메시지를 숨긴다고 가정하고 있다.

2.2. 변환영역 숨기기

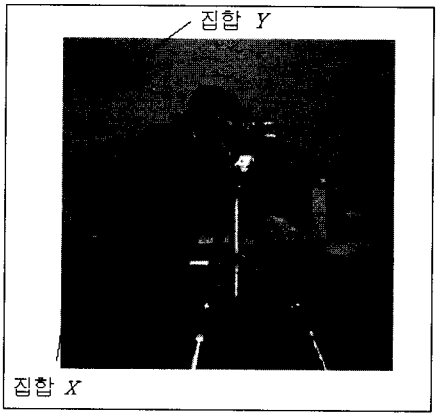
워터마킹에서와 마찬가지로 스테가노그라피에서도 영상을 일단 DFT, DCT, DWT 등 변환기술을 적용해서 영상영역에서 변환영역으로 옮기고 변환영역에서 메시지를 숨긴 다음 다시 영상영역으로 역변환하는 방법을 쓴다. 이 방법은 숨길 메시지 에너지를 영상영역에 고루 확산시키기 때문에 공격에 매우 강인하다는 장점이 있다. 그렇지만 변환 과정에서 계산 오차 및 양자화 등으로 인해 오류가 발생할 수 있어 메시지를 정확하게 찾을 수 있는 확률은 LSB 변형에 비해 낮아지게 된다. 변환영역 숨기기에서는 필연적으로 타입 I (False Positive) 에러 또는 타입 II (False Negative) 에러가 발생한다. 그러므로 이런 에러를 줄이기 위한 방안이 반드시 마련되어야 한다. 그래서 에러정정코드를 함께 사용하기도 한다.

그런데 변환영역에서 비밀 메시지를 숨길 경우에는 원본 없이 숨긴 메시지를 찾아야 하므로 반드시 블라인드 스테가노그라피 또는 블라인드 워터마킹 기술을 적용해야 한다[2, 15]. 영상에서 적용할 수 있는 블라인드 스테가노그라피 기술에는 확산대역 기법, 패치워크 기법, 특징벡터 기법 등이 있다. 물론 이들 기법은 영상영역이나 변환영역 어디에서도 적용할 수 있다. 그렇지만 강인성을 확보하기 위해 주로 변환영역에서 적용한다.

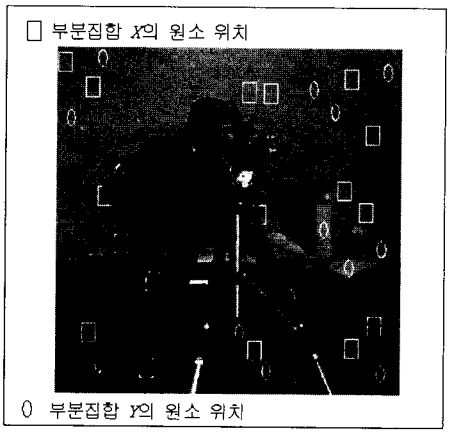
확산대역 기법에서는 영상에 주로 PN 시퀀스를 숨긴다. 오디오에서는 PN 시퀀스가 잡음을 유발하며, PN 시퀀스 값의 진폭을 매우 작게 할지라도 잡음이 귀에 분명히 들리는 단점이 있다. 그래서 대부분 심리음향모델과 연계시켜 잡음이 들리지 않게 한다. 그렇지만 영상에서는 PN 시퀀스를 숨겨도 눈에 잘 드러나지 않는다. 귀가 눈에 비해 매우 섬세한 기관이라는 것을 알 수 있다. 그래서 PN 시퀀스를 이용한 확산대역 기법은 주로 영상에서 많이 사용한다. 패치워크 방법은 통계적 특성을 이용해서 정보를 찾

는 방법이다. 패치워크 방법은 부분집합 X와 부분집합 Y를 선택해서 둘을 변조하는 방식을 택한다.

공격을 받지 않아도 블라인드 스테가노그라피는 숨긴 정보를 찾아낼 때 오류를 수반할 수 있다. 숨길 영상의 특성에 따라 오류의 정도가 달라진다. 저주파 성분이 많고 고주파 성분이 적은 영상은 정보를 검출할 때 오류를 적게 발생시키는 경향이 있다. 그렇지만 영상의 변화가 적기 때문에 조금만 값을 변화시켜도 쉽게 그 차이를 알 수 있는 단점을 지니고 있다.



〔그림 4〕 패치워크의 개념도. 두 부분집합 X와 Y의 선정 방법이 매우 중요하다.

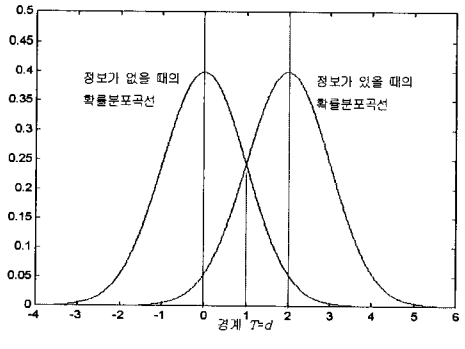


〔그림 5〕 패치워크를 위해 무작위로 두 부분집합의 위치를 선정하는 방법

그림 4의 왼편 상단에 두 개의 상자가 보인다. 왼편 상자의 값은 주변의 픽셀보다 훨씬 어둡고, 그 옆의 상자는 약간 밝게 일부러 과장해서 표시했다. 부분집합 X를 하나의 패치라고 부르며, 부분집합 Y도 하나의 패치라고 부른다. 패치 X는 원래의 픽셀 값에서 d를 빼고, 패치 Y는 반대로 원래 픽셀 값에 d를 더한다. 정보를 삽입하는 것은 이와 같이 매우 단순하다. 그렇지만 샘플의 평균과 분산을 구해야 하고, 정보가 숨겨진 위치를 정확히 알아야 하기 때문에 검출은 정교하고 복잡해진다. 즉 동기화(Synchronization) 문제가 발생한다. 모든 방법이 다 동기화 문제를 안고 있으나 특별히 패치워크가 동기화에 민감하다.

실제로는 그림 4와 같이 집합을 인접한 점들로 구성하지 않는다. 인접한 점들로 부분집합 X를 구성하면 집합 X는 서로 비슷한 통계적 특성을 갖는다. 역시 인접한 점들로 구성된 부분집합 Y의 점들도 비슷한 특성을 지닌다. 그렇지만 부분집합 X와 Y사이의 통계적 특성은 무척 다르게 된다. 그런데 패치워크는 두 부분집합의 통계적 특성이 매우 비슷해야 비로소 제 기능을 발휘할 수 있다.

그래서 패치를 만들 때는 통계적으로 샘플을 무작위 추출하게 된다. 그림 5가 무작위로 샘플을 추출해서 만든 부분집합 X와 부분집합 Y의 원소 위치를 보여주고 있다. 이렇게 샘플을 선정하면 부분집합 안의 원소들은 그 값에서 큰 차이가 있을 수 있지만, 두 부분집합 사이의 통계적 특성은 비슷해진다. 따라서 무작위 샘플 추출이 매우 중요하다.



〔그림 6〕 패치워크의 통계적 특성. $d=1$인 경우.

강제적으로 두 집합의 원소의 값에서 임의의 상수 d을 더하거나 뺀다면, 즉, $X'=X+d$와 $Y'=Y-d$로 만든다면 두 표본평균의 차의 기대값은

$$E\{\bar{X}'-\bar{Y}'\}=E\{(\bar{X}+d)-(\bar{Y}-d)\}=E\{\bar{X}-\bar{Y}\}+2d=2d$$

과 같이 된다. 그림 6은 $d=1$일 때 $\overline{X}-\overline{Y}$의 분포를 보여준다. 여기서, $E\{\overline{X}-\overline{Y}\}=0$인 것을 근거로 $E\{\overline{X'}-\overline{Y'}\}=2d$ 임을 이용해서 d를 경계 (Threshold) T로 삼아 워터마크 유무를 판별한다.

3. 스테거낼리시스 기술

스테가노그라피의 목표가 메시지 은닉 여부 자체를 철저히 숨기는 것이기 때문에 스테거낼리시스의 출발은 비밀 메시지 은닉 여부를 알아내는 일로부터 시작된다. 사실 이 과정이 가장 어렵다고 해도 과언이 아니다. 비밀 메시지가 숨겨져 있다면 그것을 찾아내 해독하거나 비밀 메시지 자체를 제거해버리는 일이 그 다음 순서이다. 그런데 비밀 메시지를 제거하는 일은 비교적 쉽다. 그러나 비밀 메시지 해독은 매우 어려운 일이다. 특히 숨긴 메시지가 평문이 아닌 비문이라면 이야기는 더욱 복잡해진다. 즉 스테고 분석은 다음과 같이 이루어진다.

1단계: 스테고 여부 탐지 → 아래의 2나 3단계를 시행하기 위한 전처리 단계
2단계: 스테고 메시지 파괴 → 3단계를 실시하기는 어렵지만 비밀통신의 무력화 가능
3단계: 스테고 메시지 검출 및 해독 → 고난도 기술을 요하는 최종 목표

스테가노그라피 기술을 적용했음을 알아내는 1단계 기술이 감지 (detection) 기술이다. 감지기술은 다시 서명 (signature) 감지와 맹목 (blind) 감지로 분류된다.

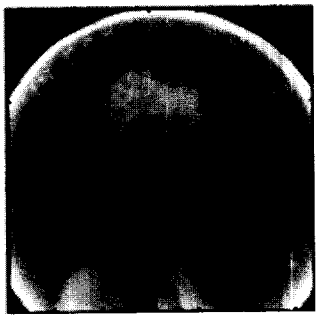
3.1. 서명 감지 기술

서명 감지는 기존의 스테가노그라피 도구를 이용할 때 자연스럽고 깨끗한 영상과 달리 특이한 현상이 나타나는지를 알아내는 기술이다. 이 방법에서는 원본과 사본을 모두 보유하고 있기 때문에 비밀 메시지를 넣기 전과 넣은 후의 차이를 분석하는 것이므로 기술 자체가 어렵지는 않다. 다만 그런 차이를 논리적으로 설명할 수 있어야 한다는 것이 과제이다. 그리고 둘 사이의 차이를 아는 것은 다음과 같이 두 가지 중요성을 지닌다.

1) 새로운 스테가노그라피 기술을 개발할 때, 그런 차이를 드러내지 않도록 설계하는데 필요한 자료를 제공한다.
2) 기존의 스테가노그라피 도구를 썼는지 알 수 있고, 그랬다면 은닉 여부를 쉽게 판별할 수 있다.

그렇지만 원본 없이도 메시지가 숨겨져 있는지 판별하는 방법이 더 바람직하다. 뒤에서 설명할 맹목 감지방법이 바로 그것이다.

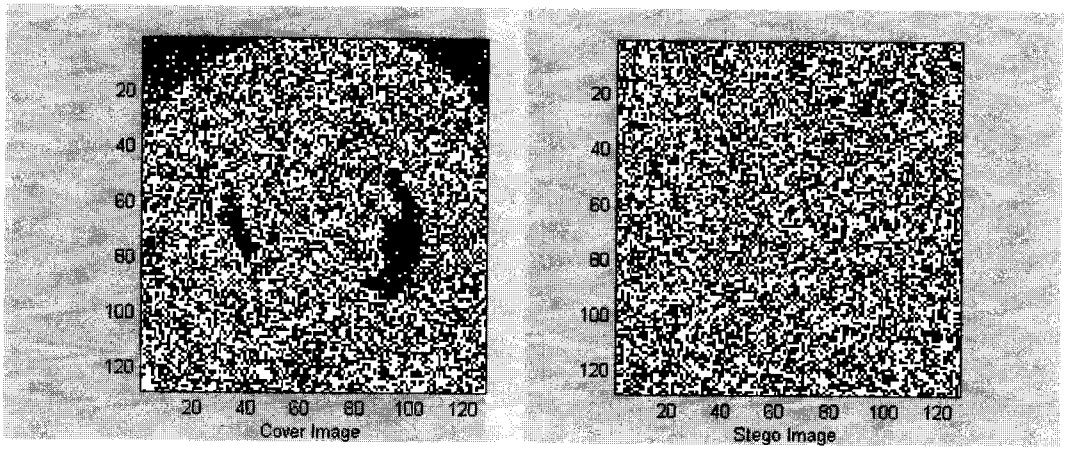
아무튼 기존의 도구들은 스테가노그라피 기술을 적용하기 전과 후에 두드러진 차이를 보인다. 예를 들면 S-Tools 같은 도구는 원본 영상의 컬러 수를 줄이면서 몇몇 컬러 팔레트로 확장시킨다. 그 팔레트를 휘도에 따라 정렬시키면 한 무리의 색깔들이 동일하게 나타나지만 실제로는 한 비트의 분산 차이를 보인다. 이런 분산 형태는 변형되지 않은 자연적인 영상에서는 매우 드물게 나타나는 현상이다. 그래서 영상에서 이런 패턴이 관측된다면 그것은 비밀 메시지를 담고 있을 가능성이 매우 높다는 것을 의미한다.



〔그림 7〕 동전 영상

예를 들어 그림 7의 동전영상의 LSB를 변경해서 메시지를 숨긴 경우에 대해 살펴보자. 그림 8의 왼편은 그림 7의 LSB를 보여주고 있다. 자연스런 영상에서는 LSB도 원본의 윤곽을 지니고 있다는 것에 주목해야 한다. 이에 비해 그림 8의 오른편은 LSB에 메시지를 넣은 스테고 영상의 LSB를 보여주고 있다. 그런데 오른편 그림은 원본 영상의 윤곽을 전혀 포함하고 있지 않음에 주목하라.

그렇지만 그림 7과 같이 윤곽이 명확한 영상을 커버 영상으로 쓰는 것은 매우 어리석은 일이다. 실제로 잡음이 많이 포함된 영상이나 텍스처 영상은 그림 8의 오른편과 같이 보이기 때문에 메시지를 숨



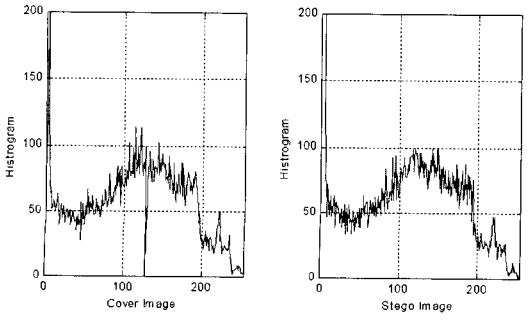
〔그림 8〕 LSB의 패턴. 왼편은 자연스러운 영상이고 오른편은 스테고 영상

기기에 적합하다.

스테가노그라피기는 철저히 "은밀함 속의 보안(Security through Obscurity)" 원칙에 의거한다. 그러므로 스테가노그라피 특성이 노출된 방법을 다시 쓰는 일은 거의 없다고 보아야 한다. 그래서 스테가노그라피에서는 쫓고 쫓기는 연구가 연속적으로 이루어질 수 밖에 없다.

3.2. 맹목 감지 기술

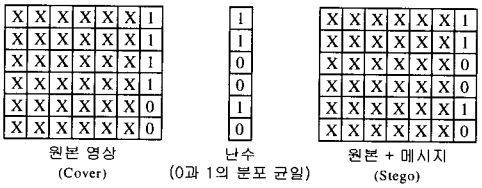
맹목 감지는 원본 영상이 없이 비밀 메시지 유무를 감지해야 하기 때문에 훨씬 고난도 기술이라고 할 수 있다. 이 연구의 핵심은 비밀 메시지가 포함되지 않은 자연스런 영상에서만 나타나는 특징 및 비밀 메시지가 포함된 영상에서만 나타나는 고유한 특징을 알아내는 일에 있다. 한편 통계적 특성을 이용해서 비밀 메시지 여부를 판별하는 방법도 쓰이고 있다.



〔그림 9〕 커버 영상과 스테고 영상의 히스토그램

3.2.1. 히스토그램 분석법

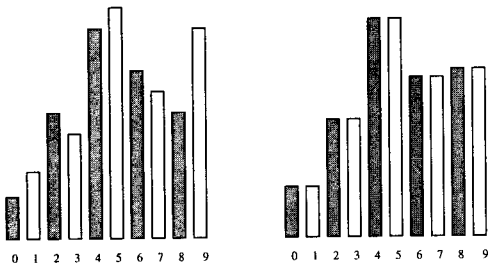
자연스러운 영상에서는 픽셀의 히스토그램이 균일하지 않지만 스테고 영상에서는 히스토그램이 훨씬 균일한 분포를 갖는다는 것이 통계적 방법에서의 기본 가설이다. 과연 그것이 사실인가? 이 점에 대해서는 명확한 해답이 아직은 없지만 실험 결과는 가설에 근접한다. 그림 7의 동전 영상에 대한 히스토그램이 그림 9와 같다. 그림 9의 왼편이 원본 영상에 대한 히스토그램이고, 오른편이 스테고 이미지에 대한 히스토그램이다. 물론 두 히스토그램 다 치우침 현상이 매우 큰 편이지만 그래도 분명히 오른편이 왼편에 비해 분포가 다소 더 균일하다고 할 수 있다. 그림 7의 영상에 매 픽셀마다 2비트의 난수를 더했더니 그림과 같이 히스토그램의 분포를 약간이나마 더 평탄하게 만든 것이다.



〔그림 10〕 스테고 영상의 분포가 보다 균일하게 된다는 가설의 근거

그렇다면 이런 현상이 나타나는 이유는 무엇인가? 그림 10은 상대적인 픽셀의 빈도수를 보여주고 있다. 왼편은 원본 픽셀 값의 분포를 보여주는 것으로 LSB 값을 볼 때 '0'이 '1'에 비해 1/3에 불과하다. 그런데 삽입할 메시지에서는 '0'과 '1'의 비율이

거의 같다. (만일 '0'과 '1'의 비율에 큰 차이가 나면 암호학적으로 바람직하지 못하다고 결론을 내릴 수 있다.) 따라서 삽입할 '0'과 '1'은 원본에서의 '0'과 '1'의 비율에 맞게 삽입될 것이므로 결국은 삽입 이후의 스테고 영상에서의 LSB 값은 오른편 그림에서와 같이 '0'과 '1'의 빈도가 비슷해진다는 것이 이 가설의 근거이다.



〔그림 11〕 메시지 은닉 전과 후의 히스토그램 추정 모양

그림 9는 픽셀의 값을 0부터 255까지의 8비트-레벨에 대한 결과이기 때문에 히스토그램이 완전히 편평해지기를 기대할 수는 없지만 그림 9가 어느 정도 수긍할 수 있는 개연성은 제공하고 있음을 알 수 있다. 그림 9가 부분적으로 사실이라면 여기서 "인접하는 색들은 서로 비슷해진다"는 가설을 만들 수 있다.

그림 11은 이런 가설을 근거로 만들어진 그림이다. 그림 11의 왼편이 스테고 영상의 히스토그램이다. 스테고 분석에서 얻을 수 있는 자료는 이것이 유일하다. 그런데 정보가 숨겨져 있는지의 여부를 확인하려면 원본이 있어야 가장 확실하다. 그런데 원본은 없다. 그렇다면 생각을 바꾸어 왼편이 원본이라고 보고 오른편이 스테고 영상이라고 가정한다. 둘 사이의 차이가 존재하는지를 봄으로써 정보의 은닉 여부를 알아내는 방법을 Westfeld와 Pfitzmann이 제시했다[14].

정보를 숨기기 전과 숨긴 후의 i번째의 히스토그램 값을 각각 n_i와 n_i^*라고 하자. 정보를 숨기기 전의 히스토그램 통계량은

$$y_i = n_{2i}$$

로 놓고, 삽입 후의 히스토그램 통계량은

$$y_i^* = \frac{n_{2i} + n_{2i+1}}{2}$$

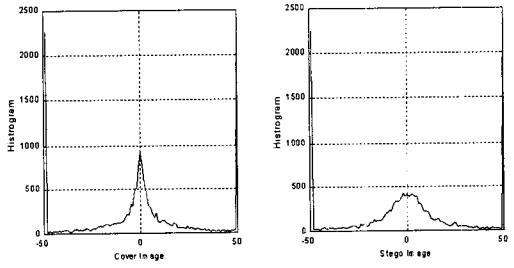
로 잡자. 이 두 통계량 사이에 차이가 있는지 알아보기 위해 테스트를 적용한다. 검정 통계량은 아래 식

$$c = x_{k-1}^2 = \sum_{i=0}^{k} \frac{(y_i - y_i^*)^2}{y_i^*} \quad (k-1)\text{자유도}$$

와 같이 주어진다. 이때 y_i^*와 y_i의 분포가 같게 될 확률은 아래 식

$$p = 1 - \frac{1}{2_z \Gamma(z)} \int_0^c e^{-x/2} x^{z-1} dx, \; z = \frac{k-1}{2}$$

에서 p로 주어진다.



〔그림 12〕 동전 영상에 DFT를 적용하고 난 후 구한 히스토그램

만일 메시지가 연속으로 삽입된 경우에는 p값이 1에 근접하다가 메시지 끝 부분에 이르러서는 p가 갑자기 0에 가깝게 떨어지는 것을 관찰할 수 있다. 그러므로 이 방법을 쓰면 메시지 은닉 여부도 알 수 있을 뿐 아니라 메시지 차이도 알 수 있게 된다. 그렇지만 메시지를 랜덤하게 숨기면 이 방법의 효용이 크게 감소한다.

한편 Provos는 DCT 변환을 한 후 두 영상의 차이에 x^2 테스트를 적용했다. 이 논문에서는 DFT를 취하고 난 후 실수부의 히스토그램을 구했다[13]. 그림 7의 동전 영상에 대해 8×8 DFT를 적용하고 히스토그램을 구하면 그림 12와 같은 모양을 얻는다. 왼편이 메시지를 삽입하기 전의 것이고 오른편이 메시지 삽입 후의 모양이다. 분명한 차이가 발생하는 것을 확인할 수 있다.

3.2.2. 이중통계량 방법

이 방법은 Fridrich, Goljan, Du에 의해 개발되었고[5] Fridrich, Goljan에 의해 일반화되었다[6]. 이들 개념은 상당히 신선하지만 너무 많은 가정들로 인해 아직은 널리 쓰이고 있지 않다. 그렇지만 이론적으로는 흥미있는 접근이라고 할 수 있다.

스테고 영상에 $M\times N$ 픽셀이 있고 각 픽셀은 8비트의 값으로 구성되어 있다고 가정하자. 스테고 분석을 위해 판별함수 (discrimination function) f를 이용해서 공간상관관계를 도출한다. 판별함수 f는

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=0}^{n-1} |x_{i+1} - x_i|$$

로 정의한다. 이 함수는 픽셀 그룹 $G(x_1, x_2, \ldots, x_n)$의 매끄러움을 나타내는 측도로 사용된다. 영상에 잡음이 많이 포함될수록 f의 값은 더 커진다.

LSB 삽입은 영상이 잡음처럼 보이게 만드는 특성이 있다. 따라서 LSB 삽입 후에는 f의 증가가 예상된다. 그런데 LSB 삽입은 다음과 같은 플리핑 함수 (flipping function) F를 이용해서 표현할 수 있다. 플리핑 함수 F_1은

$$F_1: 0\leftrightarrow 1,\ 2\leftrightarrow 3,\ 4\leftrightarrow 5,\ \ldots,\ 252\leftrightarrow 253,\ 254\leftrightarrow 255$$

로 정의된다. 플리핑 함수 F_1을 적용하면 각 픽셀 값의 LSB를 변경하는 것과 같은 효과가 있다. 예를 들어 픽셀 값 15인 '00001111'는 LSB를 플리핑시켜 '00001110'으로, 즉 14로 변경시키는 것과 같다. 이에 비해 플리핑 함수 F_2는

$$F_2: 0\leftrightarrow 2,\ 1\leftrightarrow 3,\ 4\leftrightarrow 6,\ \ldots,\ 252\leftrightarrow 254,\ 253\leftrightarrow 255$$

로 바꾸는 것을 의미한다. 플리핑 함수 F_{-1}는

$$F_{-1}: -1\leftrightarrow 0,\ 1\leftrightarrow 2,\ 3\leftrightarrow 4,\ \ldots,\ 253\leftrightarrow 254,\ 255\leftrightarrow 256$$

이 된다. 플리핑 함수 F_1는 모든 x에 대해 $F_{-1}(x) = F_1(x+1) - 1$가 성립한다. 한편 F_0 함수는 플리핑을 하지 않고 입력을 그대로 출력한다. 즉 $F_0(x) = x$가 된다. 그레이 레벨 값 x에 F을 적용하는 것은 다시 역변환이 가능한 치환 (permutation) 연산이므로 원상 회복이 가능하다.

영상 픽셀은 연속하는 n개를 묶어 픽셀 그룹 $G(x_1, x_2, \ldots, x_n)$을 형성한다. 이 영상 그룹은 다시 세 종류의 그룹으로 분류하는데 정규그룹 (regular group) R, 특이그룹 (singular group) S, 무용그룹 (unusable group) U가 그것들이다. 이들 분류의 기준은 다음과 같다.

$$\text{정규그룹: } G\in R \leftrightarrow f(F(G)) > f(G)$$
$$\text{특이그룹: } G\in S \leftrightarrow f(F(G)) > f(G)$$
$$\text{무용그룹: } G\in U \leftrightarrow f(F(G)) > f(G)$$

단, $F(G) = (F(x_1), F(x_2), \ldots, F(x_n))$을 의미한다.

예 1: 다음 영상 그룹 $G(x_1, x_2, \ldots, x_6) = [120, 121, 200, 230, 163, 170]$의 그룹을 판정하라.

해: 플리핑 함수 F_1을 적용한다면

$$F_1(G(x_1, x_2, \ldots, x_6)) = [121, 120, 201, 231, 162, 171]$$

이 된다. 여기에 판별함수 f를 적용하면

$$f(F_1(G(x_1, x_2, \ldots, x_6)) = |120-121| + |201-120| + |231-201| + |162-231| + |171-162| = 190$$

이다. 이에 반해

$$f(G(x_1, x_2, \ldots, x_6)) = |121-120| + |200-121| + |230-200| + |163-230| + |170-163| = 184$$

이다. 결국 $f(F(G)) > f(G)$이므로 이 그룹은 정규그룹이다.

마스크 M은 n개의 원소로 구성되며, 그 값은 -1, 0, +1 가운데 하나의 값을 갖는다. 플리핑 그룹 $F(G)$는

$$F(G) = (F_{M(1)}(x_1), F_{M(2)}(x_2), \ldots, F_{M(n)}(x_n))$$

으로 정의한다. 플리핑의 목적은 픽셀 값을 약간씩 변화시켜 잡음을 삽입하는 것과 같은 효과를 얻는데 있다. 플리핑은 대부분 판별함수의 값을 증가시킨다. 물론 판별함수의 값이 줄어드는 경우도 있다. 그렇지만 대부분 플리핑은 정규그룹의 수가 특이그룹의 수보다 많게 만든다. Fridrich, Goljan, Du는 이런 방법으로 정보를 은닉하는 기술을 제안했다[5].

마스크 M을 지닌 플리핑으로 골라진 정규그룹의 상대적 수를 R_M이라고 하자. 마찬가지로 특이그룹의 수를 S_M이라고 한다. 그런데 여기서 $R_M + S_M \leq 1$이고 음의 마스크에 대해서도 $R_{-M} + S_{-M} \leq 1$이 성립한다고 가정한다[6]. RS 기법에 의한 통계학적 가설검증은 특정한 영상에 대해 R_M에 대한 기대치가 R_{-M}의 기대치와 같고, 마찬가지로 S_M에 대한 기대치도 S_{-M}의 기대치와 같다는 것이다. 즉,

$$R_M \cong R_{-M},\ S_M \cong S_{-M}$$

이다.

R_M는 M에 대한 그룹 G의 매끄러움을 나타내는 지표이다. 한편 R_{-M}은 픽셀의 색을 하나씩 이동시킨 후 매끄러움을 판별하는 것이므로 여전히 비슷한 매끄러움을 지닐 것으로 예상할 수 있다. 그런 의미에서 $R_M \cong R_{-M}$가 될 것이라고 예측하는 데는 무리가 없다. 마찬가지로 $S_M \cong S_{-M}$가 성립할 것이라고 예상할 수 있다.

예 2: 영상 픽셀의 값 〔178 187 179 180 177 176 175 200 202 201 204 205 206 200 201 198〕에 대해 일 때 R_M, R_{-M}, S_M, S_{-M}을 계산하라.

해: 먼저 영상을 네 픽셀씩 4개의 그룹으로 나눈다. 각각의 픽셀 그룹은

$$G_1 = [178\ 187\ 179\ 180]$$
$$G_2 = [177\ 176\ 175\ 200]$$
$$G_3 = [202\ 201\ 204\ 205]$$
$$G_4 = [206\ 200\ 201\ 198]$$

가 된다. 먼저 G_1에 대해 마스크 M으로 플리핑 함수를 적용하면

$$\begin{aligned}(F_{M(1)}(x_1), F_{M(2)}(x_2), F_{M(3)}(x_3), F_{M(4)}(x_4)) &= (F_0(178), F_1(187), F_1(179), F_0(180)) \\ &= [178, 186, 178, 180]\end{aligned}$$

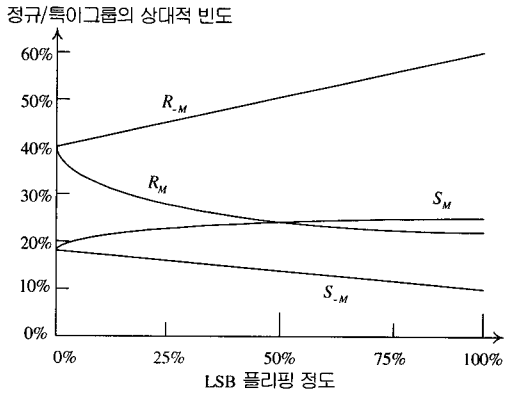
이 된다. 따라서 $f(F(G_1)) = 18$이고 $f(G_1) = 18$이므로 이 그룹은 무용그룹이다. 마찬가지로 G_2, G_3, G_4에 대해서도 마찬가지 연산을 하면 $f(F(G_2)) = 29$, $f(G_2) = 27$, $f(F(G_3)) = 7$, $f(G_3) = 5$, $f(F(G_4)) = 8$, $f(G_4) = 10$이므로 $R_M = 2/4$, $S_M = 1/4$이 됨을 알 수 있다.

이번에는 $-M$을 G_1에 적용해보자.

$$\begin{aligned}(F_{-M(1)}(x_1), F_{-M(2)}(x_2), F_{-M(3)}(x_3), F_{-M(4)}(x_4)) &= (F_0(178), F_{-1}(187), F_{-1}(179), F_0(180)) \\ &= [178, 188, 180, 180]\end{aligned}$$

이므로 $f(F(G_1)) = 18$, $f(G_1)) = 18$,로 역시 무용그룹이다. 마찬가지로 G_2, G_3, G_4에 대해서도 같은 연산을 하면 $f(F(G_2)) = 27$, $f(G_2) = 27$, $f(F(G_3)) = 3$, $f(F(G_3)) = 5$, $f(F(G_4)) = 14$, $f(G_4) = 10$이므로 $R_M = 1/4$, $S_{-M} = 1/4$이 됨을 알 수 있다.

그림 13은 자연적인 영상에서 잡음을 100% 삽입할 때 까지 예상되는 R/S 그룹의 수를 보여준다. 잡음이 포함되지 않은 자연적인 영상에서는 $R_M \cong R_{-M}$, $S_M \cong S_{-M}$가 예상되는 것이 이상적이다. 그리고 이때는 $R_M (\cong R_{-M}) \geq S_M (\cong S_{-M})$이 예상된다. 그런데 LSB에 잡음을 삽입하게 되면, 그리고 삽입하는 메시지가 많아질수록 R_M과 S_M의 간격은 좁혀진다. 그림 13에서 보면 x축이 0%일 때는 이상적으로 $R_M \cong R_{-M}$, $S_M \cong S_{-M}$이고 x축이 100%일 때는 R_M과 S_M의 간격이 좁혀진 것을 알 수 있다. 50%일 때는 $R_M \cong S_M$이 되는 것을 확인할 수 있다. 놀라운 것은 LSB의 값을 뒤흔들게 되면 R_{-M}과 S_{-M}의 간격은 더 벌어진다는 점이다.



〔그림 13〕 예상되는 이상적인 RS-도표

R_{-M}과 S_{-M}의 간격이 벌어지는 것을 $M=[0\,1\,0]$인 경우를 예로 들어 보이기로 한다. 집합 $C_i=\{2i, 2i+1\}$, $i=0, 1, 2, \ldots, 127$을 정의하고, $C_{rst}=\{G \mid G\in C_r\times C_s\times C_t\}$을 클릭 (clique), 즉 그룹의 집합으로 정의한다. 그렇다면 클릭의 총 수는 128^3가지가 존재한다. 각 클릭은 3개의 원소로 구성된 8개의 집합으로 구성된다. 예를 들어 $r=s=t=1$ 인 경우를 살펴보자. 이때의 클릭은

$$C_{1,1,1}=\{G \mid (2,2,2)\ (2,2,3)\ (2,3,2)\ (2,3,3)\ (3,2,2)\ (3,2,3)\ (3,3,2)\ (3,3,3)\}$$

와 같이 생성된다. 이것에 대해 $M=[0\,1\,0]$으로 마스킹하고 그룹을 판별하면 $R=2$, $S=2$, $U=4$가 얻어진다. 그렇지만 같은 마스크에 대해 F_1을 적용하면 $U=8$이 얻어진다. 이번에는 $r=s=1$, $t=2$이라면

$$C_{1,1,2}=\{G \mid (2,2,4)\ (2,2,5)\ (2,3,4)\ (2,3,5)\ (3,2,4)\ (3,2,5)\ (3,3,4)\ (3,3,5)\}$$

이고 이에 대해 F_1을 적용하면 역시 $R=2/8$, $S=2/8$, $U=4/8$가 얻어지지만 F_{-1}을 적용하면 $R=4/8$, $U=4/8$가 얻어진다. 따라서 이것을 정리하면 표 1이 된다. 단, 수평적으로나 수직적으로 대칭인 클릭만 모아 4개의 클릭으로 정리했다.

LSB 변경에 대한 연구의 근간은 LSB 변경의 결과로 F_1에 대해서는 언제나 $R_M\cong S_M$이 됨을 표 1에서 알 수 있다. 그렇다면 잡음이 많이 첨가될수록 R_{-M}이 증가한다는 것은 그림 13과 표 1에서 확인할 수 있다. 그리고 잡음 삽입이 증가할수록 $R_M\cong S_M$이 된다는 것도 그림 13은 물론 표 1에서도 확인이 가능하다. 아울러 잡음 증가가 R_{-M}를 증가시키면서 S_{-M}을 감소시킨다는 것도 표 1에서 확인할 수 있다.

〔표 1〕 클릭의 네 종류와 플리핑 결과 단, $M=[0\,1\,0]$

클릭의 종류	F_1			F_{-1}		
	R	S	U	R	S	U
$r=s=t$	2/8	2/8	4/8	8/8	0	0
$r=s>t$	2/8	2/8	4/8	4/8	0	4/8
$r<s>t$	4/8	4/8	0	4/8	4/8	0
$r>s>t$	0	0	8/8	0	0	8

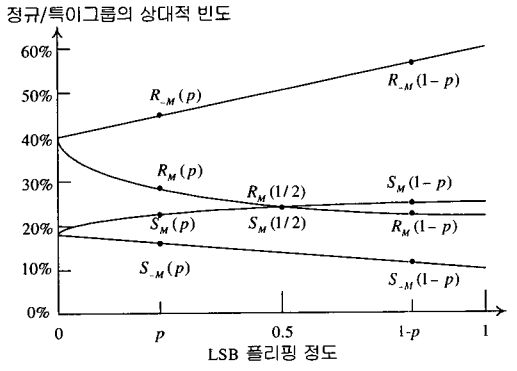
Fridrich 등의 RS 스테거낼리시스는 RS 도표의 네 곡선을 예측하는 것으로 시작된다[6]. 그들은 실험을 통해 R_{-M}과 S_{-M}의 곡선은 직선으로 모델링이 가능하고, R_M과 S_M은 2차곡선 식으로 모델링이 가능하다고 주장한다[6]. 많은 실험을 통해 모델링하면 그림 13과 같은 그림이 얻어질 수 있지만 실제 영상에서는 13과 비슷한 영상을 얻을 수도 있고, 전혀 다른 것을 얻을 수도 있다.

만일 길이는 p이지만 그 길이는 알 수 없는 메시지를 포함하는 스테고 영상이 있다고 가정하자. 메시지는 영상의 랜덤한 위치의 LSB에 숨겨졌다고 가정한다. 이제 p의 크기를 예측하는 방법에 대해 살펴보기로 하자.

우선 실험을 통해 우리는 $R_M(p)$, $S_M(p)$, $R_{-M}(p)$, $S_{-M}(p)$,을 얻게 된다 (그림 14 참조). 이제 이 스테고 영상 LSB 전체에 잡음을 삽입한다. 이미 p에 해당하는 부분에는 잡음이 삽입되었기 때문에 전체에 잡음을 섞는다는 것은 $1-p$에 해당하는 부분에 잡음을 섞는 것으로 재해석할 수 있다. 따라서 $R_M(1-p)$, $S_M(1-p)$, $R_{-M}(1-p)$, $S_{-M}(1-p)$, 을 실험을 통해 얻는다. R_{-M}과 S_{-M}은 1차식으로

모델링이 가능하다고 했으므로 각각 두개의 점이 있으면 이것을 이용해 모델링을 마친다. 따라서 각각 $R_{-M}(p)$과 $R_{-M}(1-p)$을 이용해서 R_{-M}을 모델링한다. 그런데 R_M과 S_M은 2차식으로 모델링이 가능하다고 했기 때문에 각각 3개의 점이 필요하다.

따라서 추가로 $R_M(1/2)$과 $S_M(1/2)$의 값을 구한다. LSB를 변경해가면서 이 점들은 실험을 통해 $R_M(1/2)=S_M(1/2)$인 점을 찾는다. 그런데 이 점은 특정한 LSB 변경 방법에 의존적이므로 정확한 점을 찾기 위해서 다수의 실험을 반복하는 것이 바람직하다.



〔그림 14〕 이상적인 RS-도표를 이용한 잡음크기 예측 자료

Fridrich 등은 p를 구하기 위해 아래 이차방정식

$$2(d_1-d_0)x^2+(d_{-0}-d_{-1}-3d_0)x+(d_0-d_{-0})=0$$
$$d_0=R_M(p)-S_M(p),\ d_{-0}=R_{-M}(p)-S_{-M}(p)$$
$$d_1=R_M(1-p)-S_M(1-p),$$
$$d_{-1}=R_{-M}(1-p)-S_{-M}(1-p)$$

의 근을 구한다. 이 이차방정식의 근 가운데 절대값이 작은 것을 x라 하면 p는

$$p=\frac{2x}{2x-1}$$

가 된다. 그런데 이 방법은 네 곡선 R_M, R_{-M}, S_M, S_{-M}이 필요하다. 게다가 다음 두 조건이 반드시 성립해야 한다.

1) R_M와 R_{-M}가 교차하는 좌표는 반드시 S_M와 S_{-M}가 교차하는 x좌표와 일치해야 한다.

2) $R_M(1/2)=S_M(1/2)$

위의 조건이 성립한다면 R_M과 R_{-M} 또는 S_M과 S_{-M}만으로도 p를 예측할 수 있음을 보일 수 있다. 다만 지면의 제약으로 그 식은 이 논문에서 제시하지 않겠지만 개념을 정리하면 다음과 같다. 2차 곡선과 1차곡선은 각각 3개의 점과 2개의 점으로 모델링이 가능하다. 다음 두 곡선이 $x=0$에서 교차한다는 가정 아래 p를 구하면 된다. 이 방법은 곡선을 Fridrich 등과 달리 2개만 쓴다는 점에서 훨씬 효율적일 수 있다. 그렇지만 RS-도표를 쓰는 이 방법은 너무 많은 가정을 필요로 한다는 점에서 많은 약점을 지니고 있다. 그리고 실제 많은 영상에서는 이러 가정들이 맞지 않는다는 점에 주목할 필요가 있다. 그렇지만 가정이 잘 맞는다면 매우 강력한 방법이 될 수 있다.

3.2.3. 보편적 맹목검출 기법

앞에서 설명한 검출 방법들은 삽입하는 방법에 밀접히 관련되어 있다. 따라서 삽입 방법이 달라진다면 검출 방법도 달라진다는 문제점을 지니고 있다. 이에 비해 Farid가 제안한 방법은 스테가노그라피 알고리듬에 무관하게 판정을 내린다는 장점이 있다[3]. 그의 방법은 사전에 원본 영상과 스테고 영상의 쌍을 많이 준비해서 판별방법을 훈련시켜둔다. 훈련이 충분히 된 다음에는 스테고 영상만을 주고 비밀 메시지가 들어있는지 판별하도록 한다. 정확한 판정을 위해 특징벡터와 같은 통계량 정보를 축적하는 것이 관건인데 과연 무엇을 통계량 또는 특징벡터로 삼을 것인지 향후 연구해야 할 대목이다. 훈련을 시키는 방법으로는 뉴럴 네트워크나 유전 프로그래밍 (genetic programming) 등과 같은 소프트 컴퓨팅 기법이 사용될 수 있다.

Farid가 사용한 방법은 영상에 n레벨 DWT (digital wavelet transform) 변환을 적용해서 영상을 분할한다. DWT에서는 영상을 변환할 때마다 수평, 수직, 대각 성분으로 분할된다. 레벨 i에서의 수평, 수직, 대각 성분을 각각 $H_i(x, y)$, $V_i(x, y)$, $D_i(x, y)$라고 정의한다. 각각의 서브밴드 영상에 대해 평균, 분산, 첨도 (kurtosis), 왜도 (skewness) 값을 구한다 (그림 15 참조). 레벨 i에서 수평, 수직, 대각 성분의 평균, 분산, 첨도, 왜

도를 각각 구하므로 모두 12개의 통계량이 구해지고, 전체 $(n-1)$ 레벨에 대해 통계량을 구하므로 $12(n-1)$ 개의 통계량을 구하게 된다. 여기서 통계량 왜도 $k^{(3)}$과 첨도 $k^{(4)}$는

$$k^{(3)} = E\left\{\left[\frac{x(\zeta)-\mu_x}{\sigma_x}\right]^3\right\},\quad k^{(4)} = E\left\{\left[\frac{x(\zeta)-\mu_x}{\sigma_x}\right]^4\right\}-3$$

으로 정의된다.

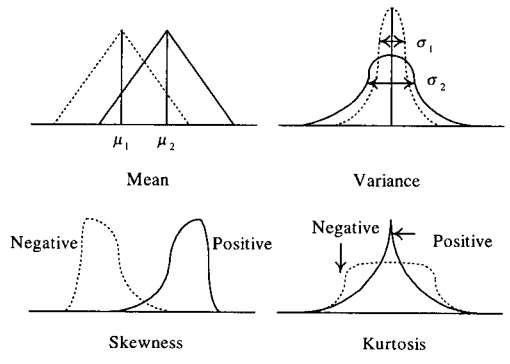
다음 그는 최적의 선형예측기를 이용해서 예측기 에러를 최소화하는 계수들을 정한다. 예를 들어 수직 선형예측기는

$$\begin{aligned} V_i(x,y) = {} & \\ & w_1 V_i(x-1,y) + w_2 V_i(x+1,y) + w_3 V_i(x,y-1) \\ & + w_4 V_i(x,y+1) + w_5 V_{i+1}(x/2,y/2) + w_6 D_i(x,y) \\ & + w_7 D_{i+1}(x/2,y/2) \end{aligned}$$

로 놓는다. 마찬가지로 수평과 대각 서브밴드에 대해서도 선형예측기를 설계해서 에러를 최소화하는 계수들을 구한다. 수직성분 $V_i(x, y)$의 예측기를 $\overline{V}_i(x, y)$라 놓으면 둘 사이에는 오차가 존재하므로 그 오차를

$$E_v(i) = \log_2(V_i(x,y)) - \log_2(\overline{V}_i(x,y))$$

라고 놓는다. Farid는 여기서도 평균, 분산, 첨도, 왜도를 구한다. 따라서 원래의 $12(n-1)$개의 통계량과 에러에서 구한 $12(n-1)$개의 통계량, 즉 도합 $24(n-1)$개의 통계량을 이용해서 특징벡터를 구성한다.



〔그림 15〕 통계량의 개념

이들 통계량을 이용해서 판별기를 훈련시키고 판별에는 Fisher의 선형판별 방식을 적용한다. 이를 위해 데이터베이스에 저장된 대량의 영상을 이용해서 원본영상과 스테고 영상을 판별하는 적절한 경계값을 정한다.

스테가노그라피 방식에 무관하게 판정한다고 해서 Farid의 방식을 보편적 판정방식이라고 부른다. Maurer의 통계적 잡음 판별방법[6]도 사실은 보편적 판정방식이라고 굳이 분류할 수 있다. 이 방법은 압축이 가능하다면 여유정보가 많다는 것을 의미하고, 여유정보가 많다는 것은 확률적 치우침 현상이 존재하므로 잡음과 거리가 멀다고 본다. 이런 성질을 이용해서 픽셀 값이 얼마나 잡음에 가까운지 판정하고 그 결과로부터 스테고 영상인지를 이론적으로는 판별할 수도 있다고 주장할 수 있다. 그러나 실제로는 거의 쓰이지 않는다. 영상 가운데는 잡음에 가까운 영상도 있기 때문에 이 방법을 적용하기에는 무리가 따른다.

4. 결 어

이 논문에서는 비밀 메시지를 삽입하는 방법의 기본 원리를 설명했다. 그러나 삽입방법은 앞으로도 무수히 개발될 것이다. 스테가노그라피의 비밀이 일단 노출되면 그 생명이 끝난다고 본다. 그래서 안정성이 입증된 방법이 개발될 때까지 꾸준히 새로운 스테가노그라피 기술은 개발될 것이다.

이 논문에서는 숨기는 방법이 아니라 숨겨진 메시지가 있는지 알아내는 방법에 중점을 두었다. 그리고 그것도 어떤 메시지가 숨겨져 있는지를 알아내는 것이 아니라 과연 메시지가 숨겨져 있는지만 알아내는데 주안점을 두었다. 그 이유로는 우선 메시지 은닉 여부만 알아내는 것도 현재 기술수준으로는 매우 어렵기 때문이다. 둘째는 메시지가 숨겨져 있다는 심증이 있을 때 거기서 메시지를 찾아내는 것은 또 다른 난제이기 때문이다. 일단 메시지가 숨겨져 있을 것으로 추측되는 영상만 골라낼 수 있어도 상대에게는 여러 측면에서 위협이 된다. 사실 상대는 스테가노그라피 사용여부를 호도하기 위해 일부러 깨끗한 영상을 다수 내보낼 수 있다. 그런데 그 많은 영상을 다 뒤지는 것은 시간 낭비일 수 있다. 그러므로 수천 수만 개의 영상 가운데서 수백 개의 영상만으로 압축할 수 있다면 그것만으로도 대단한 개가라고 할 수 있다.

본 논문에서는 특별히 스테고 분석기법으로 Westfeld와 Pfitzmann의 방법[14], Fridrich, Goljan, Du의 방법[4,5], Farid의 방법[3]을 소개했다. JPEG 영상에 대한 분석법이 Fridrich, Goljan, Du[3]에 의해 제안되었으나 지면의 한계로 이 논문에서는 소개하지 않았다. 아무튼 이 논문에서 소개한 방법들은 수긍할만한 이론적 근거를 바탕으로 삼아 개발되었다고 보기에는 아직 넘어야 할 산이 많이 있지만 이 분야 연구수준이 아직 황무지 상태에 있어 그나마 최소한 이론적 근거를 마련하려고 노력했다는 점에 점수를 주었다. 스테가노그라피에서의 보안[8] 기준을 마련하려는 시도가 있지만 이 역시 아직은 초보적인 단계에 머물고 있다.

향후 연구의 주류는 "자연스러운 영상"과 "부자연스러운 영상"을 판별하는 기준 마련이 될 것으로 본다. 스테고 분석과 같은 연구가 진행된 이후로 "자연스러움"에 대한 관심이 증가하고 있지만 만족스러운 결과가 얻어진 것은 없다. 사실 자연스럽다는 것 자체가 자연스럽지 못한 정의라고 해도 과언이 아니다. 분명한 것은 정보를 더 숨길수록 엔트로피가 증가한다. 이런 점을 이용한 연구도 아직 초보단계에 머물고 있다. 이 분야 연구가 어려운 것은 맹목검색, 즉 원본 없이 메시지가 숨겨져 있는지를 알아내야 하는 제약 때문이라고 할 수 있다. 따라서 엔트로피가 증가했다면 기준이 있어야 하고 부자연스럽다면 역시 기준이 있어야 하는데 원본이 없기 때문에 이런 비교가 어렵다는 것이 가장 큰 장애요인이라고 할 수 있다.

감사의 글

이 논문은 KAIST IDEC의 지원으로 이루어졌음을 "ISMC 연구회"의 이름으로 밝히며 지원에 감사드립니다. 그림 1을 편집해주신 강원대학교 제어계측공학과 대학원 김춘화씨에게 감사드립니다. 이 논문과 관련된 내용에 대한 질문에 성심껏 답변해주신 SUNY Binghamton의 J. Fridrich 박사에게도 감사드립니다.

참 고 문 헌

[1] R. J. Anderson, and F. A. P. Petitcolas, "On the limits of steganography," *IEEE Journal of Selected Areas in Communications*, vol. 16, no. 4, pp. 747-781, 1998.

[2] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann, 2002.

[3] H. Farid, Detecting Steganographic Messages in Digital Images, Technical Report TR2001-412, Department of Computer Science, Dartmouth College, 2001.

[4] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG compatibilty," *SPIE Multimedia Systems and Applications IV*, Denver, Colorado, USA, 2001.

[5] J. Fridrich, M. Goljan, and R. Du, "Distron-free data embedding," *Lecture Note in Computer Science*, vol. 2137, vol. Springer-Verlag, 2001.

[6] J. Fridrich, and M. Goljan, "Practical steganalysis of digital images: State of the art," *Proceedings of the Electronic Imaging*, SPIE, San Jose, California, 2002. (submitted)

[7] N. Johnson, and S. Jajodia, "Steganalysis of images created using current steganography software," *Lecture Notes in Computer Science*, pp. 273-289, 1998.

[8] S. Katzenbeisser, and F. A. P. Petitcolas, "On defining security in steganographic systems, *Proceedings of the Electronic Imaging*, SPIE, San Jose, California, 2002. (submitted)

[9] J. Kelley, "Terror groups hide behind Web encryption," *USA Today*, February 5, 2001.

[10] U. M. Maurer, "Universal statistical test for random bit generators," *Journal of Cryptology*, vol. 5, no. 2, pp. 89-105, 1992.

[11] D. McCullagh, "Secret messages come in waves," *Wired News*, February 2001.

[12] E. A. P. Petitcolas, R. J. Anderson, and M. Kuhn, "Information Hiding: A

survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062-1078, 1999.
[13] N. Provos, "Defending against statistical steganalysis," *10th USENIX Security Symposium*, Washington DC, USA, 2001.
[14] A. Westfeld, and A. Pfitzmann, Attacks on steganographic systems," *Lecture Notes in Computer Science*, vol. 1768, Springer-Verlag, pp. 61-75, 2000.
[15] 김형중, 여인권, "블라인드 워터마킹: 튜토리얼," 방송공학회 논문지, vol. 6, no. 3, pp. 270-282, 2001.

〈者 著 紹 介〉

김 형 중
종신회원
1978년: 서울대학교 전기공학과 학사
1986년: 서울대학교 제어계측공학과 석사
1989년: 서울대학교 제어계측공학과 박사
1992년~1993년: USC 방문교수
현재: 강원대학교 제어계측공학과 교수, 정보보호학회 멀티미디어보호연구회 (SIGMP) 위원장
관심분야: MPEG-21 IPMP, 디지털 워터마킹, 디지털 스테가노그라피, Quantum Computing