

일반논문 (Regular Paper)

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

<https://doi.org/10.5909/JBE.2018.23.4.526>

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

## SPAM 기반 영상 스테그아날리시스의 범용성에 대한 실험적 검증

김재영<sup>a)</sup>, 박한훈<sup>a)\*</sup>, 박종일<sup>b)</sup>

### Experimental Verification of the Versatility of SPAM-based Image Steganalysis

Jaeyoung Kim<sup>a)</sup>, Hanhoon Park<sup>a)\*</sup>, and Jong-Il Park<sup>b)</sup>

#### 요 약

많은 스테가노그래피 알고리즘들이 연구되어왔고 스테가노그래피의 연구로 인해서 스테가노그래피 알고리즘이 적용된 스테고 영상을 검출하기 위한 스테그아날리시스 또한 연구되어왔다. 특히 영상 스테그아날리시스의 경우에 ALE, SPAM, SRMQ와 같은 통계적인 특성에 기반한 수제 특징이 영상으로부터 추출되고 이 특징을 기계학습 알고리즘을 사용하여 스테고 영상의 분류에 사용하였다. 하지만 이러한 연구들은 단지 단일 영상의 크기, 비밀 정보 삽입 비율에 대해서만 고려하였으며 다양한 영상의 크기, 가로세로 비, 비밀 정보의 비율들에 대해서는 고려하지 않았다. 결과적으로 SPAM 특징이 다양한 조건하에서도 범용성있게 사용가능한지에 대한 검증은 이루어지지 않았다. 본 논문에서는 SPAM 기반의 영상 스테그아날리시스를 사용하여 영상의 크기, 가로 세로 비 비밀 정보 삽입 비율의 다양한 조건에 대한 분류율을 분석하여 SPAM 특징이 일관적인 분류 결과를 보여줄 수 있는지 확인해본다.

#### Abstract

Many steganography algorithms have been studied, and steganalysis for detecting stego images which steganography is applied to has also been studied in parallel. Especially, in the case of the image steganalysis, the features such as ALE, SPAM, and SRMQ are extracted from the statistical characteristics of the image, and stego images are classified by learning the classifier using various machine learning algorithms. However, these studies did not consider the effect of image size, aspect ratio, or message-embedding rate, and thus the features might not function normally for images with conditions different from those used in their studies. In this paper, we analyze the classification rate of the SPAM-based image steganalysis against variety image sizes aspect ratios and message-embedding rates and verify its versatility.

Keyword : Image steganography, Steganalysis, SPAM, Statistical feature, Versatility verification

a) 부경대학교 전자공학과(Department of Electronic Engineering, Pukyong National University)

b) 한양대학교 컴퓨터소프트웨어학과(Department of Computer and Software, Hanyang University)

\* Corresponding Author : 박한훈(Hanhoon Park)

E-mail: hanhoon\_park@pknu.ac.kr

Tel: +82-2-629-6225

ORCID:<https://orcid.org/0000-0002-6968-4565>

※ 이 연구는 방위사업청 및 국방과학연구소의 재원에 의해 설립된 신호정보 특화연구센터 사업의 지원을 받아 수행되었음.

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

· Manuscript received May 16, 2018; Revised July 4, 2018; Accepted July 10, 2018.

## I. 서론

스테가노그래피(steganography) 기법이란 비밀 정보를 제 3자가 비밀 정보의 존재 사실 자체를 알지 못 하도록 미디어에 숨기는 것을 의미한다. 이런 스테가노그래피 중에서도 영상 스테가노그래피는 디지털 영상 미디어에 비밀 정보를 숨기는 기법으로 삽입 영역에 따라서 공간 도메인과 DWT<sup>[1][2]</sup>, DCT 도메인과 같은 주파수 도메인으로 나누어지며 공간도메인에서 사용되는 알고리즘으로 LSB<sup>[3]</sup>, PVD<sup>[4][5]</sup>와 같은 많은 방법들이 개발되었다. 이렇게 스테가노그래피에 대한 연구가 진행되면서 스테가노그래피의 개념이 적용된 미디어를 검출하기 위한 스테그아날리시스(steganalysis)의 연구 또한 진행되어왔다. 영상에 대한 스테그아날리시스는 주로 영상의 통계적인 특성을 고려하여 스테가노그래피가 적용된 스테고 영상과 커버 영상사이의 통계적인 요소를 고려하여 수제(handcrafted) 특징(feature)을 추출하고 이를 이용하여 다양한 기계학습 알고리즘을 통해 추출된 특징을 분류함으로써 스테고 영상과 커버 영상을 분별해낸다. 대표적인 방법들로는 ALE<sup>[6]</sup>, SPAM<sup>[7]</sup>, SRMQI<sup>[8]</sup>등이 있다. SPAM특징의 경우 영상으로부터 8개의 방향에 대한 이웃 픽셀들의 픽셀 값 차이를 계산하고 이에 대한 통계적인 값을 이용하여 1000차원 이하의 특징을 생성하며 스테고 영상과 일반 영상의 차이를 SVM 알고리즘을 이용하여 분류기를 생성한다. 반면에 SRMQ특징의 경우에는 SPAM이 가진 특성에 더하여 최대, 최소의 비선형적인 요소를 포함하는 다양한 커널을 사용하여 특징을 추출하며 10000차원 이상의 특징을 이용하여 분류기를 학습시킨다. 이 때 SPAM 특징처럼 SVM 알고리즘을 사용할 경우에 학습 속도의 측면에서 어려움이 발생하기 때문에 앙상블 학습<sup>[9]</sup>를 사용하게 된다. 앙상블 학습의 경우에 SRMQ 내부의 유형에 따라서 추출된 특징을 다수의 FLD (Fisher's linear discriminant) 분류기 들을 랜덤 포레스트 방법으로 취합하여 하나의 통합된 분류기를 만들어낸다.

최근 스테그아날리시스 연구들은 대부분 딥러닝 기술을 이용하여 학습을 통해 최적의 특징을 추출하고, 스테고 영상과 커버 영상을 분류한다<sup>[10][11]</sup>. 이는 기존 수제 특징들에 비해 보다 최적화된 특징을 추출할 수 있지만, 학습 데이터

에 대한 의존성이 크고 특징 추출과 분류를 위한 학습을 병행해야 하기 때문에 방대한 양의 학습 데이터를 필요로 한다. 무엇보다 기존 연구에서 수제 특징들을 사용하더라도 충분히 높은 분류 성능을 가진다는 것이 검증되었다<sup>[7][8]</sup>.

하지만 대부분의 통계적 특성 기반 스테그아날리시스 연구들이 극히 한정된 영상의 크기(512×512)와 삽입 비율에 대한 분류 결과만을 보여주기 때문에 실제로 다양한 영상의 크기나, 영상의 가로세로 비, 비밀 정보 삽입 비율이 적용된 전반적인 영상의 분류에 대한 범용성의 검증은 이루어지지 않았다. 그래서 본 논문에서는 통계적 특성 기반 스테그아날리시스 방법의 범용성을 검증하고자 한다. 구체적으로, 통계적 특징 중에서 SPAM 2차 특징을 사용하여 다양한 영상의 크기와 삽입 비율에 대한 분류 성능의 변화를 분석한다.

## II. 관련 연구

SPAM 특징을 사용한 기존 실험은 0.25bpp(bits per pixel), 0.5bpp의 삽입 비율과 단일 크기(BOSEbase 데이터 셋의 경우 512×512)의 영상 데이터 셋에 대해 SVM을 사용하여 분류기를 학습시키고 분류결과까지 보여주었다<sup>[7]</sup>. 본 논문에서도 기존 연구와 같은 스테가노그래피 기법, 스테그아날리시스 특징과 분류기를 사용하기 때문에, 본 절에서는 n-LSB 스테가노그래피 기법, SPAM 특징과 SVM에 대해 간략히 설명한다.

### 1. n-LSB 스테가노그래피 기법

LSB(least significant bit)는 영상의 픽셀을 비트로 표현할 때 가장 하위 비트를 의미한다. 이러한 LSB는 MSB (most significant bit)와 반대로 가장 작은 가중치를 갖기 때문에 MSB의 변화가 영상에 끼치는 영향에 비해서 LSB의 변화는 극히 미미하다. n-LSB 스테가노그래피 기법들은 이러한 LSB의 특성에 기인하여 다양한 스테가노그래피 기법<sup>[3][14]</sup>이 연구되었다. n-LSB 스테가노그래피 기법의 흐름은 다음과 같다.

삽입 알고리즘

- 1) 삽입 하고자하는 비밀 정보를 비트로 변환.
- 2) 삽입 순서에 따라서 픽셀의 하위  $n$ 개의 LSB를 비밀 정보로 대체.
- 3) 비밀 정보의 삽입이 완료될 경우 종료.

추출 알고리즘

- 1) 삽입 순서에 따라서 픽셀의 하위  $n$ 개의 LSB를 추출.
- 2) 비밀 정보의 추출이 완료될 경우 종료.
- 3) 추출된 비밀 정보의 비트를 해당 미디어로 복원.

비밀 정보의 삽입이 픽셀의 비트에 적용되므로 삽입 가능한 비밀 정보의 크기는 영상의 크기와 LSB의 수  $n$ 에 비례하며  $n$ 의 수가 증가하면 그에 비례하여 커버 영상과의 스테고 영상의 차이가 증가한다. 이러한 커버 영상과 스테고 영상의 차이는 PSNR(peak signal-to-noise ratio)로 정의하며 삽입 비밀 정보의 양과 PSNR을 균형있게 조정하여 비밀 정보를 삽입한다. 본 논문에서는 실험에 사용될 스테고 영상을 생성하기 위해 1-LSB를 사용한다.

2. SPAM(Subtractive Pixel Adjacency Matrix) 특징

SPAM 특징은 이웃한 픽셀 값들이 유사한 값을 갖는 영상의 특성을 이용한 특징 중 하나이다. 영상은 부분적으로 코너, 엣지와 같은 고주파수 성분을 갖고 그 외 대부분의 전반적인 영역에서 저주파수 성분을 갖게 된다. 이러한 특성으로 외부적인 요인이 없을 경우에 이웃한 픽셀 값의 차는 0이 될 확률이 가장 높으며 그림 1의 실선과 유사한 확률 분포를 갖게 된다. 그러나, 영상에 인위적인 비밀 정보가 삽입되면 이러한 통계적인 특성은 그림 1의 점선처럼 변형되는데 SPAM 특징은 이렇나 통계적인 특성의 변형을 정량적으로 표현한다. Pevny등은 총 8가지 방향 ‘→, ←, ↓, ↑, ↘, ↙, ↗, ↖’에 대해서 고려한 1차, 2차 SPAM 특징을 제안했다<sup>[7]</sup>. 픽셀값을  $I_{i,j}$ , 픽셀값 차이를  $D_{i,j}^{\rightarrow} = I_{i,j} - I_{i,j+1}$  라고 할 때,  $D_{i,j}^{\rightarrow} \in \{-T, \dots, T\}$ 를 만족하는 픽셀들에 대한 확률을 특징으로 추출하였다. 1차 특징의 ‘→’방향에 대해서 생각해보면 영상 전체를 계산했을 때

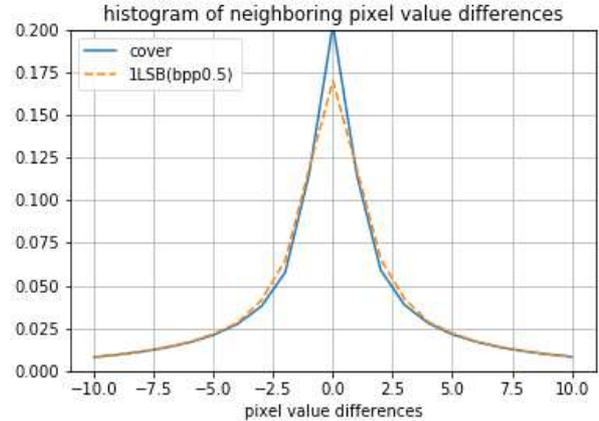


그림 1. BOSSbase 데이터 셋의 1000장의 커버 영상과 1-LSB bpp 0.5 삽입 스테고 영상의 이웃 픽셀 사이의 값 차이 분포<sup>[15]</sup>  
 Fig. 1. Distribution of neighboring pixels differences between cover images and 1-LSB bpp 0.5 stego images in 1000 BOSSbase datasets<sup>[15]</sup>

$D_{i,j}^{\rightarrow} \in \{-T, \dots, T\}$ 의 조건을 만족하는 수가  $P$ 라 하고  $D_{i,j}^{\rightarrow}$ 를 만족하면서  $D_{i,j+1}^{\rightarrow} \in \{-T, \dots, T\}$ 를 만족하는 모든 수가  $F$ 라고 하면 ‘→’방향에 대한 특징은  $M_{u,v,w}^{\rightarrow} = P(D_{i,j+1}^{\rightarrow} = u | D_{i,j}^{\rightarrow} = v) = F/P$ 가 되며  $(2T+1)^2$ 의 차원을 갖는다(식 1 참조). 2차 특징은 1차 특징에서  $D_{i,j+2}^{\rightarrow}$ 을 추가하여 동일한 방법으로 특징을 생성하며  $(2T+1)^3$ 의 차원을 갖는다(식 2 참조).

$$M_{u,v}^{\rightarrow} = P(D_{i,j+1}^{\rightarrow} = u | D_{i,j}^{\rightarrow} = v) \tag{1}$$

$$M_{u,v,w}^{\rightarrow} = P(D_{i,j+2}^{\rightarrow} = u | D_{i,j+1}^{\rightarrow} = v, D_{i,j}^{\rightarrow} = w) \tag{2}$$

$$F_{1,\dots,k} = \frac{1}{4} [M^{\rightarrow} + M^{\leftarrow} + M^{\downarrow} + M^{\uparrow}] \tag{3}$$

$$F_{k+1,\dots,2k} = \frac{1}{4} [M^{\searrow} + M^{\swarrow} + M^{\nearrow} + M^{\nwarrow}] \tag{4}$$

단일 방향에 대해서만 특징을 추출하지 않고 2차원으로 구성된 영상의 특성을 최대한 많이 추출하기 위해서 SPAM 특징은 ‘→, ←, ↓, ↑, ↘, ↙, ↗, ↖’ 8가지 방향에 대해서 특징을 각 추출하며 추출된 특징을 다음의 ‘→, ←, ↓, ↑’, ‘↘, ↙, ↗, ↖’와 같은 두 가지 부분 집합으로 분리하고 각 부분 집합의 평균 특징을 추가하여 최종 특징을 추출한다

(식 3, 4 참조). 최종 추출된 특징은 1차 특징과 2차 특징의 경우 각  $2(2T+1)^2$ ,  $2(2T+1)^3$  차원을 갖는다.

### 3. SVM(Support Vector Machine)

SVM<sup>[12][13]</sup>은 분류를 위한 기계학습 알고리즘 중 하나로 주어진 특징으로 클래스를 분류할 수 있는 최적의 초평면을 구하는 알고리즘이다.  $n$  차원의 특징이 주어질 때 초평면과 특징의 거리는 식 5와 같이 주어지며 최적의 초평면을 구하기 위해서 초평면으로부터 가장 거리가 가까운 특징 즉 서포트 벡터와 초평면 사이의 거리를 최대화하여 목적 함수를 구한다.

$$d = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n + b}{\sqrt{w_1^2 + w_2^2 + \dots + w_n^2}} \quad (5)$$

하지만 대부분의 문제들은 비선형적인 특징 분포를 갖거나 예외들(outlier)이 존재하여 선형 SVM만을 적용할 경우에 제한적이므로 주로 비선형 SVM을 사용하게 된다. 비선형 SVM의 경우에는 소프트 마진 SVM과 커널을 이용한 SVM이 있다. 소프트 마진 SVM은 각 클래스의 특징들의 대다수는 클래스의 중심에 분포하며 그 중 일부의 특징들이 오 분류되는 경우에 적용할 수 있으며 이러한 오 분류 특징들에 감점요소를 부여하여 학습을 시킨다. 커널을 이용한 SVM은 XOR의 분류문제(그림 2참조)와 같이 특징의

차원에 존재하는 단일 초평면으로는 분류가 불가능한 경우에 동차다항식(Homogeneous polynomial), 다항식 커널(Polynomial kernel), RBF(radial basis function)등의 커널을 사용하여 분류에 사용하는 특징을 고차원 공간으로 사상시켜서 초평면을 구한다. 3절에서 사용하는 SVM은 기존 논문과 동일하게 RBF 커널(식 6 참조)을 사용한 SVM을 학습시킨다.

$$k(X_i, X_j) = e^{-\gamma \|X_i - X_j\|^2} \quad (6)$$

### III. 실험 방법

본 논문에서는 단일 알고리즘을 사용할 때 영상의 크기나 중횡비 그리고 삽입 비율이 달라지면 분류결과에 어떠한 영향을 미치는지 확인한다. 기존의 영상 스테그아날리시스에서 사용한 수제 특징들의 경우에 영상 내부의 인접한 화소들의 통계적인 특성을 고려하여 설계가 되었다(그림 1 참조). 그렇기 때문에 영상의 중횡비나 영상의 크기가 추출되는 특징에 영향을 미칠 수 있음을 의미한다. 실험에서 사용한 SPAM 특징의 경우에도 수평, 수직과 대각 방향의 이웃한 픽셀값의 차이를 고려하기 때문에 중횡비가 추출되는 특징에 영향을 주며 영상의 크기가 줄어들 경우 또한 마찬가지로 영향을 끼칠 것이다. 그러므로 다양한 영상 크기 및 픽셀 변화에 대해서 수제 특징들 중 하나인 SPAM 특징의 성능을 검증해볼 필요가 있다. 이를 위해서 첫 번째로 중횡비를 고정시키고 영상의 크기를 달리하여 학습시킨 분류기의 분류결과를 비교하고 두 번째로 영상의 크기를 고정시키고 중횡비를 달리하여 학습시킨 분류기의 분류결과를 비교한다. 실험에서는 분류결과 분석을 용이하게 하기 위해서 공간도메인에서 가장 기본적인 1-LSB 스테가노그래피 기법을 사용하였고 BOSEbase 영상 셋을 사용하여 다양한 영상의 크기와 삽입 비율로 생성한 스테고 영상으로부터 추출한 SPAM 2차 특징의 분류 결과를 분석한다.

#### 1. 데이터 셋 생성

실험은 BOSSbase 데이터 셋  $512 \times 512$  영상 10000장을 원 영상,  $256 \times 256$ ,  $512 \times 128$ ,  $128 \times 512$ ,  $128 \times 128$

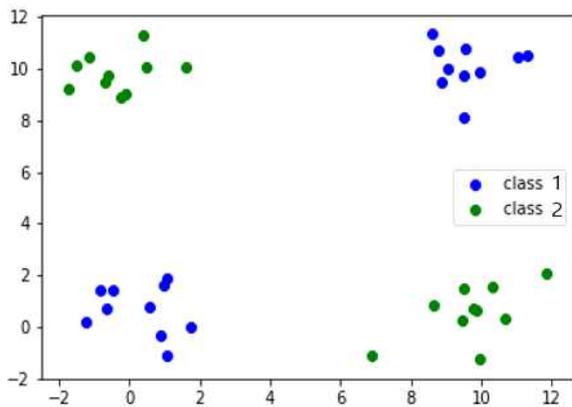


그림 2. 2차원 특징의 XOR 분포  
 Fig. 2. XOR distribution of 2 dimensional features

5개의 크기로 1, 4, 16분할하며 분할한 영상을 1-LSB 삽입 영상 스테가노그래피 알고리즘을 이용하여 영상 크기에 대해서 5%~100% 삽입 비율로 5%의 간격으로 삽입한 스테고 영상 데이터 셋을 생성한다. 여기서 삽입 비율은 전체 영상의 픽셀에 대해서 비트의 변화 유무를 나타내며 1-LSB 삽입 방법을 사용하기 때문에 1 비트의 대체가 발생한다. 그래서 삽입 비율은  $\text{bpp}(\text{bits per pixel})$ 로 나타낸다. 이 때 데이터의 삽입은 픽셀마다 독립적인 확률로 영상 전체에 동일한 비율로 적용되어 데이터 셋을 생성한다. 데이터 셋 생성을 위해서 다음의 1-LSB 방법을 사용하였다.  $C(y,x)$ 를  $(y,x)$ 에서의 픽셀 값이라고 할 때 생성되는 스테고 영상의 픽셀 값은  $S(y,x)$ 로 대체된다. 이 때 두 픽셀 값 사이의 관계는 삽입되는 정보에 따라서 달라지는데 삽입되는 비밀 정보가  $b = \{0, 1\}$  일 때 생성되는 스테고 영상의 픽셀 값은 다음  $S(y,x) = 2 \times \lfloor C(y,x)/2 \rfloor + b$  와 같이 변한다. 삽입되는 비밀 정보의 값에 따라서 영상의 픽셀 값이 변할 확률은 50%이므로 생성된 스테고 영상에서 대체된 픽셀의 비율은  $\text{bpp}$ 에 따라서 2.5%~50%이다.

## 2. 분류기 생성

RBF커널을 이용한 SVM을 사용하며 세부 파라미터  $C=10000$ ,  $\gamma = \frac{1}{686}$ 를 갖는다. 학습에 사용하는 트레이닝 셋은 다음을 사용하여 20개의 분류기를 생성한다.

- 1) 5%, 10%, 15%, 20%의 삽입 비율을 갖는 영상 셋
- 2) 3.1절에서 생성한 15가지 데이터 셋의 90%사용
  - 분할한 영상의 평균 특징 1~9000번 영상 사용

생성된 분류기는 다음과 같이 나타낸다.  $clf_{d,r}$   $d$  분할한 영상,  $r$ 은 중횡비를 나타낸다. 예를들어  $512 \times 512$  영상으로 학습시킨 경우  $clf_{1,1}$ 이며  $128 \times 512$  영상으로 학습시킨 분류기의 경우  $clf_{4,1.4}$ 로 나타낸다. 각 크기가 다른 영상으로부터 추출된 테스트 셋의 2차 SPAM 특징을 분류하는데 사용된다.

## 3. 분류

생성한 전체 분류기 20개에 대해서 분류를 한다. 분류에 사용하는 테스트 셋은 트레이닝 셋에서 사용하지 않은 9001~10000번 영상을 삽입률과 영상 크기에 따라서 생성한 데이터 셋으로부터 추출한 SPAM 2차 특징을 사용한다. 영상의 크기에 따라서  $512 \times 512$  크기의 영상은 1000개,  $256 \times 256$  크기의 영상은 4000개,  $128 \times 128$  영상은 16000개의 SPAM 특징을 갖는다. 영상의 크기에 따라서 생성된 SPAM 특징은 50%는 스테고 영상 50%는 커버 영상을 사용해서 분류에 사용한다.

## IV. 실험 및 결과

실험의 결과는 오검출률에 대해서 고려하며 분류에 사용한 데이터 셋의 스테고 영상과 커버 영상의 비율이 1대1이므로 삽입 비율과 영상의 크기에 대한 영상 전체의 False Positive와 False Negative의 비율의 합을 나타낸다. 전체 분류 결과는 표 4와 같으며 세부적인 분석 결과는 다음 절에서 나타낸다.

### 1. 영상 크기별 분류 결과

트레이닝 셋의 각  $\text{bpp}$  0.05, 0.10, 0.15, 0.20에 따라서 학습된 분류기를 영상의 크기가 다른 테스트 셋을 사용하여 분류를 하였다. 테스트 셋의 분류결과는 표 1과 같고 표 1의 평균값을 그래프로 나타내는 것은 그림 3과 같으며  $\text{bpp}$  0.05 마다 분류한 결과를 도시하였다. 분류에 사용한 영상은 중횡비를 1대1로 유지하고 영상의 크기만을 달리하여 영상 크기를 1/4로 축소시켜서 사용하였다. 실험에서는  $256 \times 256$  크기의 영상을  $\text{bpp}$  0.15의 삽입 비율로 삽입시킨 영상으로 학습시킨 분류기(그림 3, 표 1  $\text{bpp}$  0.15 참조) 오검출율이 28.74%로 가장 뛰어난 성능을 보여주었다. 범용적인 측면에서 볼 때 학습에 사용한 트레이닝 셋의 스테고 영상의  $\text{bpp}$ 가 0.15이상일 때 좋았으며 특히 영상의 크기

1)  $512 \times 512$ ,  $256 \times 256$ ,  $512 \times 128$ ,  $128 \times 512$ ,  $128 \times 128$ 의 크기로 분할한 데이터 셋 각 10000, 40000, 40000, 40000, 160000장이다.

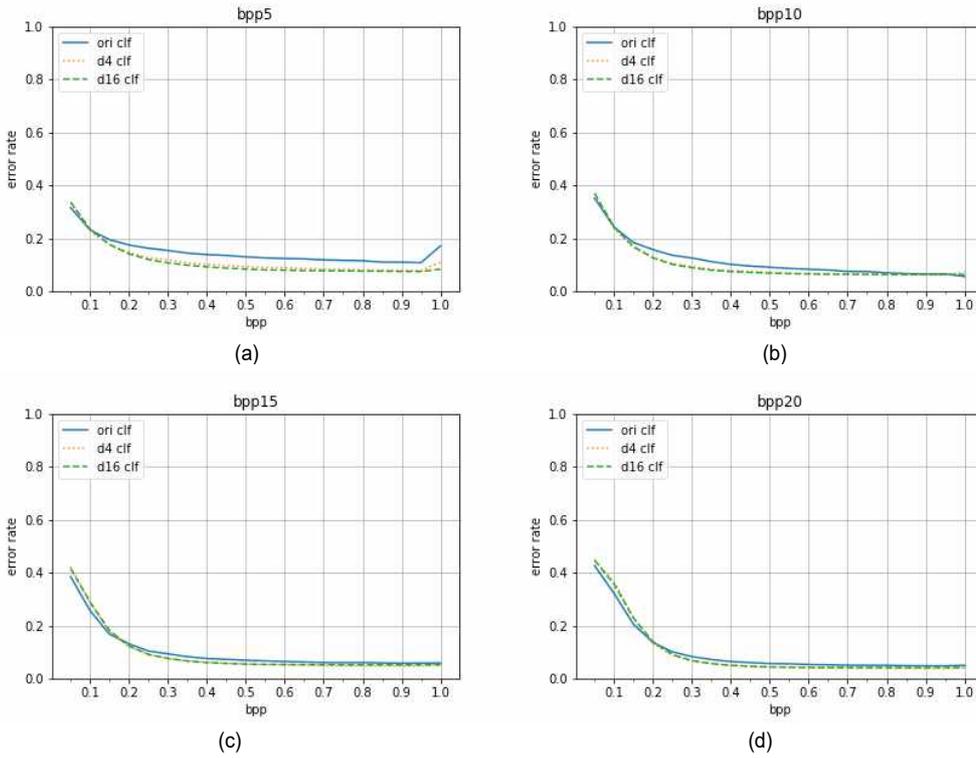


그림 3. 삽입 비율에 대한 각 분류기의 분류 결과  
 Fig. 3. Classification result of each classifier about embedding ratios

표 1. 학습에 사용한 스테그 영상의 bpp와 영상 크기별 분류 결과  
 Table 1. The classification result of the stego images used for learning by bpp and image size

learning bpp	classifier	error rate for each image sizes[%]			
		512 × 512	256 × 256	128 × 128	average
0.05	$df_{1,1}$	7.55	13.94	23.48	14.99
	$df_{4,1}$	7.26	11.16	17.28	11.90
	$df_{16,1}$	8.37	11.07	14.33	11.26
0.10	$df_{1,1}$	5.58	10.38	18.73	11.56
	$df_{4,1}$	6.28	9.30	15.09	10.22
	$df_{16,1}$	7.56	9.94	12.96	10.15
0.15	$df_{1,1}$	4.78	8.95	16.97	10.23
	$df_{4,1}$	6.02	8.92	13.80	9.58
	$df_{16,1}$	7.53	9.32	12.20	9.68
0.20	$df_{1,1}$	5.89	8.31	16.28	10.16
	$df_{4,1}$	6.90	8.5	13.46	9.62
	$df_{16,1}$	7.69	9.44	11.99	9.71

가 중간값을 갖는  $256 \times 256$ 으로 학습시킨 분류기의 성능이 좋았다. 이 결과로 볼 때 영상의 크기가 추출되는 SPAM

특징에 영향 미치는 것으로 보이며 이에 대한 분석을 다방면으로 해보았다. 첫 번째로 학습에 사용한 트레이닝 셋의

커버 영상과 스테고 영상의 평균 특징  $f_{tr} = \frac{2}{N} \sum_{n=1}^{\frac{N}{2}} f_n$  과 BOSSbase의 셋 10000장에서 생성한 스테고 영상과 커버 영상의 특징  $f_{BOSS} = \frac{1}{10000} \sum_{n=1}^{10000} f_n$  의 유클리디언 거리를 측정해 보았으나 뚜렷한 거리차는 없었다. 두 번째로는 [7]에서 10000차원 이상의 고차원 특징을 스테그아날리시스에 사용할 때 추출되는 특징의 빈공간수 즉 특징의 값이 0이 되어서 분류에 영향을 주지 못하는 공간이 증가하는 것에 대해서 고려하는 것에 착안하여 SPAM 특징의 분석에 사용해 보았다. 영상의 크기가 줄어들면서 SPAM 2차 특징으로 추출될 수 있는 표본의 수가 크게 감소한다는 점으로부터 실험에서 분류기 학습에 사용한 트레이닝 셋으로부터 추출된 686차원의 특징에서 영상의 크기에 따라서 빈공간이 얼

마나 생성되는지에 대해서 확인해 본 결과는 표 2와 같다. 결과에서 볼 수 있는 것처럼 영상의 크기가  $512 \times 512$  인 경우에는 빈 공간의 수가 거의 0에 수렴한다. 하지만 영상이 분할되는 경우에 빈 공간의 수가 증가하며  $128 \times 128$  크기로 16분할 된 커버 영상과 스테고 영상에서 추출된 특징의 경우에 686차원 중에서 약 8%, 5%정도로 크게 나타났다. 결과적으로 특징의 빈 공간수가 분류의 실패에 영향을 미치며 실험에 사용한 분류기들 중에서 빈공간의 수가 중간 정도를 갖는  $df_{4,1}$  학습기가 테스트 셋의 평균 오검출율이 가장 낮게 나온 것으로 보인다. 마찬가지로 오차율이 영상의 크기와 반비례하는 것을 볼 수 있는데 이것 또한 영상의 크기가 갖는 특징의 빈공간 수와 영상의 크기가 반비례하여 분류의 결과에 영향을 미친 것으로 보인다.

표 2. 영상 크기별 686차원 SPAM 2차 특징의 빈 공간 수  
Table 2. Number of empty spaces of 686 second order dimensional SPAM feature by image size

images size	bpp				
	0	0.05	0.10	0.15	0.20
$512 \times 512$	0.70	0.04	0.03	0.03	0.02
$256 \times 256$	8.78	2.71	2.52	2.36	2.25
$128 \times 512$	6.95	0.93	0.83	0.76	0.71
$512 \times 128$	13.90	5.83	5.42	5.12	4.92
$128 \times 128$	52.86	32.24	31.10	30.24	29.56
average	6.07	1.91	1.78	1.68	1.62

표 3. 영상의 종횡비별 분류 분류 결과  
Table 3. Classification result of aspect ratio

learning bpp	classifier	error rate for each image sizes[%]			
		$256 \times 256$	$128 \times 512$	$512 \times 128$	average
0.05	$df_{4,1}$	11.16	10.57	11.60	11.11
	$df_{4,1:4}$	11.84	10.46	12.16	11.49
	$df_{4,4:1}$	11.22	10.36	11.48	11.02
0.10	$df_{4,1}$	9.30	8.09	9.62	9.00
	$df_{4,1:4}$	9.42	8.00	9.81	9.08
	$df_{4,4:1}$	9.55	8.16	9.69	9.130
0.15	$df_{4,1}$	8.92	7.51	8.86	8.43
	$df_{4,1:4}$	8.58	7.58	8.93	8.36
	$df_{4,4:1}$	8.74	7.65	9.06	8.48
0.20	$df_{4,1}$	8.5	7.98	8.75	8.41
	$df_{4,1:4}$	8.58	7.75	8.63	8.32
	$df_{4,4:1}$	8.86	8.00	9.05	8.64

표 4. 전체 분류 결과

Table 4. Total classification result

learning bpp	classifier	error rate for each image sizes[%]					average
		512 × 512	256 × 256	128 × 512	512 × 128	128 × 128	
0.05	$df_{1,1}$	7.55	13.94	12.45	14.38	23.48	14.36
	$df_{4,1}$	7.26	11.16	10.57	11.60	17.28	11.57
	$df_{4,1:4}$	7.36	11.84	10.46	12.16	18.86	12.14
	$df_{4,4:1}$	7.37	11.22	10.36	11.48	17.80	11.65
	$df_{16,1}$	8.37	11.07	10.14	11.16	14.33	11.01
0.10	$df_{1,1}$	5.58	10.38	8.60	11.20	18.73	10.90
	$df_{4,1}$	6.28	9.30	8.09	9.62	15.09	9.68
	$df_{4,1:4}$	6.05	9.42	8.00	9.81	15.46	9.75
	$df_{4,4:1}$	5.92	9.55	8.16	9.69	15.33	9.73
	$df_{16,1}$	7.56	9.94	8.67	10.08	12.96	9.84
0.15	$df_{1,1}$	4.78	8.95	7.83	9.75	16.97	9.66
	$df_{4,1}$	6.02	8.92	7.51	8.86	13.80	9.02
	$df_{4,1:4}$	5.84	8.58	7.58	8.93	14.07	9.00
	$df_{4,4:1}$	6.19	8.74	7.65	9.06	13.93	9.11
	$df_{16,1}$	7.53	9.32	8.45	9.43	12.20	9.39
0.20	$df_{1,1}$	5.89	8.31	7.48	9.06	16.28	9.40
	$df_{4,1}$	6.90	8.5	7.98	8.75	13.46	9.12
	$df_{4,1:4}$	6.87	8.58	7.75	8.63	13.53	9.07
	$df_{4,4:1}$	6.89	8.86	8.00	9.05	13.43	9.25
	$df_{16,1}$	7.69	9.44	8.56	9.50	11.99	9.44

## 2. 영상의 중횡비별 분류 결과

영상의 통계적인 특성을 이용한 특징의 경우에 영상의 픽셀 수는 중요한 요소이며 SPAM 특징은 이러한 통계적인 특성을 수직, 수평, 대각등의 다양한 방향에 따라서 세분화하여 특징을 획득하였다. 이러한 요소로 인하여 특징 중횡비로 학습을 한 분류기를 사용할 때 중횡비가 다른 영상을 분류할 때의 결과를 비교해 보았다. 중횡비에 대한 분류 결과는 표 3과 같다. 표 3의 결과에서 볼 수 있는 것처럼 bpp와 영상의 중횡비에 따라서 다르게 학습시킨 모든 분류기가 중횡비가 1:4인 경우에 가장 좋은 성능을 보였다. 이는 앞서 4.1절에서와 마찬가지로 표 2의 특징의 빈공간의 수가 미친 영향으로 보인다. 특징의 빈공간 수가 적을수록 좀 더 정확한 분류를 할 수 있었다. 하지만 1:4의 중횡비와 4:1 중횡비의 영상이 특징의 빈공간 수가 차이가 나는 것은 BOSSbase 영상 셋이 횡으로 뺀 배경을 갖는 영상을 많이 포함하고 있는 것이 영향을 미친 것으로 보인다. 하지만 결

과적으로 4.1절에서 영상의 크기에 대한 분류결과의 차이와 비교할 때 가장 큰 차이가 0.05  $df_{4,1:4}$ ,  $df_{4,4:1}$  분류기의 경우에 1.4%정도로 작은 오차를 보여주었다.

## V. 결론

본 논문은 [7]~[9]와 같은 스테그아날리시스 논문들의 실험이 단일 영상의 크기와 단일 삽입 비율에 대해서만 실험하였기 때문에 학습시킨 조건 외의 다양한 입력에 대해서 분류결과의 변화를 분석하기 위한 실험을 진행하였다. 그리고 스테그아날리시스들 중에서도 통계적인 수제 특징중 하나인 SPAM 2차 특징이 범용적으로 사용하기에 적절한지 판단하기 위해서 다양한 영상의 크기와 삽입 비율에 대해서 실험을 진행하였다. 실험 결과 대부분의 경우에서 오검출율이 50%이하로 나타났지만 영상의 크기가 다른 경우에 유의미한 오차율의 차이를 보였다(표1, 4참조). 작게는

3% 많게는 10%정도의 오차율의 차이가 발생하였고 영상의 중횡비에 따라서는 2%이내의 차이를 보였다(표3, 4참조). 결과적으로 스테고 영상과 커버 영상의 분류에 있어서 특징을 추출할 수 있는 표본의 수인 영상의 크기가 더 중요하며 영상의 크기가 같을 경우에 영상의 중횡비는 큰 차이를 보이지 않았다. 실제로 영상을 획득하는 카메라의 경우 주로 16:9, 4:3, 1:1, 2:1 정도의 중횡비를 갖기 때문에 실험에서 사용한 4:1 정도의 비율에 대한 결과로 볼 때 범용적인 측면에서도 중횡비는 분류기의 학습에 많은 영향을 끼치지 않을 것으로 보인다. 또한 영상의 크기와 오차율의 차이가 추출된 특징의 빈공간의 발생정도와 비례하였는데 이러한 관계가 정확할 경우에 영상의 크기가  $512 \times 512$  보다 큰 영상의 경우에 특징의 빈공간 수가  $512 \times 512$  크기의 영상보다 더 적을 것이고 이는 오차율의 차이가 거의 없을 수도 있다는 것을 의미한다. 즉 통계적인 특성을 사용한 특징의 모델에 따라서 절대적인 영상의 크기 이하일 경우에 오차율이 급증할 수 있다는 것이다. 실험결과를 볼 때 범용성 있는 스테그아날리시스 분류기를 학습시키기 위해서는 추출되는 특징의 빈공간의 수가 중간값을 갖는 영상의 크기로 학습시킬 경우에 전체 영상의 크기에 대해서 좋은 결과를 보일 수 있을 것으로 보인다. 추가적인 연구로는 앞서 말한 통계적인 특징들이 영상의 크기가 증가할 경우에 빈공간의 수가 0에 수렴할 것인데 이 때 영상의 크기와 오차율의 관계에 대한 연구도 필요할 것이다.

### 참 고 문 헌 (References)

- [1] J. Kim and H. Park, "A statistical approach for improving the embedding capacity of block matching based image steganography, Journal of Broadcast Engineering, Vol. 22, No. 5, 2017.
- [2] W. Sohn and L. N. T. dung, "A blind watermarking scheme using singular vector based on DWT/RDWT/SVD," Journal of Broadcast Engineering, Vol. 21, No. 2, pp. 149-156, 2016.
- [3] J. Seon-Su, "A study of optimal image steganography based on LSB techniques," Journal of the Korea Industrial Information Systems Research, Vol. 20, No. 3, pp. 29-36, 2015.
- [4] D. Wu and W. Tsai, "A steganographic method for images by pixel-value differencing," Pattern Recognition Letters, Vol. 24, pp. 1613-1626, 2003.
- [5] J. Kim and H. Park, "Image steganography using layered pixel-value differencing," Journal of Broadcasting Engineering, Vol. 22, No. 3, 2017.
- [6] G. Cancelli, G. Doerr, I. J. Cox and M. Barni, "Detection of  $\pm 1$  LSB steganography based on the amplitude of histogram local extrema," ICIP, 2008.
- [7] T. Pevny, P. Bas and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," IEEE Transactions on Information Forensics and Security, Vol. 5, No. 2, pp. 215-224, 2010.
- [8] J. Fridrich, Member, IEEE and J. Kodovsky, "Rich models for steganalysis of digital images," IEEE Transactions on Information Forensics and Security, Vol. 7, No. 3, pp. 868-882, 2012.
- [9] J. Kodovsky, J. Fridrich, Member, IEEE and V. Holub, "Ensemble classifiers for steganalysis of digital media," IEEE Transactions on Information Forensics and Security, Vol. 7, No. 2, pp. 432-444, 2012.
- [10] J. Ye, J. Ni, Member, IEEE and Y. Yi, "Deep learning hierarchical representations for image steganalysis," IEEE Transactions on Information Forensics and Security, Vol. 12, No. 11, pp. 2545-2557, 2017.
- [11] L. Pibre, J. Pasquet, D. ienco and M. Chaumont, "Deep learning is a good steganalysis tool when embedding key is reused for different images even if there is a cover source-mismatch," Society for Imagng Science and Technology, 2016.
- [12] S. Kim and D. S. Han, "Real time traffic light detection algorithm based on color map and multilayer HOG-SVM," Journal of Broadcast Engineering, Vol. 22, No. 1, pp. 62-69, 2017.
- [13] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, Vol. 20, pp. 273-297, 1995.
- [14] C. Chan and L. M. Cheng, "Hiding data in images by simple LSB substitution," Pattern Recognition, Vol. 37, pp. 469-474, 2004.
- [15] P. Bas, T. Filler and T. Pevny, "'Break our steganographic system': The ins and outs of organizing boss," in Proc. Int. Workshop Inf. Hiding, pp. 59-70, 2011.

---

저 자 소 개

---



**김 재 영**

- 2012년 : 부경대학교 전자컴퓨터정보통신공학부
- 2016년 ~ 현재 : 부경대학교 일반대학원 석사과정
- ORCID : <http://orcid.org/0000-0003-2348-3626>
- 주관심분야 : 스테가노그래피, 기계 학습 등



**박 한 훈**

- 2000년 : 한양대학교 전자통신전파공학과 공학사
- 2002년 : 한양대학교 대학원 전자통신전파공학과 공학석사
- 2007년 : 한양대학교 대학원 전자통신전파공학과 공학박사
- 2008년 ~ 2011년 : NHK방송기술연구소 박사후연구원
- 2012년 ~ 현재 : 부경대학교 전자공학과 부교수
- ORCID : <http://orcid.org/0000-0002-6968-4565>
- 주관심분야 : 증강현실, 인간컴퓨터상호작용, 3차원 영상처리/비전 등



**박 종 일**

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