Journal of The Institute of Electronics and Information Engineers Vol.52, NO.8, August 2015 http://dx.doi.org/10.5573/ieie.2015.52.8.097
ISSN 2287-5026(Print) / ISSN 2288-159X(Online)

논문 2015-52-8-11

DCT 계수의 마코프 특징을 이용한 내용 적응적 스테가노그래피의 스테그분석

(Steganalysis of Content-Adaptive Steganography using Markov Features for DCT Coefficients)

박 태 희*, 한 종 구**, 엄 일 규***

(Tae Hee Park, Jong Goo Han, and Il Kyu Eom©)

요 약

내용 적응적 스테가노그래피는 복잡한 텍스쳐 또는 잡음 영역과 같이 통계적 모델로는 기술하기 어려운 영역에 비밀 메시지를 은닉한다. 이러한 메시지를 검출하기 위해서는 인접 화소간의 국부적인 의존성을 정교하게 모델링해야 하기 때문에 종종 고차원의 특징벡터 추출이 필요하다. 이러한 스테그분석 방법은 계산량이 많을 뿐만 아니라 비밀 메시지의 검출 정확도가 은닉 영역과 사용된 왜곡 척도에 의존한다는 문제점을 가진다. 본 논문에서는 적은 수의 특징 벡터를 이용하여 비밀 메시지의 검출율을 높일 수 있는 개선된 내용 적응적 스테가노그래피의 스테그분석 방법을 제안하고자 한다. 먼저 이산 코사인 변환 계수의 차이를 이용한 특징이 내용 적응적 스테가노그래피의 분석에 유용함을 보이고, 이에 대한 1차 마코프 확률을 특징으로 사용하는 방법을 제시한다. 추출된 특징 벡터는 앙상블 분류기로 입력되어 커버 영상과 스테고 영상을 분류하기 위해 학습된다. 실험 결과 내용 기반 적응적 스테고 영상들에 대해 적은 수의 특징 벡터를 사용함에도 불구하고 기존의 방법에 비해 검출율과 정확도가 우수함을 확인할 수 있었다.

Abstract

Content-adaptive steganography methods embed secret messages in hard-to-model regions of covers such as complicated texture or noisy area. Content-adaptive steganalysis methods often need high dimensional features to capture more subtle relationships of local dependencies among adjacent pixels. However, these methods require many computational complexity and depend on the location of hidden message and the exploited distortion metrics. In this paper, we propose an improved steganalysis method for content-adaptive steganography to enhance detection rate with small number features. We first show that the features form the difference between DCT coefficients are useful for analyzing the content-adaptive steganography methods, and present feature extraction mehtod using first-order Markov probability for the the difference between DCT coefficients. The extracted features are used as input of ensemble classifier. Experimental results show that the proposed method outperforms previous schemes in terms of detection rates and accuracy in spite of a small number features in various content-adaptive stego images.

Keywords : content-adaptive steganography, DCT coefficient, Markov chain, difference image, ensemble classifier

* 정회원, 동명대학교 메카트로닉스공학과
(Dept. Mechatronics Eng., TongMyong University)
** 학생회원, *** 정회원, 부산대학교 전자공학과
(Dept. Electronics Eng., Pusan National University)
© Corresponding Author(E-mail: ikeom@pusan.ac.kr)
※ 이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2012R1A1A2042034)
Received ; March 2, 2015 Revised ; June 10, 2015 Accepted ; July 25, 2015

Ⅰ. 서 론

오늘날 디지털 영상은 인터넷을 매개로 하여 빈번한 사용과 전송이 이루어지고 있다. 따라서 디지털 영상은 비밀 메시지를 전송하기 위한 중요한 수단이 되고 있다[1]. 스테가노그래피(steganography)는 디지털 영상 내에 비밀 메시지를 은닉함으로써 인증되지 않은 수신자로부터 메시지의 내용 및 존재 여부를 감추는 대표적인 정보은닉 기법이다[2]. 따라서 스테가노그래피는 개인의 사생활이나 기업의 경영 활동 그리고 국가의 안전 보장을 위해 유용하게 사용될 수 있다. 그러나 최근에는 상업 스파이에 의한 기술 문서의 유출, 스팸 메일속의 악성 코드에 의한 PC의 감염 등 스테가노그래피 기술에 의한 기밀문서 및 개인정보의 유출사례가 증가하고 있다. 이에 따라 기밀문서 및 개인정보의 악의적인 유출을 막기 위한 수단으로 비밀 메시지의 내용을 추출하거나 존재 유무를 감지하고자 하는 스테그분석(steganalysis)에 관한 연구가 활발히 진행되고 있다[3].

스테그분석은 시각적으로는 비밀 메시지의 존재를 거의 감지할 수 없으나 비밀 메시지의 삽입으로 인해 공간영역 또는 변환영역에서 화소 및 변환계수들의 통계적 특성과 의존성이 변하는 원리를 이용한다. 따라서 스테그분석에서는 통계적 차이를 크게 하는 특징벡터를 추출하기 위해 영상의 히스토그램 또는 특성 함수로부터 특징벡터를 추출하는 방법[4-9]들이 제안되었다.

스테그분석에 관한 대표적인 연구로써 2010년 Pevny 등[10]은 정보 은닉에 의해 발생한 스테고 잡음을 추출하기 위해 인접 화소간의 차분영상을 구하고, 이에 대한 마코프 체인을 구성하는 방법을 제안하였다. 즉, 차분영상에 대한 1차 및 2차 마코프 체인을 이용하여 686개의 천이확률(transition matrix)을 특징벡터로 추출하였다. 이 방법은 거의 모든 스테가노그래피 방법에 대하여 높은 검출율을 얻을 수 있었다. 그러나 HUGO(Highly Undetectable SteGO)[11]와 같이 내용 적응적 스테가노그래피 방법에 대하여는 매우 낮은 검출율을 보이는 단점을 가지고 있다. 특히 최근의 스테가노그래피 기법들은 영상의 내용에 기반하여 복잡한 텍스쳐 또는 잡음 영역에 적응적으로 비밀 메시지를 숨기므로 통계적인 특성의 차이를 이용하여 비밀 메시지를 감지하기는 상당히 어렵다[12,13].

따라서 본 논문에서는 DCT 계수의 차이를 이용하여 마코프 체인을 구성하고, 이에 대한 천이확률을 특징으로 하는 내용 적응적 스테가노그래피에 특화된 스테그분석 방법을 제안하고자 한다. 먼저 영상을 8×8블록으로 분할하고, 각 블록에 대해 이산 코사인변환을 수행한다. 그리고 수평, 수직 및 대각 방향의 이웃 블록으로부터 이산 코사인변환 계수의 최대 차이를 구한 후, 이로부터 49개의 마코프 특징벡터를 추출한다. 추출된 특징벡터는 앙상블(ensemble) 분류기[14]에 입력되어 커버 영상과 스테고 영상을 학습하고 분류한다. 본 논문의 방법은 매우 적은 수의 특징으로 기존의 방법과 우수하거나 동등한 스테그분석 성능을 보인다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 내용 적응적 스테가노그래피 및 스테그분석 방법을 기술하고, Ⅲ장에서는 제안 방법을 설명한다. Ⅳ장에서는 실험을 통하여 제안 방법의 성능을 분석한다. 마지막으로 Ⅴ장에서 결론을 맺는다.

Ⅱ. 내용 적응적 스테가노그래피 및 스테그분석

최근의 내용 적응적 스테가노그래피 기법들은 영상의 내용에 따라 비밀 메시지를 숨기는 방법을 사용한다. 즉, 영상 내에서 복잡한 텍스쳐, 에지 또는 잡음에 가까운 영역을 선별하여 비밀 메시지를 은닉한다.

커버 영상을 X, 스테고 영상을 Y라 할 때 비밀 메시지는 다음과 같은 왜곡척도 $D(X,Y)$가 최소가 되는 형태로 은닉된다. 즉,

$$D(X,Y)=\sum_{i=1}^{N}\rho_i(x_i,y_i), \tag{1}$$

여기서 $\rho_i(x_i,y_i)$은 커버 영상의 i번째 화소 x_i을 y_i로 변화시키는 가중치를 의미한다. 내용 적응적 스테가노그래피는 항상 전체 영상의 왜곡을 최소화하는 방법으로 비밀 메시지를 삽입하므로 통계적 특성의 차이를 이용하여 은닉된 메시지의 존재 여부를 감지하기는 매우 어려워진다.

식 (1)의 왜곡 척도 함수는 공간 영역이나 이산 코사인 변환 영역, 웨이블릿 변환 영역과 같이 다양한 영역에서 정의될 수 있다. 대표적인 내용 적응적 스테가노그래피 방법으로는 HUGO[11], UNIWARD(UNIversal WAvelet Relative Distortion)[12], WOW(Wavelet Obtained Weights)[13] 등이 있다. HUGO는 공간 영역

의 이웃 화소값에 대해 왜곡척도가 정의되며, WOW 및 UNIWARD는 웨이블릿 변환 계수에 대해 서로 다른 왜곡 척도를 사용함으로써 비밀 메시지를 은닉한다.

그림 1은 576×768 크기의 영상에 대해 0.5bpp의 비트율로 은닉된 비밀 메시지의 위치를 보인 것이다. 그림 1(a)의 커버 영상은 많은 에지와 몇 개의 텍스쳐로 이루어져 있다. 그림 1(b)는 LSB Matching 방법에 의한 은닉 메시지의 위치를 나타내고 있는데, 비밀 메시지가 영상 전체에 랜덤하게 분포되어 있음을 알 수 있다. 이는 LSB Matching 방법이 가산 잡음 모델에 의한 스테가노그래피 방법을 사용하고 있기 때문이며, 가산 잡음 모델에 의한 스테고 영상들은 앞서 살펴본 SPAM (subtractive pixel adjacency matrix)의 방법[10]으로 쉽

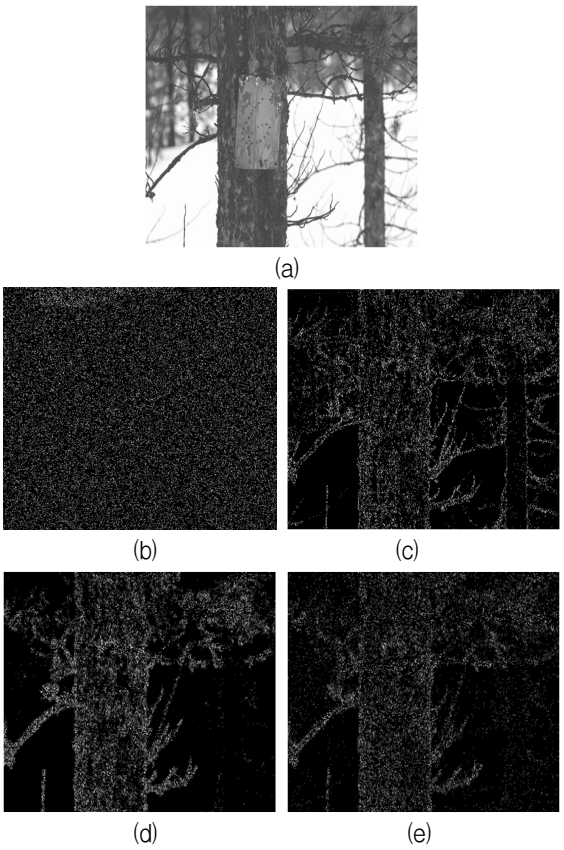


그림 1. 다양한 스테가노그래피 방법에 의해 은닉된 비밀 메시지의 위치 (엠베딩율 0.5 bpp),
(a) 커버 영상, (b) LSB matching, (c) HUGO,
(d) WOW, (e) S-UNIWARD

Fig. 1. The locations of secret messages embedded by various steganography methods (0.5 bpp embedding rate)
(a) cover image, (b) LSB matching, (c) HUGO,
(d) WOW, (e) S-UNIWARD.

게 검출이 가능하다.

반면 내용 적응적 스테가노그래피 방법인 HUGO에 의한 비밀 메시지는 그림 1(c)에 보이는 바와 같이 주로 에지 영역에 분포되어 있으며, WOW와 UNIWARD 방법은 각각 그림 1(d) 및 (e)에서 보는 바와 같이 에지 영역에 비해 텍스쳐 영역에 더 골고루 분포되어 있다. 이와 같이 내용 기반의 스테가노그래피는 기존의 가산 잡음 모델 기반의 방법과는 달리 비밀 메시지의 삽입 영역이나 사용된 왜곡 척도와는 무관하게 대부분 시각에 덜 예민한 영역에 은닉되어 비밀 메시지의 검출이 더욱 어려워진다. 이에 따라 최근에는 내용 적응적 스테가노그래피 방법을 위한 스테그분석 방법들이 제안되고 있다.

2011년 Liu[15]는 YASS와 DCT 영역의 정보 은닉 방법을 검출하기 위한 방법을 제안하였다. 또한 HUGO를 공격하기 위한 스테그분석 방법으로 Fridrich 등[16]은 수평, 수직, 대각 방향의 인접 화소로부터 3차 오차 영상을 구하고 이로부터 동시 발생 행렬을 특징으로 하는 33,963개의 벡터를 추출하였다. 2013년 Holub 등[17]은 불필요한 특징벡터의 수를 줄이고 비밀 메시지의 은닉 영역과는 무관한 스테그분석을 위해 수평, 수직 방향의 인접 화소로부터 선형 및 비선형 오차 영상을 구하고 이를 8개의 랜덤 방향으로 투영한 후 이들의 히스토그램으로부터 12,870개의 특징벡터를 추출하는 방법을 제안하였다. 반면 2014년 Denemark 등[18]은 S-UNIWARD(UNIWARD의 공간 영역 버전)을 공격하기 위한 방법으로써 영상을 변화가 많은 영역과 그렇지 않은 영역으로 분류하고 각각의 3차 오차 영상을 구한 후 히스토그램과 동시 발생 행렬로부터 1,183개의 특징벡터를 추출하였다. 그러나 커버 영상의 종류가 달라지거나 UNIWARD와는 다른 왜곡 척도를 사용하는 HUGO 및 WOW에 대해서는 좋지 못한 성능을 가진다.

위의 방법들은 비밀 메시지의 비트율이 증가함에 따라 평탄한 영역까지 메시지가 은닉되어 공간 적응적 특성을 점점 잃게 된다. 따라서 공간 영역에서 인접 화소간의 의존성을 정밀하게 모델링하기 어려울 뿐 아니라 정확한 모델링을 위해 매우 많은 특징 벡터를 요구하게 된다. 이러한 단점을 극복하기 위하여 본 논문에서는 영상을 블록 단위로 분할하고, 인접 블록 간의 주파수 변화의 의존성을 매우 적은 수의 특징벡터로 모델화함으로써 다양한 비트율에서도 스테그분석의 성능을 향상

시킬 수 있는 방법을 제안하고자 한다.

Ⅲ. 제안 방법

2.1. 주파수 영역 블록 간 차이 영상 구성

본 논문에서는 내용 적응적 스테가노그래피 방법이 영상의 에지 및 텍스쳐 영역과 같이 주파수가 높은 영역에 주로 비밀 메시지를 삽입한다는 것에 착안하여, 영상의 블록 간 주파수의 차이를 변화를 특징으로 이용하는 스테그분석 방법을 제안하고자 한다.

그림 2는 비트율 0.5bpp에서 은닉된 비밀 메시지의 위치와 공간 영역 및 주파수 영역에서 인접 화소간의 차이 영상을 나타낸 것이다. 그림 2에서 보는 바와 같이 공간 영역에 비해 이산 코사인 변환 계수의 차이 영상이 은닉 메시지의 위치를 더 잘 표현함을 확인할 수 있다. 따라서 본 논문에서는 그림 3과 같이 $M\times N$ 크기를 가지는 영상을 8×8블록으로 분할하고, 각 블록에 대해 이산 코사인 변환을 수행한다. 그리고 영상의 내용을 억제하고 스테고 잡음을 추출하기 위해 이웃한 세 방향의 이산 코사인변환 블록으로부터 수평, 수직 및 대각 성분의 계수의 차이를 구한다. 이산 코사인변환 블록을 $C_{k,l}(i,j)$라 하고, (i,j)를 블록 내의 좌표값 $(i=j=1,2,\cdots,8)$이라 할 때, 두 블록에 대한 수평, 수직, 대각 방향의 계수의 차이는 다음과 같이 구한다. 즉,

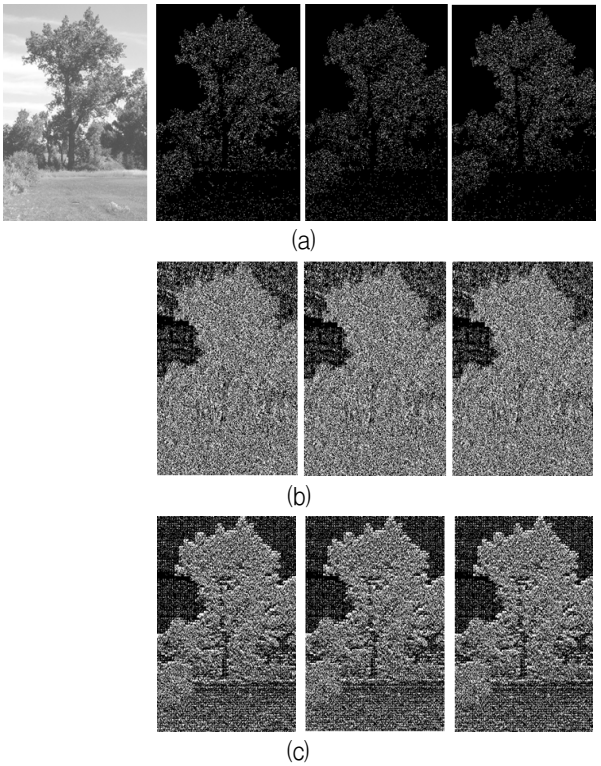


그림 2. 은닉 메시지의 위치와 공간 영역 및 주파수(이산 코사인 변환) 영역의 차이 영상 (a) 은닉 메시지, (b) 공간 영역 차이 영상 (c) 주파수 영역 차이 영상 (왼쪽으로부터 HUGO, S-UNIWARD, WOW 방법의 구성된 차이 영상)

Fig. 2. The difference images and the location of hidden messages in spatial and frequency (discrete cosine transform) domain, (a) hidden messages, (b) difference images in spatial domain, (c) difference images in frequency domain (from left, the difference images by HUGO, S-UNIWARD, and WOW.

$$R^{h}_{k,l}(i,j)=C_{k,l+1}(i,j)-C_{k,l}(i,j), \tag{2}$$

$$R^{v}_{k,l}(i,j)=C_{k+1,l}(i,j)-C_{k,l}(i,j), \tag{3}$$

$$R^{d}_{k,l}(i,j)=C_{k+1,l+1}(i,j)-C_{k,l}(i,j), \tag{4}$$

여기서 $R^{h}_{k,l}(i,j)$, $R^{v}_{k,l}(i,j)$, $R^{d}_{k,l}(i,j)$은 각각 수평 및 수직 그리고 대각 방향의 계수 차이를 나타낸다.

그림 4는 식 (2), (3), 및 (4)에 의해 구해진 차이 영상을 나타내고 있다. 그림 4로부터 세 개의 차이 영상은 서로 다른 방향으로의 스테고 잡음 정보를 제공함을 알 수 있다. 즉, 비밀 메시지가 수평 에지에 많이 밀집되어 있다면 수직 방향의 차이 영상은 수평 및 대각 방향의 차이 영상보다 더 큰 값을 가지게 된다. 이러한 사실은 이산 코사인 변환 블록의 차이 영상은 그림 1의 내용 기반 적응적 스테가노그래피에 의해 삽입된 스테고 잡음의 위치 정보를 잘 표현할 수 있음을 의미한다.

따라서 본 논문에서는 에지와 텍스쳐에 적응적인 스테그분석을 위해 세 개의 차이 영상으로부터 최대 차이를 가지는 차이 블록 $R^{m}_{k,l}(i,j)$을 다음과 같이 구한다.

$$R^{m}_{k,j}(i,j)=\max\left\{R^{h}_{k,l}(i,j),R^{v}_{k,l}(i,j),R^{d}_{k,l}(i,j)\right\} \tag{5}$$

식 (5)의 최대 차이를 가지는 이산 코사인 계수를 특징으로 사용함으로써 가장 높은 주파수 차이를 가지는 이산 코사인변환 계수를 특징으로 사용할 수 있다.

2.2. 마코프 체인 구성 및 특징 추출

본 논문에서는 인접 블록간 이산 코사인변환 계수의 차이에 대한 마코프 체인을 구성하고, 이에 대한 확률값을 스테그분석을 위한 특징으로 사용한다. 마코프 천이 확률값은 SPAM[10] 등의 다양한 스테그분석 방법에서 사용되고 있다. 그러나 일반적으로 이산 코사인 변환 계수의 값은 큰 범위에 걸쳐 존재하고, 그에 대한 차이값 또한 넓은 범위를 가진다. 따라서 계수의 차이에 대한 마코프 상태는 매우 많아지게 되고 이로 인해 특징 벡터의 수는 기하급수적으로 증가하게 된다. 따라서 본 논문에서는 SPAM의 방법과 마찬가지로 계수의 차이의 범위를 임의의 임계값 이내, 즉 $[-T, T]$로 제한하는 방법을 사용한다. 즉,

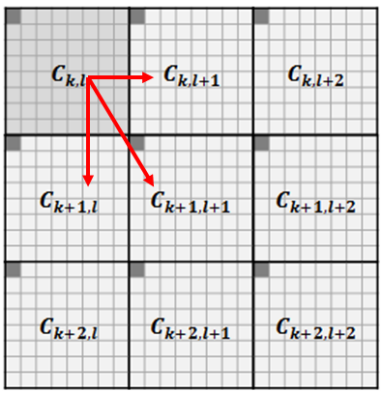


그림 3. $C_{k,l}$을 중심으로 하는 인접한 이산 코사인 변환 블록

Fig. 3. Adjacent blocks of discrete cosine transform centered by $C_{k,l}$

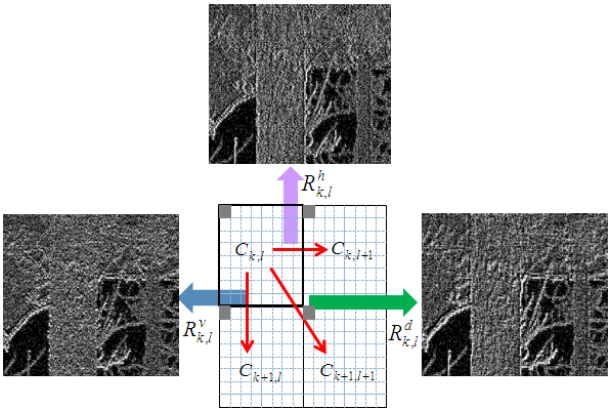


그림 4. 인접 DCT 블록에서 수평, 수직 및 대각 방향의 차이 영상

Fig. 4. The difference images between horizontal, vertical and diagonal direction in neighboring DCT blocks.

$$tR^m_{k,l}(i,j) = \begin{cases} R^m_{k,l}(i,j) & , \ |R^m_{k,l}(i,j)| \le T \\ T \cdot sgn(R^m_{k,l}(i,j)) & , \ otherwise \end{cases}, \quad (6)$$

여기서 $tR^m_{k,l}(i,j)$ $(-T \le tR^m_{k,l}(i,j) \le T)$는 임계값에 의해 제한된 이산 코사인변환 계수의 차이값을 나타낸다.

임계값 이내로 제한된 $tR^m_{k,l}(i,j)$에 대해 1차 마코프 전이 확률은 다음의 식과 같이 정의된다.

$$M(s,t) = \Pr[tR^m_{k,l}(i,j) = s | tR^m_{k,l+1}(i,j) = t], \quad (7)$$

여기서 s와 $t(-T \le s,t \le T)$는 제한된 이산 코사인 변환 계수값의 차이이고, $M(s,t)$는 $tR^m_{k,l+1}(i,j) = t$에서 $tR^m_{k,l}(i,j) = s$로 천이할 확률이다.

마코프 천이 확률에 의한 특징벡터의 수는 임계값 T에 의존하며, $(2T+1)^2$개의 벡터를 가지게 된다. 본 논문의 방법은 각각 수평, 수직, 대각 방향 계수 차이의 최대값을 이용하기 때문에 하나의 마코프 체인만 구성하여도 세 방향에 대한 주요 특징을 추출할 수 있다. 따라서 본 논문에서는 $T=3$을 이용하여 총 49개의 특징 벡터를 추출하며, 이를 커버 영상 및 스테고 영상을 검출하기 위한 특징 벡터로 사용한다.

Ⅳ. 실험 및 결과

4.1. 실험 환경 설정

다양한 영상에서 제안 방법의 성능을 평가하기 위해 본 논문에서는 압축되지 않은 BMP 포맷의 NRCS (Natural Resources Conservation Services)[19] 및 Bossbase1.01[20] 데이터베이스 영상을 각각 1,000개씩 사용하였다. NRCS는 식물, 동물 및 토양, 물 등의 사진으로 구성된 1576개의 그레이 영상으로써 512×768, 520×768 및 1024×6838 등 다양한 크기를 가진다. 또한 Bossbase1.01은 다양한 건축물과 바다, 동물 등의 사진으로 구성된 크기 512×512의 그레이 영상이다.

비밀 메시지의 삽입을 위해 내용 적응적 스테가노그래피 방법인 HUGO 및 WOW 그리고 S-UNIWARD 임베딩 방법을 사용하여 2,000개의 커버 영상에 대해 각각 0.05 bpp, 0.1 bpp, 0.25 bpp, 0.3 bpp, 0.5 bpp 그리

고 1.0 bpp의 삽입률을 갖는 다양한 스테고 영상을 만들었다. 또한 제안된 스테그분석 방법의 성능을 비교하기 위해 인접 화소와의 차분 영상으로부터 686개의 마코프 특징 벡터를 추출하는 SPAM[10]과 내용 선택 잔류 영상으로부터 1,183개의 동시 발생 특징벡터를 추출하는 CSR(content adaptive residuals)[18]의 성능을 평가하였다. 이때 CSR에서 사용된 파라메타의 값은 $T_h = 10$, $\alpha = 0.4$, t_s와 t_L은 각각 0.05와 0.06이다. 추출된 특징벡터들은 앙상블 분류기[14]를 사용하여 커버 및 스테고 영상을 분류하였다.

스테그분석 방법의 성능 평가를 위한 첫 번째 척도로써 본 눈문에서는 TP(true positive)율과 TN(true false)율 및 정확도(accuracy, ACC)[21]를 사용하였다. TP율과 TN율은 각각 전체 영상으로부터 커버 영상 및 스테고 영상을 정확하게 판단한 정도를 의미한다. 또한 정확도는 커버 영상과 스테고 영상 모두를 정확하게 판단한 확률로써 평균적인 검출율을 나타낸다.

4.2. 스테그분석의 정확도 평가

표 1은 다양한 비트율을 가지는 HUGO 방법의 스테고 영상에 대해 본 논문의 방법과 SPAM, CSR 방법의 검출 정확도를 나타낸 것이다. 표 1의 결과로부터 제안 방법은 모든 비트율에서 SPAM과 CSR 방법에 비해 검출율이 상당히 개선되었음을 알 수 있다. 특히 HUGO의 공격에 취약한 SPAM의 경우 최소 9.2%에서 최대 20.1% 이상이 향상되었으며, CSR에 대해서도 5.7%에서 19.8%의 성능이 개선되었다.

스테그분석은 악의적 목적을 가진 비밀 메시지의 감지가 주된 목적이므로 성능 평가시 TN율이 매우 중요하다. 이러한 점에서 제안 방법은 모든 비트율에서 TN율이 TP율보다 더 나은 성능을 보이고 있으며, 특히 0.3 bpp 이하의 저 비트율에서 평균 15% 정도의 검출 성능 개선을 보이고 있다.

표 1. HUGO 영상에 대한 검출율 비교(단위 : %)
Table 1. Detection rate comparison for HUGO images. (Unit : %)

	HUGO								
	SPAM[10] (686)			CSR[18] (1183)			제안방법 (49)		
bpp	TP	TN	ACC	TP	TN	ACC	TP	TN	ACC
0.1	48.5	52.9	**50.7**	49.6	52.6	**51.1**	57.4	76.8	**67.1**
0.3	54.0	54.7	**54.4**	51.4	58.0	**54.7**	83.7	65.3	**74.5**
0.5	54.2	64.5	**59.4**	54.8	67.3	**61.1**	71.4	83.2	**77.3**
1.0	65.0	78.7	**71.9**	69.3	81.4	**75.4**	80.5	81.6	**81.1**

표 2 및 3은 다양한 삽입 비트율에 대하여 WOW 방법 및 S-UNIWARD 방법에 대한 스테그분석 성능을 평가한 것이다. 실험 결과 WOW 영상에 대해 SPAM과 CSR은 0.5 bpp 이하의 비트율에서 60% 이하의 매우 낮은 검출 정확도를 보인다. 이는 저 비트율에서 스테고 영상을 거의 감지할 수 없음을 의미한다. 또한 제안 방법은 SPAM과 비교하여 15%에서 21.9%, 그리고 CSR 방법과 비교하여 11%에서 19.8%의 검출 정확도의 개선이 이루어졌다. 특히 1.0 bpp의 비트율에서 TN율이 93% 이상으로 매우 높음을 알 수 있다.

반면 S-UNIWARD를 공격하기 위해 제안되었던 CSR은 테스트 커버 영상의 유형과 학습 영상의 개수가 달라짐에 따라 성능이 크게 저하됨을 알 수 있다. 그러나 제안 방법은 스테가노그래피의 종류와 사용된 왜곡 척도에 무관하게 모든 비트율에서 기존의 방법에 비해

표 2. WOW 영상에 대한 검출율 비교 (단위 : %)
Table 2. Detection rate comparison for WOW images. (Unit : %)

	WOW								
	SPAM (686)			CSR (1183)			제안방법 (49)		
bpp	TP	TN	ACC	TP	TN	ACC	TP	TN	ACC
0.1	46.7	55.2	**51.0**	50.5	60.4	**55.5**	55.3	76.7	**66.0**
0.3	49.9	56.8	**53.5**	54.5	57.7	**56.0**	78.4	68.6	**73.5**
0.5	54.8	60.9	**57.9**	55.9	63.8	**59.9**	79.6	71.4	**75.5**
1.0	65.6	77.7	**71.7**	68.7	78.9	**73.8**	92.6	94.5	**93.6**

표 3. S-UNIWARD 영상 대한 검출율 비교 (단위 : %)
Table 3. Detection rate comparison for S-UNIWARD images.(Unit : %)

	S-UNIWARD								
	SPAM (686)			CSR (1183)			제안방법 (49)		
bpp	TP	TN	ACC	TP	TN	ACC	TP	TN	ACC
0.1	52.6	49.9	**51.3**	55.0	63.1	**59.1**	55.7	80.0	**67.9**
0.3	50.8	61.9	**56.4**	63.5	69.2	**60.4**	74.0	76.1	**75.1**
0.5	54.1	71.8	**63.0**	62.1	77.7	**69.9**	76.4	79.6	**78.0**
1.0	71.5	87.7	**79.6**	73.0	88.6	**80.8**	82.8	90.7	**86.8**

6%에서 최대 18.7% 검출 정확도가 높아졌다.

제안 기법의 성능을 평가하기 위한 두 번째 척도로써 본 논문에서는 평균 오류 확률을 평가하였다. 커버 영상으로부터 스테고 영상을 감지한 확률을 P_{FP}라 하고, 스테고 영상으로부터 커버 영상을 잘못 감지한 확률을 P_{FN}이라 하면, 평균 오류 확률은 다음과 같이 정의된다. 즉,

$$P_{error} = 1 - \frac{1}{2}(P_{FP} + P_{FN}) \tag{8}$$

그림 5는 세 가지 방법(HUGO, WOW, S-UNIWARD)으로 생성된 스테고 영상에 대한 오류 확률을 비교 평가한 것을 나타내고 있다. 그림 5로부터 제안 방법의 평균 오류 확률은 기존의 방법들에 비해 모든 비트율에서 저하되었으며, 특히 1.0 bpp의 고비트율에서 0.2 이하로 매우 낮고 WOW 영상의 경우 0.132의 더 낮은 오류 확률로 개선됨을 알 수 있다.

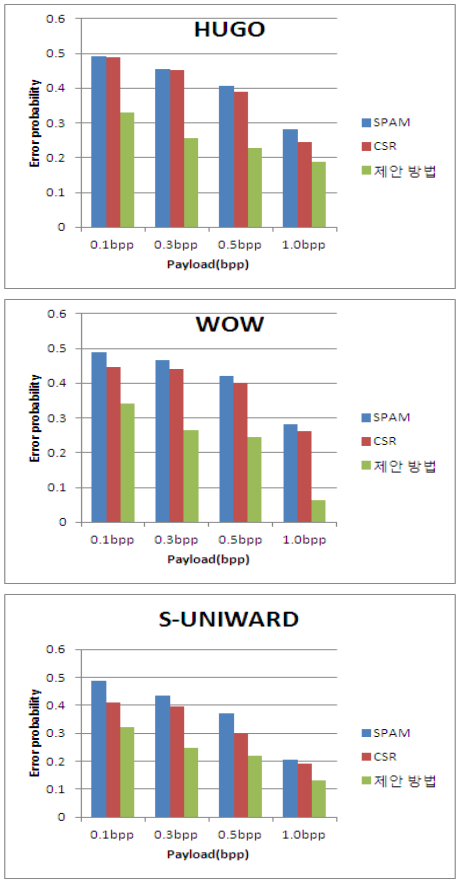


그림 5. 기존 방법과 제안 방법의 오류 확률
Fig. 5. Error probability for conventional methods and proposed method.

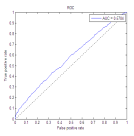
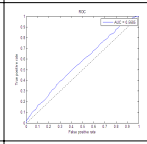
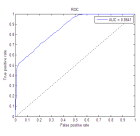
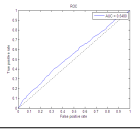
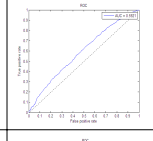
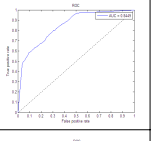
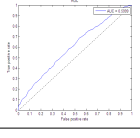
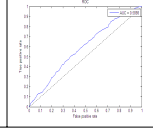
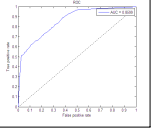
또한 제안 방법의 성능을 평가하기 위한 세 번째 척도로는 AUC(area under the curve)[22]를 사용하였다. AUC는 ROC(receiver operating characteristic) 곡선 아래의 면적을 의미한다. 일반적으로 스테그분석의 정확도는 ROC 곡선이 왼쪽 모서리에 가까울수록, 그리고 AUC가 1에 가까울수록 높아진다.

표 4는 모든 비트율에서 AUC를 비교한 것이며, 표 5는 비트율 0.3 bpp에서 기존의 방법과 제안 방법의

표 4. 기존 방법과 제안 방법의 AUC 비교
Table 4. Comparison of ACU between conventional methods and proposed method.

삽입방법 \ 스테그분석		SPAM[14]	CSR[18]	제안방법
HUGO	0.1bpp	0.5118	0.5051	0.7240
	0.3bpp	0.5704	0.5692	0.8641
	0.5bpp	0.6415	0.5982	0.8854
	1.0bpp	0.8036	0.8102	0.9120
WOW	0.1bpp	0.5103	0.5211	0.7382
	0.3bpp	0.5480	0.5921	0.8449
	0.5bpp	0.6048	0.6478	0.8771
	1.0bpp	0.7993	0.8021	0.9726
S-UNIWARD	0.1bpp	0.5169	0.5256	0.7621
	0.3bpp	0.5999	0.5898	0.8599
	0.5bpp	0.6866	0.6910	0.8869
	1.0bpp	0.8789	0.8789	0.9440

표 5. 기존 방법과 제안 방법의 ROC 곡선 비교 (0.3bpp)
Table 5. Comparison of ROC curves between conventional methods and proposed method (0.3bpp).

삽입방법 \ 스테그분석	SPAM[9]	CSR[12]	제안방법
HUGO			
WOW			
S-UNIWARD			

ROC를 나타낸 것이다. 제안 방법은 모든 스테고 영상에 대해 기존의 방법에 비해 AUC의 값이 높으며 특히 1.0 bpp에서 0.9 이상의 매우 높은 값을 가짐을 알 수 있다.

Ⅳ. 결 론

본 논문에서는 복잡한 에지 영역이나 텍스쳐 영역에 비밀 메시지를 은닉하는 내용 기반 스테가노그래피의 개선된 스테그분석을 위해 영상을 블록 단위로 분할하고, 이웃한 DCT 블록 계수의 최대 차이 영상으로부터 마코프 특징 벡터를 추출하는 방법을 제안하였다. 다양한 삽입률의 스테고 영상에 대해 실험한 결과 제안 방법은 모든 비트율에서 스테가노그래피의 종류와 사용된 왜곡 척도에 무관하게 기존의 방법에 비해 6%에서 최대 18.7% 만큼 검출 정확도가 높아졌다. 향후 저 비트율의 스테고 영상을 보다 더 정확하게 검출할 수 있는 스테그분석에 관한 연구가 필요하다고 생각된다.

REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "*Information hiding - a survey*," Proceedings o fthe IEEE, vol. 87, no. 7, pp. 1062-1078, 1999.

[2] C. Abbas, C. Joan, C. Kevin and M. Paul, "*Digital image steganography : Survey and analysis of current methods*," Signal Processing, vol. 90, no. 3, pp. 727-752, 2010.

[3] A. Nissar, and A. H. Mir, "*Classification of steganalysis techniques: A study*", Digital Signal Processing, vol. 20, no, 6, pp. 1758-1770, 2010.

[4] H. Farid, S. Lyu, "*Detecting hidden messages using higher order statistics and support vector machines*", Lecture Notes in Computer Science, vol. 2578, pp. 340-354, 2002.

[5] Y. Wang, P. Moulin, "*Optimized feature extraction for learning-based image steganalysis*," IEEE Transactions on Image Processing, vol. 2, no. 1, pp. 31-45, 2007.

[6] X. Luo, F. Liu, S. Lian, C. Yang, S. Gritzalis, "*On the typical statistic features for image blind steganalysis*," IEEE Journal on Selected Areas in Communications, vol. 29, no. 7, pp. 1404-1422, 2011.

[7] M. Tang, J. Hu, M. Fan and W. Song, "*A steganalysis by adjacency pixel bits structure*," Computers and Electrical Engineering, vol. 39, no. 2, pp. 488-498. 2013.

[8] Q. Liu, A. H. Sung, B. Ribeiro, M. Wei, Z. Chen, J. Xu, "*Image complexity and feature mining for steganalysis of least significant bit matching steganography,*" Information Sciences, vol. 178, pp. 21 - 36, 2008.

[9] T. H. Park, S. H. Hyun, J. H. Kim and I. K. Eom, "*Steganalysis using joint moment of wavelet subbands*," Journal of IEEK, vol. 48-SP, no. 3, pp. 71-78, 2011.

[10] T. Pevny, P. Bas and J. Fridrich, "*Steganalysis by subtractive pixel adjacency matrix*," IEEE transaction on Information Forensics and Security. vol. 5, no. 2, pp. 215-224, 2010.

[11] T. Pevný, T. Filler, and P. Bas, "*Using high-dimensional image models to perform highly undetectable steganography*," Lecture Notes in Computer Science, vol. 6387, pp. 161-177, 2010.

[12] V. Holub and J. Fridrich, "*Digital image steganography using universal distortion.*" Proceedings of the first ACM workshop on Information hiding and multimedia security, pp. 59-68, 2013.

[13] V. Holub, J. Fridrich and T. Denemark, "*Universal distortion function for steganography in an arbitrary domain,*" EURASIP Journal on Information Security, vol. 2014:1, pp. 1-13, 2014..

[14] J. Kodovský, J. Fridrich, and V. Holub, "*Ensemble classifiers for steganalysis of digital media*," IEEE Transaction on Information Forensics and Security. vol. 7, no. 2, pp. 432-444, 2012.

[15] Q. Liu, "*Steganalysis of DCT-embedding based adaptive steganography and YASS,*" Proceedings of the thirteenth ACM Multimedia Workshop on Multimedia and Security, pp. 77-86), 2011.

[16] J. Fridrich, J. Kodovsky, V. Holub, and M. Goljan, "*Steganalysis of content-adaptive steganography in spatial domain*," Lecture Notes in Computer Science, vol. 6958, pp. 102-117, 2011.

[17] V. Holub, J. Fridrich, "*Random projections of residuals for digital image seganalysis,*" IEEE Transactions on Information Forensics and Security, vol. 8, no. 12, pp. 1996-2006, 2013.

[18] T. Denemark, J. Fridrich and V. Holub, "*Further Study on the Security of S-UNIWARD*," Proceedings of SPIE Media Watermarking, Security, and Forensics, vol. 9028, pp. p. 902805-1-902805-13, 2014.
[19] NRCS Photo Gallery, http://photogallery.nrcs.usda.gov/res/sites/photogallery
[20] Bossbase1.01, http://www.agents.cz/boss
[21] I. H. Witten, and E. Frank, Data Mining, Elservier, 2005.
[22] A. Westfeld, "ROC curves for steganalysis," Proceedings of the third WAVILA Challenge, pp. 39-45, 2007.

저 자 소 개

박 태 희(정회원)
1993년 부경대학교 정보통신 공학과 학사 졸업.
1996년 부경대학교 전자공학과 석사 졸업.
1999년 부산대학교 전자공학과 박사 졸업.
현재 동명대학교 메카트로닉스공학과 조교수
<주관심분야 : 영상 신호처리, 스테그분석, 영상 접합 분석, 영상 포렌식, 멀티미디어 보안>

엄 일 규(정회원)-교신저자
1990년 부산대학교 전자공학과 학사 졸업.
1992년 부산대학교 전자공학과 석사 졸업.
1998년 부산대학교 전자공학과 박사 졸업.
현재 부산대학교 전자공학과 교수
<주관심분야 : 영상 신호처리, 웨이블릿 변환 신호처리, 영상 포렌식, 멀티미디어 보안>

한 종 구(학생회원)
2004년 인제대학교 전자공학과 학사 졸업.
2007년 요코하마 국립대학교 물리 전자공학과 석사 졸업.
2014년 부산대학교 전자공학과 박사과정 수료.
<주관심분야 : 영상 신호처리, 웨이블릿 변환 신호처리, 영상 접합 분석>