

특집논문 (Special Paper)

방송공학회논문지 제23권 제4호, 2018년 7월 (JBE Vol. 23, No. 4, July 2018)

<https://doi.org/10.5909/JBE.2018.23.4.462>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

## 영상 스테가노그래피의 개념과 판별

이 재 훈<sup>a)</sup>, 김 찬 란<sup>a)</sup>, 이 상 화<sup>b)†</sup>, 박 종 일<sup>a)</sup>

## Image Steganography and Its Discrimination

Jae Hoon Lee<sup>a)</sup>, Chanran Kim<sup>a)</sup>, Sang Hwa Lee<sup>b)†</sup>, and Jong-Il Park<sup>a)</sup>

### 요 약

스테가노그래피란 숨기려는 정보를 미디어 데이터에 교묘하게 삽입하여 정보의 존재를 전혀 눈치 채지 못하도록 하는 데이터 은닉 기술의 하나이다. 데이터 암호화의 경우에는 정보 데이터에 직접 복호하기 어렵게 복잡한 부호화 기법을 이용하여 변형하는 것으로, 정보의 존재를 쉽게 인지할 수 있지만 복호를 어렵게 하는데 초점이 맞춰져 있다. 그러나 스테가노그래피는 다른 미디어에 정보를 자연스럽게 숨겨서 정보의 존재를 눈치 채지 못하게 함으로써, 더욱 강력하게 정보를 보호할 수 있다. 이 논문에서는 영상에 다른 영상 정보를 숨기는 영상 스테가노그래피의 개념을 소개하고, 스테가노그래피 부호화 여부를 판별하는 기법을 제안한다. 주로 영상 픽셀의 LSB (Least Significant Bit)를 조작하는 공간영역의 기법과, JPEG 압축시 DCT 계수를 조작하는 주파수 영역의 영상 스테가노그래피를 소개하고, 이들의 스테가노그래피 부호화 여부와 복호결과를 판별하는 새로운 방법을 제안한다.

### Abstract

Steganography is to hide information in a media data so naturally that the other users can not notice the existence of hidden information. Data encryption focuses on the complex encoding methods of information data in order to make it difficult to decode the information even though the other users notice the existence of information. On the other hand, steganography methods concentrate on natural hiding information into other media. Since the other users do not recognize the existence of hidden information, the information can be better protected. This paper introduces the concept of image steganography that an image information is concealed into another image (cover image), and proposes a new discrimination method of steganography. This paper explains the spatial methods with LSB manipulation and frequency methods using DCT coefficients. Finally, this paper proposes a new discrimination method of image steganography by inspecting that an image information is correctly decoded.

Keyword : Image steganography, LSB manipulation, DCT coefficients, steganography discrimination

a) 한양대학교 컴퓨터소프트웨어학과(Department of Computer Science, Hanyang University)

b) 서울대학교 뉴미디어통신공동연구소(Institute of New Media and Communications, Seoul National University)

† Corresponding Author : 이상화(Sang Hwa Lee)

E-mail: lsh529@snu.ac.kr

Tel: +82-2-880-8394

ORCID:<https://orcid.org/0000-0003-0228-0459>

※ 이 논문의 연구 결과중 일부는 “IPTU 2018”에서 발표한 바 있음.

※ 본 연구는 방위사업청 및 국방과학연구소에 의해 설립된 신호정보 특화연구센터의 지원을 받아 수행되었음.

※ This work was supported by the research fund of Signal Intelligence Research Center supervised by the Defense Acquisition Program Administration and the Agency for Defense Development of Korea.

· Manuscript received May 2, 2018; Revised June 11, 2018; Accepted, June 11, 2018.

Copyright © 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”



그림 1. 영상 스테가노그래피의 개념  
 Fig. 1. Image steganography concept

## 1. 서론

스테가노그래피는 전달하려는 정보를 이미지나 오디오 등의 미디어에 숨기되, 외부에 정보 자체의 존재 여부가 전혀 드러나지 않도록 하는 정보 은닉 기술의 총칭이다. 기존의 암호화 방법(cryptography)은 비밀키를 이용하여 다른 사람은 풀 수 없는 형태로 데이터를 변형하는 기술로서, 변형된 정보가 전송되는 과정에서 외부에 쉽게 노출이 되지 않음만 해독할 수 없도록 데이터를 변형하는데 중점을 둔다. 반면에 스테가노그래피는 처음부터 비밀 메시지의 존재 유무를 완벽하게 숨기려고 하는 것이 스테가노그래피 기술의 핵심 특징이다. 스테가노그래피에서 정보를 숨겨 전달하는 매체는 다양하지만 그 중에서 많은 정보를 숨기면서도 자연스러움을 유지하기 때문에 영상에 정보를 숨기는 영상 스테가노그래피가 널리 사용된다<sup>1-3)</sup>.

그림 1은 영상 스테가노그래피의 개념을 보여준다. 그림 1에서 정보를 숨기려는 영상을 커버(cover) 영상이라고 하고, 정보가 숨겨진 영상을 스테고(stego) 영상이라고 하며 스테고 영상을 만드는 과정을 스테가노그래피 부호화(stegano-encoding)라 한다. 그리고 스테고 영상에서 복호화 알고리즘을 통해 숨겨진 원 영상(original)을 알아내는 과정을 스테가노그래피 복호화(stegano decoding)라 한다. 일반적으로 픽셀의 LSB(Least Significant Bit)를 조작하는 공간 영역의 영상 스테가노그래피에서는 커버 영상의 픽셀 수 대비 25%에 이르는 용량의 영상을 숨길 수 있는데, 이 경우에도 원래의 커버 영상과 스테고 영상 간에 40dB 정도의 PSNR을 갖기 때문에 육안으로는 도저히 차이를 구분할

수 없다.

영상 스테가노그래피의 기술범주는 크게 두 가지 영역으로 나뉘는데, 각각 공간 영역에서의 기술과 주파수 영역에서의 기술에 해당한다. 공간 영역에서의 영상 스테가노그래피는 데이터를 숨기기 위해 영상의 픽셀값에 직접 데이터를 숨기는 방법으로서, 대표적으로는 커버 영상의 각 픽셀의 LSB를 이용하여 정보 영상의 픽셀 값을 나누어 숨긴다. 이 방법은 많은 양의 데이터를 빠르고 간단하게 숨길 수 있고, 스테고 영상이 시각적으로 차이가 없도록 하는 강력한 방법이다. 또한 숨기려는 픽셀의 위치를 무작위로 변경하거나 순서는 값의 크기를 간단하게 조작하기만 하더라도 원래의 정보를 거의 찾을 수 없다. 다만, 공간 영역의 기법은 픽셀 값을 직접 조작하므로, 영상 압축과 같은 후처리가 수반될 경우 원본 정보영상을 복원할 수 없다.

주파수 영역에서의 영상 스테가노그래피는 숨기려는 데이터를 주파수 영역에서 조작하는 방법으로, 주로 이미지를 압축하는 방법 중 하나인 JPEG에서 DCT(Discrete Cosine Transform) 계수를 조작하여 이진수를 숨기는 방법으로 이루어진다. DCT 계수를 조작하는 방법들도 영상의 왜곡이 적어 사람의 눈으로 구별하기 어렵지만, 조작이 가능한 DCT 계수는 영상의 픽셀 개수보다는 훨씬 적기 때문에, 숨길 수 있는 데이터의 양이 상대적으로 적고 영상의 특성에 따라 크기가 동일한 커버 영상이라 하더라도 숨길 수 있는 데이터의 양이 일정하지 않다.

2절에서는 공간 영역과 주파수 영역에서의 영상 스테가노그래피 기법들에 대해 좀 더 자세하게 살펴보도록 한다. 3절에서는 임의의 영상으로부터 스테가노그래피로 부호화

된 스테고 영상 여부를 판별하는 영상 스테가노그래피 판별 기법을 제안한다. 순수한 커버 영상과 스테고 영상은 육안으로는 구분되지 않기 때문에, 스테고 영상에 대한 분석을 통하여 스테가노그래피 부호화 여부를 판별하는 기법이 필요하다. 이 때 영상의 엔트로피 특성과 픽셀 값의 분산에 대한 상관관계, DCT 계수의 분포와 영상의 분산 특성 등을 학습한 판별 기법을 제안한다. 특히, 픽셀 값을 조작하는 공간 영역 기법은 간단하면서도 판별하기 어려운 부호화 기법이 매우 많기 때문에, 실질적으로 부호화 여부를 직접 확인하거나 정확한 부호화 기법을 찾아내는 것은 거의 불가능하다. 이를 극복하기 위하여 본 논문에서는 부호화 라이브러리에서 임의로 선택한 방법으로 숨겨진 영상 정보를 복원해보고, 그 복원 결과가 올바르게 복원된 영상 정보인지를 판별함으로써, 부호화 방법과 정보 복원을 동시에 판별하는 기법을 제안한다. 따라서 본 논문에서는 공간영역으로 부호화된 스테고 영상으로부터 영상 스테가노그래피 부호화 여부를 판별하고 숨겨진 영상 정보를 동시에 복원할 수 있는 기법을 다룬다. JPEG DCT 계수를 조작하는 주파수 영역에 대해서는 DCT 계수의 분포와 영상 특성을 분석하여 직접 부호화 여부를 판별하는 기법을 제안한다.

## II. 영상 스테가노그래피 기법

### 1. 공간 영역에서의 영상 스테가노그래피 기법

공간 영역에서의 스테가노그래피는 영상의 픽셀 값을 이용하여 정보를 숨겨 커버 영상과 시각적으로 똑같지만 영

상 정보가 담긴 스테고 영상을 만들어낸다. 보통 영상의 한 픽셀은 8비트로 표현이 되며 영상 스테가노그래피는 LSB에 해당하는 1~3비트의 정보를 다룬다. 이는 전체 이미지의 2~3% 정도의 차이에 해당하며 값을 바꾸더라도 인간의 눈으로 감지할 수 없을 정도의 차이이다. 이러한 인간 눈의 한계를 이용하여 공간 영역에서 LSB를 활용한 영상 스테가노그래피를 수행한다.

우선 LSB를 이용하기 위해서는 두 영상이 필요하다. 하나는 겉으로 보이는 커버 영상, 다른 하나는 전송의 목적이 되는 원 정보영상이다. 동일한 크기의 두 영상이 있다고 가정을 한다면 LSB에 원 영상 정보를 모두 넣기 위해서는 커버 영상을 두 배로 늘려야 한다. 커버 영상을 두 배로 확장시키면 공간 비율이 4:1이 되어 커버 영상 4픽셀의 2개씩 하위비트(LSB)에 원 영상 1픽셀에 해당하는 값을 전부 담을 수 있다. 그림 2는 LSB를 이용한 스테가노그래피 부호화 과정을 그림으로 표현한 것이다. 우선 커버 영상의 한 픽셀 데이터를 스테고 영상의 이웃하는 4개의 픽셀에 대응시킨다. 그리고 각 4개 픽셀의 LSB 2비트 데이터를 원 영상의 한 픽셀 데이터로 대체시켜주면 스테고 영상이 만들어지며 커버 영상 대비 값의 차이가 최대 3만큼밖에 나지 않아 육안으로 구별할 수 없게 된다.

하지만 단순히 원 영상의 정보를 그대로 커버 영상의 LSB에 저장하는 것은 쉽게 숨겨진 정보를 확인할 수 있는 취약점이 존재한다. 정직하게 원 영상을 그대로 부호화하는 것보다 변형시켜서 표현하는 것이 데이터를 복원하기 어렵게 한다. 대표적인 방법으로는 커버 영상과 원 영상의 픽셀 값 차이를 구한 차 영상을 이용하거나 위치를 무작위로 섞어서 숨긴다<sup>[47]</sup>. 차 영상을 만들게 되면 음의 부호를

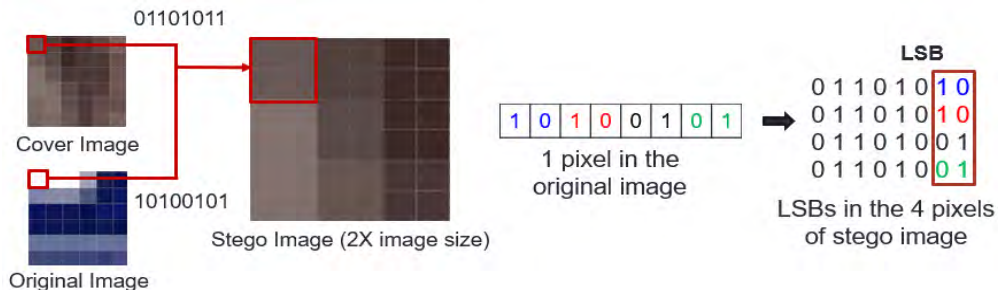


그림 2. LSB 스테가노그래피 인코딩 과정  
Fig. 2. LSB steganography encoding process

고려해야 하고 비트수를 좀 더 줄이기 위하여 차이 값에 modulus 2나 5, 또는 XOR 연산 기법들을 활용하여 공간 영역에서의 스테가노그래피 부호화를 수행한다<sup>[8]</sup>. 식 (1)~(4)는 각각 차 영상, modulus 2, modulus 5, 그리고 XOR 연산에 대한 수식이며  $D_{x,y}$ 는 차 영상,  $C_{x,y}$ 는 커버 영상 그리고  $O_{x,y}$ 는 원 영상에 대한 기호이다.

$$D_{x,y} = C_{x,y} - O_{x,y} \quad (1)$$

$$D_{x,y} = (D_{x,y} + 1)/2 \quad (2)$$

$$D_{x,y} = (D_{x,y} + 2)/5 \quad (3)$$

$$D_{x,y} = C_{x,y} \oplus O_{x,y} \quad (4)$$

가장 단순하게 modulus 2 연산을 취하는 것은 차 영상이 가질 수 있는 데이터의 범위인 -255~255를 8비트로 표현할 수 없기 때문에 2로 나누어 -127~128로 축소시키는 방법이다. 그리고 거기에 127을 더해, 최종적으로 0~255의 8비트 범위로 만들어 스테가노그래피 부호화를 가능하게 만든다. 그리고 modulus 5 연산은 차 영상이 가질 수 있는 -255~255의 범위에 해당하는 데이터 값을 5로 나누어 1개의 값에 5개의 값이 대응되게 하고 대응되는 값의 수가 동일하도록 2를 더하여 총 7비트로 표현하는 방법이다. 마지막으로 XOR 연산을 이용하여 차 영상을 만드는 방법은 부호 비트를 필요로 하지 않아 연산 결과 값을 그대로 사용할 수 있어 간단하다.

그 밖에 커버 영상에 원 영상의 데이터를 숨길 때 대응되는 위치가 복잡하면 복원하기가 더 힘들어진다. 따라서 임의의 순열 행렬(random permutation matrix)을 이용하여 변형된 LSB의 위치를 섞게 되면 보안성이 한 층 더 높아진 스테가노그래피 부호화가 가능하게 된다<sup>[8]</sup>. 순열 행렬(permutation matrix)은 이중 장치의 역할을 하는 key 값으로 송신 측과 수신 측이 초기에 생성하여 서로 갖고 있어야 하고 스테고 영상을 부호화하거나 원 영상을 복호화할 때에 필요하다. 아래 식 (5)와 식 (6)에서  $I$ 는 기존 영상,  $X$ 는 임의의 순열로 바꾼 영상이다.  $P_r$ 는 행에 대한 임의의 순열 행렬이



그림 3. 원 영상과 스테가노그래피 차 영상 기법 영상 (modulus 2, 5 and XOR)

Fig. 3. Original image and subtracted images by modulus 2, 5 and XOR methods

고,  $P_c$ 는 열에 대한 임의의 순열 행렬이다. 임의의 순열 행렬은 각 행과 열의 단 하나의 원소만 1이고 다른 모든 원소가 0인 정방 행렬로 이것을 목적이 되는 행렬에 곱하게 되면 그 행렬의 행의 순서를 바꿔주는 역할을 한다.  $P_c$ 를 전치 행렬로 이용하는 것은 열의 순서를 바꿔주기 위함이다.

$$(P_c(P_r I)^T)^T = P_r I P_c^T = X \quad (5)$$

$$P_r (P_c X^T)^T = P_r X P_c^T = I, \quad (6)$$

where  $P_r \cdot P_r = E, P_c \cdot P_c = E$

## 2. 주파수 영역에서의 영상 스테가노그래피 기법

가장 대표적인 주파수 영역에서의 스테가노그래피 알고리즘은 JPEG 압축 시 DCT 이후 양자화를 거친 계수들의 LSB를 조작하여 데이터를 숨기는 방법들이다. JSteg, F3, F4, F5<sup>[9]</sup>가 JPEG 압축 시 이뤄지는 대표적인 영상 스테가노그래피 기법들이며, 양자화된 DCT 계수들에 데이터를 숨긴 후 Huffman 코딩 또는 Arithmetic 코딩을 진행하여

압축된 영상을 생성한다. 그리고 나면 원본 영상과 차이가 작은 스테고 영상을 생성하게 되며, 그 안에는 숨겨진 정보가 존재하게 된다. 수신자는 디코딩을 통해 양자화된 DCT 계수를 복원한 다음 적용된 스테가노그래피 알고리즘에 맞게 계수들을 읽으면 그 원 정보를 알아낼 수 있다.

JSteg 알고리즘<sup>[10]</sup>은 JPEG에서 사용되는 가장 기초적인 스테가노그래피 알고리즘으로 DCT이후 양자화 된 테이블에서 각 계수들의 LSB를 숨기고자 하는 데이터의 비트로 교체하는 형식이다. 이 때 0과 1을 제외한 나머지 DCT 계수들에 대해서만 변환을 진행한다. 그런데 DCT 계수들의 빈도수는 0일 때 가장 많고 계수의 절댓값이 클수록 빈도수가 줄어들며, 양수와 음수 계수에 대해 빈도수가 대칭을 이루는 통계적인 특성을 갖고 있다. 따라서 JSteg을 적용시키게 될 경우 이 특성이 파괴되어 DCT 계수의 분포를 육안으로 관찰할 경우 스테고 영상을 쉽게 구별해 낼 수 있다.

이러한 DCT 계수 분포 대칭성 왜곡을 없애기 위해 발전한 알고리즘이 F3이다. F3은 0을 제외한 DCT 계수들의 LSB가 숨기고자 하는 데이터의 비트와 다를 때 계수의 절댓값을 감소시켜서 부호화하는 방법이며 그림 4는 F3 알고리즘으로 커버 영상을 스테고 영상으로 부호화 하는 예시이다. 예를 들어 커버 영상에서 JPEG 계수 -2의 LSB는 0을 가지며 만일 숨기고자 하는 데이터의 현재 비트가 1이라면 절댓값을 감소시켜 계수를 -1로 만들고 반대로 0이라면 그대로 두어서 부호화를 진행한다. 또한 커버 영상의 JPEG 계수가 0일 경우에는 데이터를 넣게 되면 대칭성이 파괴될 수 있으므로 부호화하지 않고 다음 JPEG 계수로 넘어가도록 한다. 그러면 수신자가 스테고 영상을 읽을 때 0

을 제외한 JPEG 계수들의 LSB만 읽으면 숨겨진 정보를 추출해낼 수 있다.

F3은 절댓값이 작은 DCT 계수들의 빈도수가 높고 음수와 양수 계수의 빈도수 대칭성이 유지되는 특성을 가진다. 하지만 DCT 계수 -1이나 1의 LSB가 숨길 데이터의 비트와 달라 절댓값을 감소시켜 0이 되어버릴 경우 숨긴 데이터의 비트인지 원래 이미지의 DCT 계수 0인지 구분할 수 없기 때문에 다음 DCT 계수에 대하여 다시 변환 과정을 진행해야 한다. 이 때문에 스테가노그래피 부호화를 하게 되면 DCT 계수 0의 빈도수가 상당히 많아지는 단점이 존재한다.

F4 알고리즘은 F3 알고리즘에서 DCT 계수 0의 빈도수가 많아지는 취약점을 해결하기 위해 보완된 기법으로 DCT 양수 계수들은 F3과 동일한 방법으로 부호화를 수행하지만 음수 계수의 경우에는 반대로 LSB가 숨길 데이터의 비트와 같을 때 절댓값을 감소시킨다. 그림 5는 F4 알고리즘으로 부호화 하는 과정이며 음수 계수들에 대해서 반대로 동작하는 것이 F3과의 차이점이다. 특히 JPEG 계수 -1과 1이 숨길 데이터 비트에 의해 0으로 바뀌는 값이 달라 상대적으로 스테고 영상의 JPEG 계수 0이 만들어지는 빈도수가 줄어들게 되어 커버 영상의 DCT 계수 분포와 유사해지게 된다.

이 외에도 F5 알고리즘도 존재하지만 이는 숨길 수 있는 정보의 양을 늘리기 위해 나온 방법이며 F4 알고리즘과 유사한 특성을 가졌다. 따라서 이러한 기법들로 부호화 된 영상을 판별하기 위해서는 DCT 계수의 빈도수 변화 특성과 이에 따른 영상 특성 변화를 자세히 분석할 필요가 있다.

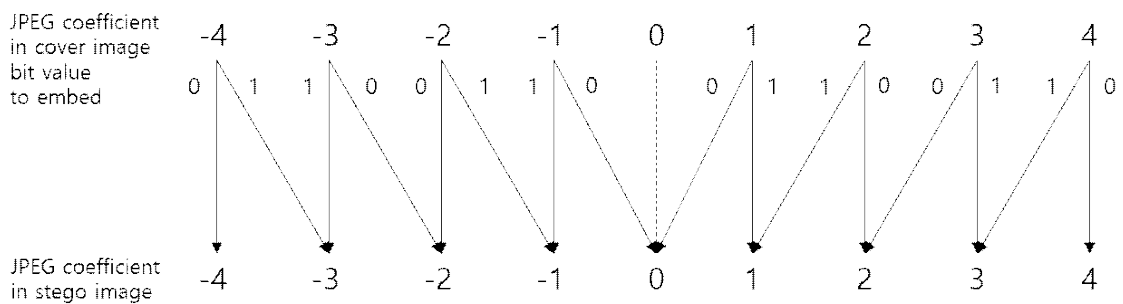


그림 4. F3 부호화 과정  
Fig. 4. F3 encoding process

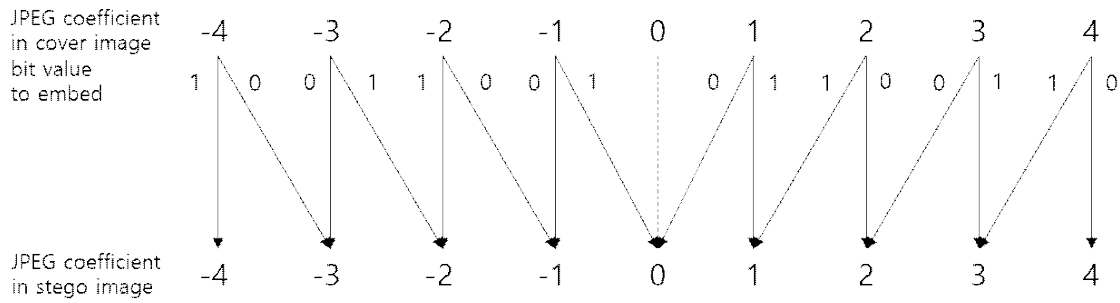


그림 5. F4 부호화 과정  
 Fig. 5. F4 encoding process

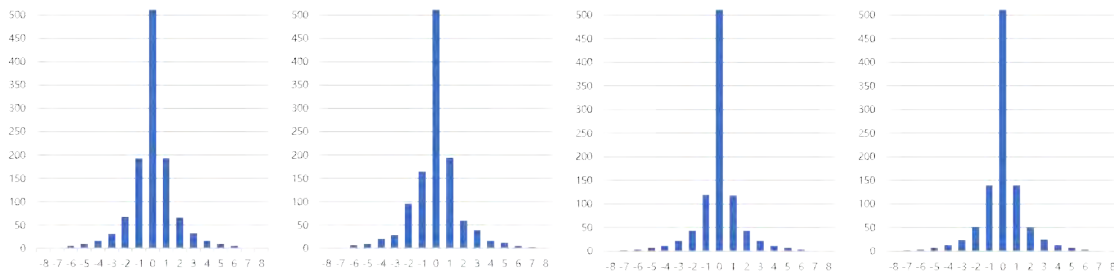


그림 6. DCT 계수 빈도수 분포(왼쪽부터 일반 영상, Jsteg, F3, F4)  
 Fig. 6. distribution of DCT coefficients frequency (original image, Jsteg, F3, F4)

그림 6은 일반 영상과 주파수 영역에서의 각 스테가노그래피 기법을 적용시킨 영상의 DCT 계수 빈도수 분포를 보여주고 있다.

### III. 영상 스테가노그래피 판별 기법

#### 1. 공간 영역에서의 영상 스테가노그래피 판별 기법

공간 영역에서의 스테가노그래피 기법들은 다양하고 의심받지 않을 정도로 고도의 미세한 정보 조작이 이루어진다. 예를 들어 임의의 순열 기법을 이용하였을 경우 key를 모르는 사용자가 모든 조합을 이용하여 원 영상이 제대로 복원되었는지를 확인하는 것은 사실상 불가능하다. 따라서 LSB를 조작하는 영상 스테가노그래피 기법에서는 부호화 여부를 바로 판별하는 것이 매우 어렵다<sup>[11]</sup>. 스테고 영상의 하위비트 특성을 분석하여 스테가노그래피 부호화 여부를 직접 판단하는 기법에 대한 연구가 지금까지 진행되어 온

주요 방향이었으나, 영상의 특성과 숨겨진 데이터의 양에 따라서 판별 성능이 크게 달라지고 정확도도 70% 수준이라는 한계를 갖고 있다<sup>[14]</sup>. 본 논문에서는 스테고 영상을 직접 분석하여 부호화여부를 판별하는 기존의 기법과는 전혀 다르게 스테가노그래피로 복호시킨 결과 영상을 분석함으로써, 올바르게 복호된 영상과 잘못 복호된 영상이 갖는 다른 특성을 이용하여 부호화가 올바르게 이루어졌는지를 판별하는 방법을 제안한다. 이를 통하여 스테가노그래피 부호화 방법과 올바르게 복호된 정보를 동시에 얻을 수 있으며, 일단 복호를 통하여 스테가노그래피 부호화 여부를 확실하게 판별할 수 있다.

잘 복호된 일반적인 영상과 잘못 복호된 노이즈 영상은 그림 7에서 볼 수 있듯이 영상의 복잡한 정도가 다르다<sup>[12]</sup>. 따라서 이 차이를 극명하게 나타내기 위해 16X16 블록에 대해 지역적 복잡도 특성에 해당하는 영상의 국소 엔트로피 값을 조사하였다. 그림 8은 잘 복호된 정상적인 영상과 그렇지 못한 비정상적인 영상의 16X16 블록에 해당하는 국소 엔트로피 값의 평균과 분산에 대한 분포 그래프다. x축은 국

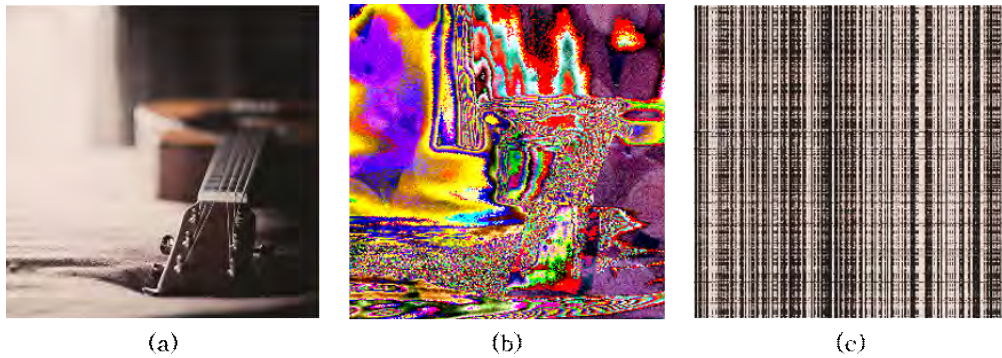


그림 7. 잘 복호된 원 영상(a), 잘못 복호된 영상(b), 순열이 잘못된 영상(c)  
 Fig 7. original image(a), wrong decoded image by irrelevant method(b), wrong decoded image by wrong permutation matrix(c)

소 엔트로피의 평균을 뜻하고 y축은 국소 엔트로피의 분산을 뜻한다. 그림 8을 보면 그림 7에서 보이는 노이즈에 해당하는 비정상적인 영상의 random dot에 의해 엔트로피가 높게 나타나며 블록 간의 엔트로피 편차는 작게 나타나는 것을 확인할 수 있었다. 반면 정상적인 영상의 경우에는 엔트로피의 평균이 영상 특성에 따라 다양하며, 엔트로피 분산은 블록 내 영상 정보에 따라서 크게 나타남을 알 수 있었다. 이는 정상적인 영상이라면, 블록 내에서는 픽셀 값의 유사성이 높고, 블록 간에는 변화가 많다는 영상의 특성을 반영한 결과

라 할 수 있다. 반대로 비정상적인 경우에는 모든 블록이 random dot 영상처럼 보이므로, 엔트로피가 클 뿐만 아니라, 블록간의 엔트로피 편차가 매우 작게 나타난다.

따라서 본 논문에서는 잘 복호된 영상과 잘못 복호된 영상 간에 복잡도 특성이 다름을 이용하여 기계 학습 방법으로 판별 기법을 제시하였다. YCbCr 색 공간에서 영상을 16X16 블록으로 나누어 엔트로피를 계산하고 그 평균과 분산을 구한 6차원의 데이터를 학습 모델로 설정하고 SVM(Support Vector Machine)을 활용한 기계 학습 방법으로 학습하여 판

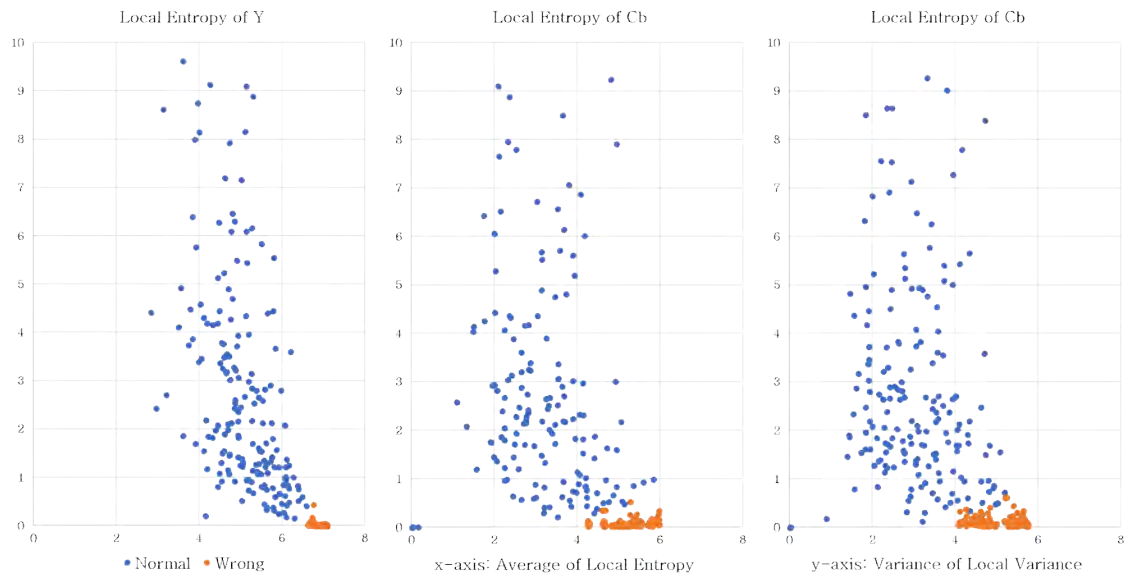


그림 8. 영상의 각 채널에 대하여 국소 엔트로피 값들의 평균과 분산  
 Fig 8. distribution of local entropy average and variance distribution in each channel of image

별 알고리즘을 구성하여 실험을 수행하였다. 아래 식(7)은 학습 모델에 대한 수식이며 YCbCr 색 공간의 각 채널별로 국소 엔트로피 평균과 분산 값으로 구성되어 있다.  $m$ 은 국소 엔트로피의 평균값으로 총  $n$ 개의 블록에 해당하는 영역의 엔트로피 값의 평균으로 계산하며 식(8)과 같이 계산된다.  $(i, j)$ 는 영상의  $k$ 번째 블록에 해당되는 영역의 좌표를 의미하며  $\sigma^2$ 는 국소 엔트로피 분산 값을 나타낸다.

$$X(i) = \{m_Y(i), \sigma_Y^2(i), m_{Cb}(i), \sigma_{Cb}^2(i), m_{Cr}(i), \sigma_{Cr}^2(i)\} \quad (7)$$

$$m = \frac{1}{n} \sum_{k=1}^n Entropy(k), \quad (8)$$

$$Entropy(k) = \sum_{(i,j) \in k} -p_{(i,j)} \log_2 p_{(i,j)}$$

$$\sigma^2 = \sum_{k=1}^n (Entropy(k) - m)^2 \quad (9)$$

## 2. 주파수 영역에서의 영상 스테가노그래피 판별 기법

JPEG 압축 시 영상의 DCT 계수 빈도수 분포는 영상의 복잡한 정도에 따라 다르게 나타난다<sup>[13]</sup>. 영상의 단순하고 복잡한 정도를 계산하는 대표적인 방법으로는 영상 픽셀

값들의 분산을 계산하는 것이 있다. 그 중 이미지를 일정 블록으로 나누어 국소 분산을 계산하고 평균을 취하는 것을 이미지의 국소 분산이라 한다. JPEG 압축은 이미지를 8X8블록으로 분할하여 DCT를 수행하며 해당 블록의 국소 분산 값이 클 경우 DCT 계수 중 0의 빈도수가 상대적으로 작은 특성이 나타나며 이 두 성분의 상관 계수를 조사하면 -0.802로 어느 정도 상관관계를 갖는다는 것을 확인할 수 있었다<sup>[13]</sup>. 여기서 F3 알고리즘을 이용하여 정보를 숨길 경우 DCT 계수 0의 빈도수가 변화하게 되며 아래 그림 9는 F3 알고리즘을 적용시켜서 각 영상마다 최대로 숨길 수 있는 양 만큼 정보를 숨긴 스테고 영상의 DCT 계수 0의 빈도수 변화를 나타낸 분포이다. 파란색 점들은 원 영상이며 주황색 점들은 F3 스테고 영상을 의미한다. 따라서 스테가노그래피가 적용되면 DCT 계수 0의 빈도수가 변하게 되어 원본 영상에서 가지는 상관관계가 변하고 이 특성을 이용하면 주파수 영역에서 영상 스테가노그래피가 적용된 여부를 판별할 수 있게 된다.

JSteg 알고리즘의 경우 육안으로 DCT 계수의 분포를 살펴보고 비대칭적인 분포를 확인하여 스테고 영상을 손쉽게 구별해 낼 수 있지만, F3와 F4 알고리즘의 경우에는 DCT 계수의 분포가 일반적인 영상과 비슷하므로 육안으로 판별하기 어렵다. 따라서 본 논문에서는 F3, F4로 부호화 된 영

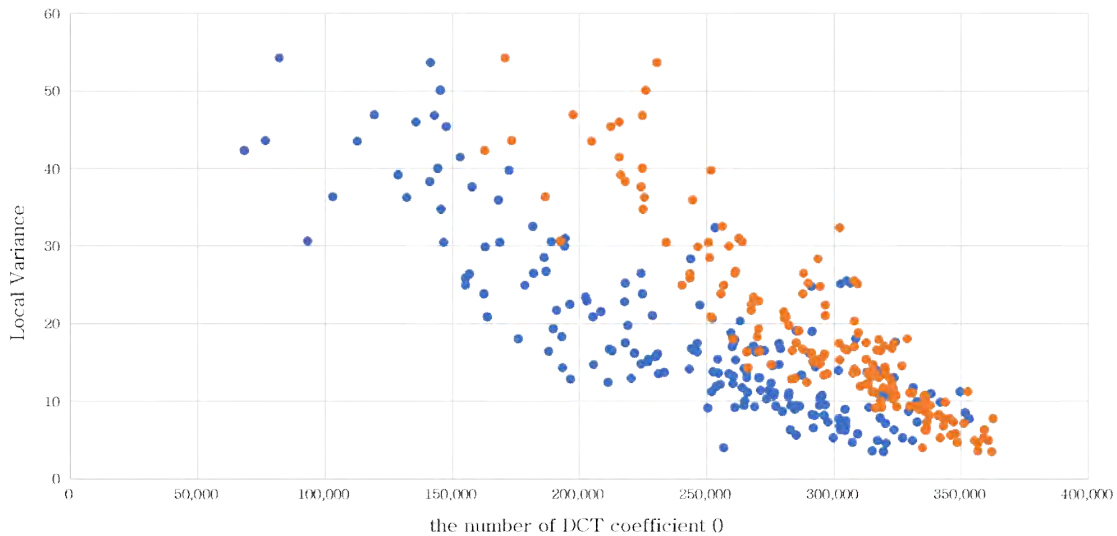


그림 9. 원본 영상과 F3 알고리즘을 적용시킨 영상의 DCT 계수 0의 빈도수와 국소 분산  
 Fig. 9. distribution of frequency of DCT coefficient 0 and local variance in original image and F3 encoded image

상을 판별하기 위해 영상의 복잡도와 DCT 계수 0의 빈도수 관계 특성과 기계 학습 방법을 이용하여 기법을 제시하였다. YCbCr 색 공간에서 JPEG 압축 시 Y 색상에 대하여 8X8 블록의 국소 분산 값의 평균과 양자화 된 테이블의 DCT 계수 0의 빈도수를 학습 모델로 설정하고 SVM을 활용해 판별 알고리즘을 구성하여 실험을 수행하였다. 아래 식 (10)은 학습에 사용된 모델에 대한 수식이며  $LV_Y$ 는 Y 색상에 대한 국소 분산 값의 평균을 의미하고  $N(0)$ 는 양자화된 DCT 계수 0의 빈도수를 의미한다. 국소 분산의 평균값은 식 (11) 처럼 8X8 블록으로 나누어진 영상의 각 영역들의 분산을 계산한 다음 그 값들의 평균으로 나타낸다.

$$X(i) = \{LV_Y(i), N(0)\}, LV = \frac{1}{n} \sum_{k=1}^n Variance(k) \quad (10)$$

$$Variance(k) = \sum_{i,j}^{(i,j) \in k} (x_{(i,j)} - mean(k))^2, \quad (11)$$

$$mean(k) = \frac{1}{64} \sum_{i,j}^{(i,j) \in k} x_{(i,j)}$$

#### IV. 실험 결과 및 분석

본 논문에서는 실험을 위해 규칙성이 없고 편향되지 않으며 임의의 이미지 500장을 수집하였다. 특히, 영상 구조가 복잡하고 다양한 기하학적 모양을 갖는 판별이 어려운 영상을 다수 포함하였다. 스테가노그래피 알고리즘을 적용

시킨 스테고 영상을 만들어 총 1000장의 학습 데이터를 구성하였고 이를 각 방법에 맞게 학습시킨 다음 판별을 수행하였다.

공간 영역에서는 영상 스테가노그래피 부호화 여부 판별을 위해서 500장의 정상 복호화 영상과 500장의 비정상 복호화 영상을 이용하여 실험을 수행하였다. 임의의 영상이 부호화된 스테고 영상이라면 복호화를 시켰을 때 그림 8의 (a)와 같이 정상적인 복호화 영상이 나타날 것이며 일반 영상을 복호화 시킨다면 그림 8의 (b)나 (c)와 같은 비정상적인 영상이 생성되기 때문에 이 차이를 구분하도록 학습 진행하였다. 이를 위해 각 영상들을 16X16 블록으로 나누어 엔트로피의 평균과 분산 값을 YCbCr 색공간의 각 채널에 대하여 계산하고 SVM 학습 방법으로 학습시킨 후 임의의 실험 영상을 이용하여 판별하였다.

주파수 영역에서의 영상 스테가노그래피 판별의 경우는 500장의 정상 영상으로부터 F3, F4 알고리즘이 적용된 스테고 영상 500장을 각각 생성하였다. 이 때, 공간 영역에서의 스테가노그래피와 달리 영상에 숨길 수 있는 정보량이 작기 때문에 최대한 원 영상과 스테고 영상이 차이가 날 수 있도록 98에 해당하는 압축 quality로 JPEG 압축 시 영상마다 최대한 숨길 수 있는 정보의 양을 활용하여 스테가노그래피 부호화를 수행하였다. 그리고 각 영상들을 8X8 블록으로 나누어 Y 채널에 대하여 국소 분산의 평균값을 계산하고 DCT 계수 0의 빈도수 값을 SVM 학습 방법으로 학습하고 판별하였다.

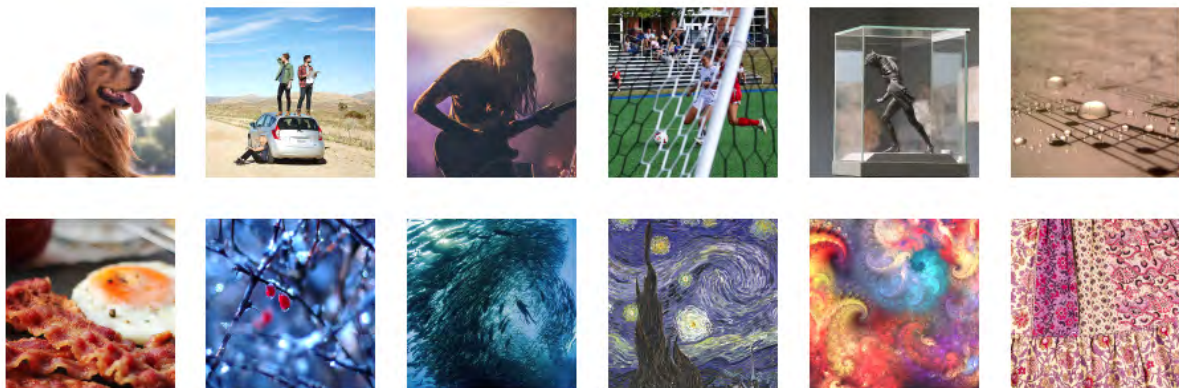


그림 10. 실험 데이터 영상 표본  
Fig. 10. Samples of image dataset

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (12)$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision+recall}$$

$$precision = \frac{TP}{TP+FP}, \quad recall = \frac{TP}{TP+FN} \quad (13)$$

표 1의 결과처럼 공간 영역에서의 영상 스테가노그래피 판별 성능은 Simple LSB 기법에 대하여 92.0%, 임의 순열 기법에 대하여 92.2% 정도로 나타났고 주파수 영역에서는 F3 기법에 대해 85.8%, F4 기법에 대해서는 81.8%의 성능을 보여주었다. 또한 정밀도(precision)와 재현율(recall)을 고려한 정확도(F-measure)도는 Simple LSB 기법이 0.923으로 가장 높게 나타났으며 임의 순열 기법과 F3, F4 순서로 낮게 나타났다. True Positive는 스테고 영상을 스테고 영상으로 잘 판별한 경우를 의미하고, True Negative는 스테가노그래피가 적용되지 않은 원본 영상을 원본 영상으로 잘 판별한 경우를 의미한다. 그리고 False Positive는 원본 영상을 스테고 영상으로 잘못 판별한 경우이며, False Negative는 스테고 영상을 원본 영상으로 잘못 판별한 경우를 의미한다. 전반적으로 공간 영역에서 스테가노그래피 부호

화 여부는 제안한 판별 기법으로 잘 판별해 내는 것을 확인할 수 있었다. 하지만 스테고 영상을 판별하는 주파수 영역에서 성능이 상대적으로 낮게 나타나는 이유는 공간 영역에서보다 숨기는 정보의 양이 적어 상대적으로 왜곡이 적기 때문이다. 아래 그림 11은 공간 영역에서의 스테가노그래피 부호화 판별에 실패한 영상(a, b)과 주파수 영역에서 스테고 영상 판별에 실패한 영상(c, d)들이다. 그림 11의 (a)와 (b)는 비정상적인 복호화 영상이지만 영상 내에 특징이 존재하기 때문에 판별에 어려움이 존재했고 그림 11의 (b)와 (c) 영상은 특징이 거의 없어 숨길 수 있는 데이터의 양이 적기 때문에 원 영상 대비 스테고 영상의 왜곡이 적어 주파수 영역에서 이를 판별하는 데 어려움이 있었다.

## V. 결론

이 논문에서는 스테가노그래피의 개념과 공간 방법과 주파수 방법에 대한 영상 스테가노그래피 기술을 소개하였다. 공간 영역에서의 스테가노그래피는 커버 영상 각 픽셀 값의 하위비트 (LSB) 1~3개에 숨기려는 영상의 픽셀 값 8

표 1. 영상 스테가노그래피 판별 결과  
Table 1. Image steganography discrimination result

	True Positive(TP)	True Negative(TN)	False Positive(FP)	False Negative(FN)	Accuracy	F1 score
Simple LSB	476	444	56	24	92.0%	0.923
Random Permutation	456	401	99	44	92.2%	0.881
F3	461	397	103	39	85.8%	0.867
F4	399	419	81	101	81.8%	0.814

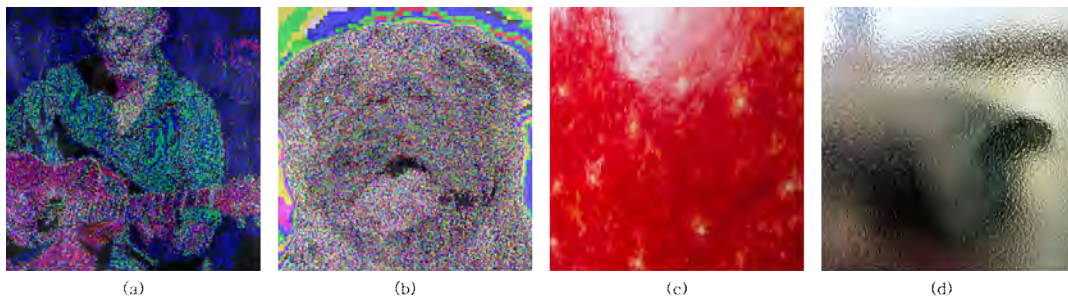


그림 11. 공간 영역에서 잘못 판별한 영상의 예(a, b)와 주파수 영역에서 잘못 판별한 영상의 예(c, d)  
Fig. 11. misidentified images in spatial domain(a, b) and frequency domain(c, d)

비트를 나누어 숨기는 방법으로서 다양한 조작이 결합되어 복화하기 어렵도록 한다. 주파수 영역에서는 주로 JPEG 압축 시 나오는 양자화된 DCT 계수의 값에 1 비트씩 약속한 방법으로 부호화를 수행하여 정보를 숨긴다. 영상 스테가노그래피의 부호화 방법은 매우 많고, 간단한 LSB 조작만으로도 데이터를 완벽하게 은닉할 수 있으므로, 그 부호화 방법을 찾는 것은 거의 불가능하다. 본 논문에서는 스테가노그래피로 부호화가 되었는지를 직접 판별하는 것이 아니라, 일단 임의의 한 방법으로 영상을 복호해보고 그 결과가 올바른 영상으로 복원되었는지를 자동으로 판별하는 기법을 제안하였다. 이를 통하여 무한한 경우를 갖는 영상 스테가노그래피로 부호화된 숨겨진 데이터를 자동으로 검출할 수 있다. 본 논문에서 소개한 영상 스테가노그래피 및 판별 기법은 서로 상충되는 주제이지만, 목적에 맞게 데이터 보안과 복호 기술로서 유용하게 활용될 수 있을 것으로 기대된다.

### 참 고 문 헌 (References)

- [1] Niels Provos and Peter Honeyman, "Hide and seek: An introduction to steganography," IEEE Security & Privacy, Vol.9, No.3, pp.32-44, May-June 2003.
- [2] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," Signal processing, Vol.90, No.3, pp.727-752, March 2010, <https://doi.org/10.1016/j.sigpro.2009.08.010>.
- [3] Mansi S. Subhedar and Vijay H. Mankar, "Current status and key issues in image steganography: A survey," Computer Science Review, Vol.13-14, pp.95-113, November 2014, <https://doi.org/10.1016/j.cosrev.2014.09.001>.
- [4] Chi-Kwang Chan and L. M. Cheng, "Hiding data in images by simple LSB substitution," Pattern Recognition, Vol.37, No.3, pp.469-474, March 2004, <https://doi.org/10.1016/j.patcog.2003.08.007>.
- [5] Ran-Zan Wang, Chi-Fang Lin, and Ja-Chen Lin, "Image hiding by optimal LSB substitution and genetic algorithm," Pattern Recognition, Vol.34, No.3, pp.671-683, March 2001, [https://doi.org/10.1016/S0031-3203\(00\)00015-7](https://doi.org/10.1016/S0031-3203(00)00015-7).
- [6] H. C. Wu, N. I. Wu, C. S. Tsai, and M. S. Hwang, "Image steganographic scheme based on pixel value differencing and LSB replacement methods," IEEE Proc. Of Vision, Image and Signal Processing, Vol.152, No.5, pp.611-615, October 2005, <http://dx.doi.org/10.1049/ip-vis:20059022>.
- [7] Gandharba Swain, "A steganographic method combining LSB substitution and PVD in a block," Procedia Computer Science, Vol.85, pp. 39-44, 2016, <https://doi.org/10.1016/j.procs.2016.05.174>.
- [8] Chanran Kim, et al., "Image Steganography Using Random Permutation and Image Difference," Journal of Broadcast Engineering, 231-234, 2016, June.
- [9] Westfeld, Andreas, "F5-a steganographic algorithm", International workshop on information hiding," International workshop on information hiding, pp. 289-302, April 2001.
- [10] Upham, D. "Steganographic algorithm JSteg," Software available at <http://zooid.org/~paul/crypto/jsteg>, 1993.
- [11] Kim, C. R., Lee, S. H., Lee, J. H., & Park, J. I., "Blind Decoding of Image Steganography Using Entropy Model," Electronics Letters, 2018, doi: 10.1049/el.2017.4276.
- [12] Chanran Kim, et al., "Image Steganography Discrimination based on Local Entropy Distribution," 29th Workshop on Image Processing and Image Understanding, 370-375, 2017, February.
- [13] Jaehoon Lee, et al., "Study on the Correlation between Image Complexity and DCT Coefficients Distribution in Image Steganography," Conference of Institute of Electronics and Information Engineers, 810-813, 2017, June.
- [14] J. Fridrich, M. Goljan, and R. Du, "Detecting LSB steganography in color and gray-scale images," IEEE Multimedia, Oct.-Dec. pp. 22-28, 2001.

---

저 자 소 개

---



**이 재 훈**

- 2017년 : 한양대학교 소프트웨어전공 공학사
- 2017년 ~ 현재 : 한양대학교 컴퓨터소프트웨어학과 공학석사과정
- ORCID : <https://orcid.org/0000-0003-2212-6723>
- 주관심분야 : 영상처리, 컴퓨터비전, 스테가노그래피, 스포츠 트래킹, 딥러닝



**김 찬 란**

- 2014년 : 세종대학교 컴퓨터공학 공학사, 물리학 이학사
- 2018년 : 한양대학교 컴퓨터소프트웨어학과 공학석사
- 2018년 ~ 현재 : 줌인터넷(주) 부설연구소 SA(Search & Artificial Intelligence)팀 담당
- ORCID : <https://orcid.org/0000-0002-0414-4613>
- 주관심분야 : 혼합현실, 스테가노그래피, 컴퓨터비전, 머신러닝, 딥러닝



**이 상 화**

- 1994년 : 서울대학교 전자공학과 공학사
- 1996년 : 서울대학교 전자공학과 공학석사
- 2000년 : 서울대학교 전기컴퓨터공학부 공학박사
- 2005년 ~ 2011년 : 서울대학교 BK21정보기술사업단 BK조교수
- 2011년 ~ 현재 : 서울대학교 연구조교수
- ORCID : <https://orcid.org/0000-0003-0228-0459>
- 주관심분야 : 비디오/영상 신호처리, 패턴인식 및 자동화, 컴퓨터비전



**박 종 일**

- 1987년 : 서울대학교 전자공학과 공학사
- 1989년 : 서울대학교 전자공학과 공학석사
- 1995년 : 서울대학교 전자공학과 공학박사
- 1992년 ~ 1994년 : 일본 NHK 방송기술연구소 객원연구원
- 1995년 ~ 1996년 : 한국방송개발원 선임연구원
- 1996년 ~ 1999년 : 일본 ATR지능영상통신연구소 연구원
- 1999년 ~ 현재 : 한양대학교 공학대학 컴퓨터공학부 교수
- ORCID : <https://orcid.org/0000-0003-1000-4067>
- 주관심분야 : 증강현실, 계산사진학, 3차원 컴퓨터비전, 인간컴퓨터상호작용