JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 19, No. 2, pp. 371-376, Feb. 2018

도약기반의 하이브리드 히스토그램 시프팅을 이용하는 데이터 은닉 방법

최 용 수[1] , 이 달 호[2*]
[1]성결대학교 파이데이아대학(멀티미디어)
[2*]가천대학교 IT대학 전자공학과

Data Hiding Method Utilizing Skipping Based Hybrid Histogram Shifting

YongSoo Choi[1], DalHo Lee[2*]
[1]Division of Liberal Arts & Teaching, Sungkyul University. Anyang, Korea
[2*]Department of Electronics, IT College, Gachon University, SungNam, Korea

[요 약]

시스템 보안기술에서 정보은닉분야는 일반적으로 많이 사용되는 콘텐츠를 매개체로 하며 정보를 삽입하는 기술들로 개발이 되어진다. 제안하는 기술은 기술적인 스테가노그라피기술로서 콘텐츠가 가진 신호 값들의 물리적/통계적 변화를 통해 일정 정보를 은닉하는 기술을 사용한다. 최근 가역 데이터 은닉 분야에서 히스토그램 시프팅에 기반 한 다양한 연구들이 있었다. 단일 피크 히스토그램 시프팅에서 다중 피크 히스토그램 기법 등의 적용으로 데이터 은닉의 용량은 점진적으로 증가하였다. 본 논문에서는 도약(Skipping)을 포함하는 히스토그램 시프트 방법을 채용하는 관점에서 은닉의 효과를 분석하고자 한다. 또한 은닉용량의 향상을 위한 범용적인 방법으로 다중 분기 데이터 은닉을 제안한다. 위의 제안은 수식을 이용한 예제로 증명이 되었으며 추가적인 향상 방안을 도출할 수 있었다.

[Abstract]

In the system security technology, the information hiding field is developed as technologies for embedding information, which are generally used as contents media. The proposed technique is a technical steganography technique which uses a technique of concealing certain information through physical / statistical change of signal values of contents. Recently, there have been various studies based on histogram shifting in reversible data concealment. In multi - peak histogram shifting, the capacity of data concealment gradually increased by applying multiple peak histogram method. In this paper, we analyze the effect of concealment in terms of adopting the histogram shift method including skipping. In addition, we propose multi - branch data concealment as a general method to improve concealment capacity. The above proposal has proved to be an example using mathematical expressions, and further improvement measures could be derived.

색인어 : 가역, 데이터 은닉, 도약, 다중분기, 스테가노그라피
Key word : Reversible, Data hiding, Skipping, Multi divergence, Steganography

http://dx.doi.org/10.9728/dcs.2018.19.2.371

Received 19 February 2018; **Revised** 26 February 2018
Accepted 27 February 2018

***Corresponding Author; DalHo Lee**

Tel: +82-31-750-5320
E-mail: dhlee@gachon.ac.kr

Ⅰ. 서 론

워터마킹, 스테가노그라피 등 데이터 은닉 분야에서는 수많은 알고리즘들이 개발이 되어왔으며 정지영상, 동영상, 오디오 등 다양한 멀티미디어 매체들에 이 기법들이 선택적으로 적용되어져왔다. 또 신호가 가지는 다양한 표현영역에서의 중복성 등을 다루기 위해 시간영역, 주파수영역 등 다양한 표현공간에서의 데이터 은닉 기법들로 확장되어져 왔다. 전술한 바와 같이 공간/주파수 영역에서의 연구들은 저작권 보호기술을 위한 디지털 워터마킹(Digital Watermarking)과 비밀 메시지 전송을 위한 스테가노그라피(Steganography)기술 두 가지로 나뉜다.

본 논문에서는 스테가노그라피를 이용한 데이터 은닉(Data Hiding)기법에 대해 다루고자 하며 특히 정지영상에 대한 가역 데이터 은닉방법에 대해 연구하고자 한다. 가역 데이터 은닉 기법은 영상의 변형을 통해 사용자의 메시지를 삽입하며, 변형된 영상을 취득한 적법한 사용자는 삽입 시 사용된 부가정보 등을 이용하여 매체 속에 숨겨진 메시지를 검출한 후 원래의 영상을 복원해낸다[1, 2]. 가역 데이터 은닉의 가장 큰 장점은 메시지 검출 후 원본이 복원되므로 메시지의 재삽입이 가능하게 되므로 새로운 정보(소유권 등)를 은닉하는 것이 가능하다. 즉, 최초의 저작권정보를 유통단계에 따라 변경해 가는 응용이 가능한 것이다. 가역 스테가노그라피 기술은 크게 두 가지의 요구사항을 가져야 한다. 첫째는 원본 신호(영상)에 가해지는 변형이 비지각적(Imperceptible)이어야 한다는 것이다. 신호처리분야에서는 이를 비교하는 지표로 PSNR(Peak Signal to Noise Ratio)를 사용하고 있으며 PSNR이 높을수록 원본 콘텐츠와 변형 콘텐츠의 유사도가 높은 것으로 평가한다. 둘째는 높은 정보은닉 용량이다. 정보은닉 기술은 콘텐츠 인증에 관한 중요한 기술로 사용되므로 높은 은닉용량은 저작권 인증에 관해 더 풍부한 정보들을 삽입할 수 있음을 의미한다.

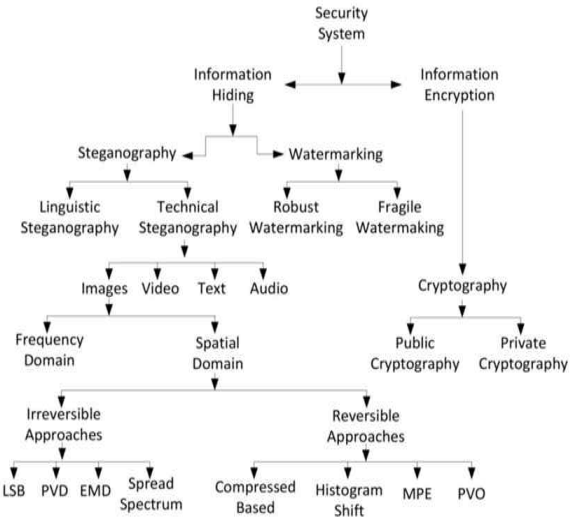

그림 1. 정보 보안 기술의 분류
Fig. 1. Classification of Information Security Techniques

이 논문은 다음과 같은 구성으로 이루어져 있다. 2장에서는 히스토그램 시프팅 기반 가역 정보은닉기술의 개념 및 방법에 대해 살펴본다. 3장에서는 제안하는 도약기법의 분석 및 한계 도출 그리고 다중 발산(2^n divergence)을 이용하는 하이브리드 히스토그램 시프팅 기반 데이터 은닉에 대해 설명한다. 4장에서는 제안하는 방법들의 효과에 대해 정리하고 향후 개선방향에 대해 논의한다.

Ⅱ. 히스토그램 시프팅 정보은닉 기법

히스토그램 변형기반 가역 정보은닉 기법들은 대부분 그림 2와 같은 순서로 수행된다. 그림 2에서는 이웃하는 두 픽셀의 차이 값 히스토그램을 은닉을 위한 자원으로 사용을 하였다. 초기의 논문들에서는 영상의 모든 픽셀 값들에 대한 히스토그램을 변경하는 것으로 데이터를 은닉하였으나 Peak 히스토그램의 빈도를 높이기 위해 주변 픽셀과의 차이 값, 특정 픽셀과 주변 블록평균과의 차이 값 또는 특정 픽셀과 예측 값의 차이 값 등의 히스토그램을 변형하는 방법으로 진화되어져 왔다[3,5-9]. 히스토그램 변형기반 정보은닉, 검출 및 영상복원의 순서는 다음과 같다.

2-1 정보은닉

1) 이웃하는 두 픽셀의 차 값들에 대한 히스토그램 생성
2.1) 최대(Peak), 최소(Zero) 히스토그램 Indices(Idx_P, Idx_Z)를 설정
2-2) 특정영역([Idx_P+1,.,Idx_Z-1])의 히스토그램을 Shifting
- 특정영역에 해당되는 픽셀 쌍들의 모든 차 값이 1 증가되도록 픽셀 값을 변경

3) Message의 비트열 값(0또는 1)에 따라 Idx_P위치의 히스토그램 변형
- 0 삽입의 경우 픽셀 쌍의 값은 보존, 1 삽입의 경우 픽셀 쌍의 찻값이 1 증가하도록 픽셀 값을 변경

2-2 메시지 검출 및 히스토그램 복원

1) 이웃하는 두 픽셀의 차 값들에 대한 히스토그램 생성
2) 최대 Index(Idx_P)와 Idx_P+1 위치의 히스토그램을 통해 Message 비트열 검출 및 최대 히스토그램 빈도 복원
- 픽셀 쌍의 차 값이 Idxp의 인덱스일 경우 0을 검출, Idxp+1일 경우 1을 검출하고 픽셀 쌍의 차이 값이 1 감소하도록 픽셀 값을 변경

3) 특정영역([Idx_P+2,...,Idx_Z])의 히스토그램을 원래의 위치로 역 Shifting
- 특정영역에 해당되는 픽셀 쌍들의 모든 차 값이 1 감소하도록 픽셀 값을 변경

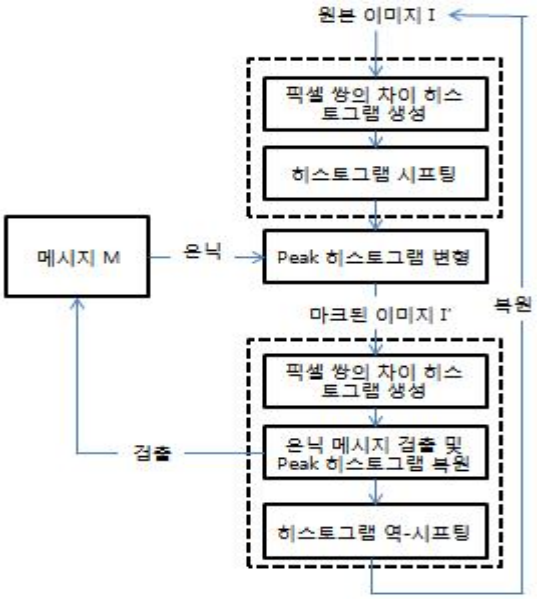

그림 2. 히스토그램 시프팅을 이용한 가역 정보은닉
Fig. 2. Reversible Information Hiding based on Histogram Shifting

그림 2는 아래의 그림 3과 같이 실제 히스토그램 변형의 예시로 도시가 가능하다. 입력 영상을 I로 가정하고 I를 이웃하는 두 개의 픽셀 쌍들로 나눈 다음 Pixel Difference Value(d) $d = abs(I(i) - I(i-1))$를 연산하여 히스토그램을 얻는다.

그림 4는 픽셀 쌍의 차이(d)를 구함에 있어 절대 값을 취하지 않는 방법이다. 이 방법은 각각의 영역에서의 최대 히스토그램을 얻게 되므로 보다 많은 삽입 용량을 얻을 수 있다. 즉, $d = I(i) - I(i-1)$.

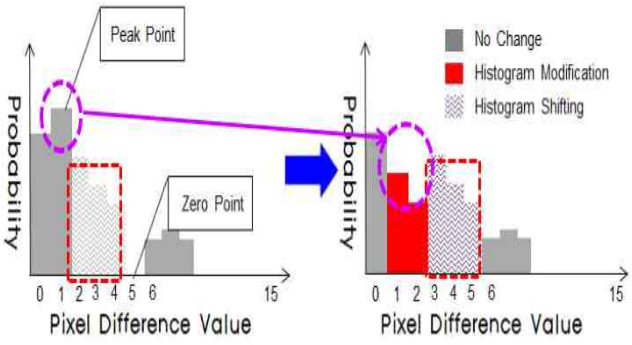

그림 3. 히스토그램 변형의 시각적 도시
Fig. 3. Graphical Example of Histogram Modification

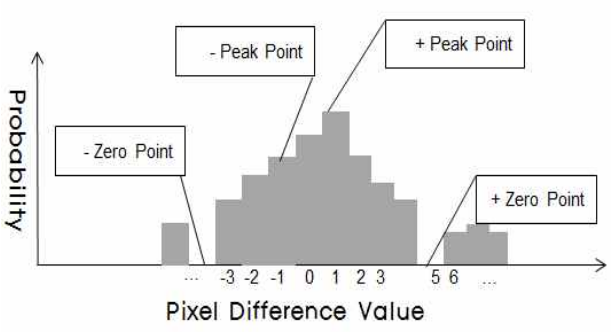

그림 4. 복수의 히스토그램 피크 변형
Fig. 4. Multiple Peak Histogram Modification

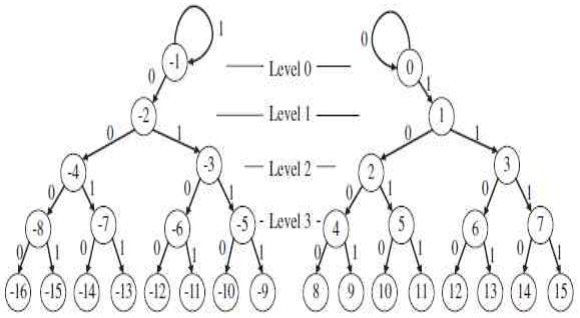

그림 5. 이중 바이너리 트리를 이용한 히스토그램 변형
Fig. 5. Histogram Modification using Dual Binary Tree

그림 5는 부가적인 정보로서 바이너리 트리를 사용하는 방법이며 결과적으로 다수의 상위 빈도 히스토그램에 대한 변형을 수행하는 방법이다. 오른쪽의 이진트리(Binary Tree)만을 이용한다면 Multiple Peaks Histogram Modification이 되며 좌우의 트리를 모두 사용하게 되면 양수와 음수 영역에 대해 각각 Multiple Peaks Histogram Modification을 수행하게 되는 것이다[6-8].

Ⅲ. 제안하는 도약 기반의 하이브리드 히스토그램 시프팅 방법

본 논문에서는 단일 Peak Histogram변형을 기본 참조모델로 사용하며 메시지 삽입에 따른 히스토그램 변형의 위치를 선택적으로 조절하는 도약(Skipping) 방법을 제안한다. 제안 방법은 정보은닉에 사용되지 않는 히스토그램의 시프팅에 있어 최소한의 오류를 발생하도록 시프팅 위치를 선택하도록 설계 되었다. 즉, 동일한 메시지 삽입 시 발생하는 PSNR의 저하를 최소화할 목적을 달성하게 된다. 여기서 히스토그램의 빈도 분포는 그림 6과 같으며 Peak Histogram H(1)에 삽입되는 메시지(M)은 각각 50%의 0과 1로 구성되어 있다고 가정한다. 또 P_0, P_1 각각은 히스토그램에서 첫 번째, 두번째 Index위치를 위미하며 해당 Index에서의 Probability를 의미하며 그림 6에서

는 P_0가 Peak Probability를 가진다는 의미로 H(1)으로 표기를 하였다.

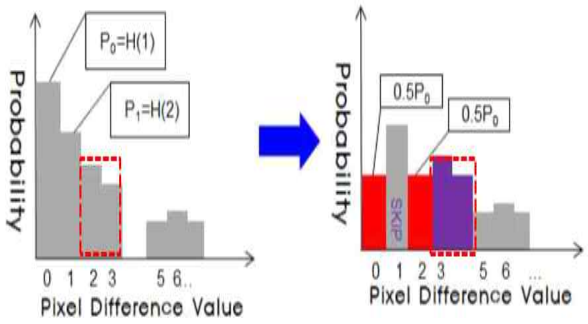

그림 6. 도약을 적용한 히스토그램 변형
Fig. 6. Histogram Modification using Skipping Method

여기서 Idxp는 인덱스 0이며 Idxz는 4이다. 전통적인 방법에서 Idxp의 다음영역인 H(2)~H(4)가 시프트 되어진다. 하지만 제안방법에서는 H(2)를 도약하는 히스토그램 인덱스로 가정하였다. 전통적인 방법을 사용하였을 때의 변형에 의한 오류 MSE(Mean Squared Error)과 도약(Skipping)방법을 적용했을 때의 오류 MSE_S를 구하면 다음과 같다.

$$\begin{aligned} MSE &= 0.5P_0(0^2)+0.5P_0(1^2)+P_1(1^2)+.../(P_0+P_1+...) \\ MSE\_S &= 0.5P_0(0^2)+P_1(0^2)+0.5P_0(2^2)+.../(P_0+P_1+...) \end{aligned} \quad (1)$$

제안하는 방법이 기존의 방법보다 개선되기 위해서는 MSE_S<MSE 가설을 만족하여야 한다. 즉, 식 1을 다시 전개하면 식 2와 같이 P_0가 P_1의 약 66% 이하여야 한다는 결론을 얻게 된다.

$$\begin{aligned} &2P_0/(P_0+P_1+...) < 0.5P_0+P_1/(P_0+P_1+...) \\ &\therefore 1.5P_0 < P_1 \end{aligned} \quad (2)$$

앞에서는 Skipping Step 크기를 1로 가정한 경우이고 Step=2로 가정을 하면 $P_0 < 0.25P_t$ 조건을 만족하여야 한다. 이때 $P_t = H(2)+H(3) = P_1+P_2$이다. 결국 도약의 Step수가 증가할수록$P_0$의 빈도는 상대적으로 감소해야 하므로 가정에 위배되므로 사용이 불가하다. 히스토그램 시프팅이 높은 은닉용량을 위해 최대치의 히스토그램 인덱스를 사용한다는 기본적인 알고리즘에도 위배된다.

그렇다면 동일한 은닉용량을 가지는 일반적인 히스토그램 방법과 도약기법과 2^n Divergence를 병행 사용하는 경우가 대안으로 제시될 수 있다. 제안된 도약기반의 히스토그램 시프팅을 적용하기 위해서는 영상에 삽입될 메시지(M)를 지정된 n-bits 단위별로 10진수로 변환한 후 10진 메시지열의 값을 차이 값 히스토그램 변형에 이용하는 방법이다. 즉, 전통적인 방법에서는 Peak Histogram의 변형이 0, 1단계로 이루어졌지만 제안하는 방법에서는 0, 1, 2, 3의 단계(n=2일 경우)로 이루어진다. 이때 10진 메시지 0~3의 빈도는 각각 25%로 동일하다고 가정한다. 상위 두 개의 히스토그램 Peak(P_0,P_1)의 빈도가 x로 동일하다고 가정하면,

$$\begin{aligned} MSE &= \frac{0.5P_0(0^2)+0.5P_0(1^2)+0.5P_1(1^2)+0.5P_1(2^2)...}{(P_0+P_1+...)} \\ MSE\_D &= \frac{\begin{array}{l}0.25P_0(0^2)+0.25P_0(1^2)+0.25P_0(2^2) \\ +0.25P_0(3^2)+P_1(3^2)+...\end{array}}{(P_0+P_1+...)} \end{aligned} \quad (3)$$

두 개의 Histogram Peak(P_0,P_1) 각각에 기존의 히스토그램 변형방법으로 메시지를 삽입한다면 삽입용량은 $2x$가 되며 MSE=$3x+.../(P_0+P_1+...)$가 된다.

하나의 Peak Histogram P_0에 0~3으로 구성된 10진 메시지를 이용하여 히스토그램을 변형하면 삽입용량은 $2x$가 되며 MSE_D=$12.5x+.../(P_0+P_1+...)$가 된다. 즉, 4*MSE_D≒MSE가 성립하게 된다.

여기서 도약(Skipping)방법을 적용한 경우 발생하는 에러의 추이를 연산해보도록 한다. Skipping Step은 1이며 삽입용량은 위의 실험방법과 동일하다고 가정한다. 단, P_0,$P_1 = x$, $P_{rem} = y$로 가정한다.

$$P_{rem} = \sum P_2 + ... + P_{Idx_Z - 1} \quad (4)$$

두 개의 Peak Histogram(P_0,P_1)에 기존의 히스토그램 변형방법으로 메시지를 삽입한다면 삽입용량은 $2x$가 되며 MSE=$3x+4y.../(P_0+P_1+...)$가 된다.

2^n Divergence 히스토그램 변형 방법과 함께 Skipping Step의 크기를 1로 가정하면 삽입용량은 $2x$가 되며 MSE_D=$6.5x+16y.../(P_0+P_1+...)$가 된다. 즉, 2*MSE_D≒MSE가 성립하게 된다. 대부분의 영상에서 Peak Histogram 주변의 몇 개 인덱스를 제외한 나머지 히스토그램의 빈도 값들은 매우 작다. 그러므로 위의 식에서 발생하는 빈도 $P_{rem}(=y)$ 값에 의한 오류는 무시하였다.

2^n Divergence 히스토그램 변형 방법과 함께 Skipping Step의 크기를 2로 설정하였다. 이때 MSE와 MSE_D의 크기를 비교해보면 MSE=$7x+4y.../(P_0+P_1+...)$이며 MSE_D=$12.5x+25y.../(P_0+P_1+...)$ 이므로 1.8*MSE_D≒MSE가 성립된다. 즉, 동일한 삽입용량을 가지는 경우 Multiple Peak 히스토그램을 사용하는 방법과 2^n Divergence를 가지는 히스토그램 변형방법은 화질의 차이가 크지만 도약(Skipping)을 적용하는 하이브리드 히스토그램 시프팅을 적용하는 경우 Skipping Step크기가 증가함에 따라 화질측정단위를 의미하는

오류(MSE)의 크기가 감소하므로 사용이 가능함을 알 수 있다.

Ⅳ. 결론 및 향후 연구방향

제안한 방법에서는 기존의 히스토그램 변형에 의한 가역 정보은닉 방법에서 삽입용량을 증가시키고 화질저하를 감소시키기 위해 도약(Skipping)을 채용한 히스토그램 변형방법을 사용하였으며 히스토그램의 Multiple Peaks에 메시지를 삽입하는 대신 하나의 Histogram Peak에 2^n divergence를 가지는 메시지 삽입 방법을 제안하였다. 3장의 증명들에 의해 도약 방법만을 이용하는 것은 화질저하 개선에 아무런 효과가 없음을 밝혔으며 2^n Divergence와 함께 도약 방법이 적용되면 화질개선이 발생함을 확인할 수 있었다.

본 논문에서 증명된 사실들을 바탕으로 Skipping Step 크기와 2^n divergence(n-bits binary to decimal number)의 변화에 따라 기존의 히스토그램 변형과 유사하거나 또는 감소된 오류를 생성할 수 있음을 예측할 수 있었다. 또한 본 논문에는 메시지 열에 포함된 원소 값의 빈도가 모두 일정한 것으로 가정하였지만 10진으로 표현된 메시지열의 빈도에 따른 정렬과 같은 추가적인 과정을 사용함으로서 성능향상을 보장할 수 있음을 알 수 있다. 향후 실제 영상들의 데이터베이스에 대해 히스토그램의 분포에 따른 도약기반의 히스토그램 시프팅 적용의 타당성을 실제로 측정해보는 과정이 필요할 것이다.

감사의 글

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2017 R1D1A1B03031465)

참고문헌

[1] Z. Ni, Y.Q. Shi, N. Ansari and W. Su "Reversible Data Hiding," *IEEE Trans. on Circuits and Systems For Video Tech.* , vol. 16, no. 3, pp. 354-362, 2006.

[2] J. Tian, “Reversible Watermarking by Difference Expansion”, *In Proc. of Workshop on Multimedia and Security*, pp. 19-22, December 2002.

[3] H.M. Yoo, S.K. Lee, Y.H. Suh and J.W. Suh “Reversible Data Hiding Using the Histogram Modification of Block Image,” *In Proc. of 16th International Conf. on Neural Information Processing*, Part I, pp. 829-837, 2009.

[4] Vasily Sachnev and YongSoo Choi, "Ternary Bose - Chaudhuri - Hocquenghem (BCH) with t = 2 code for steganography," Journal of DCS, Vol. 17, No. 6, 2016

[5] Juhi Gupta, Priya Gupta, and S.C. Gupta, "Reversible data hiding technique using histogram," Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on, 2015.

[6] W.L. Tai, C.M. Yeh, and C.C. Chang, "Reversible Data Hiding Based on Histogram Modification of Pixel Differences," *Optics Communications*, Vol.285, Issue 2, pp. 101-108, 2012.

[7] Wien Hong, "Adaptive reversible data hiding method based on error energy control and histogram shifting," *Optics Communications*, Vol.285, Issue 2, pp. 101-108, 2012.

[8] Lincy Rachel Mathews and Arathy C. Haran V., "Histogram Shifting Based Reversible Data Hiding," *International Journal of Engineering Trends and Technology*, Vol. 10, Number 10, pp. 481-485, 2015.

[9] Ming Li and Yang Li, "Histogram shifting in encrypted images with public key cryptosystem for reversible data hiding," Signal Processing, Volume 130, pp. 190-196, 2017

최용수(YongSoo Choi)

1998년 강원대학교 제어계측공학과(공학사)
2000년 강원대학교 제어계측공학과(공학석사)
2006년 강원대학교 제어계측공학과(공학박사)

2006년～2007년 연세대학교 첨단융합건설연구단 연구교수.
2007년～2013년 고려대학교 정보보호대학원 연구교수.
2013년～현재 성결대학교 교양교직부(멀티미디어)

※관심분야 : Digital Forensics, Information Hiding, Multimedia Watermarking, Steganography

이달호(DalHo Lee)

1982년 서울대학교 제어계측공학과 공학사
1985년 서울대학교 제어계측공학과 공학석사
1992년 서울대학교 제어계측공학과 공학박사

1992년～현재 가천대학교 전자공학과 교수

※관심분야 : 시스템 식별, 필터링 기법, INS 응용, Data Hiding