

# 런LENGTH 코딩(Run Length Coding) 기법을 이용한 이진 이미지에서의 데이터 은닉

論 文
8-2-6

## Binary Image Data Hiding Using Advanced Run-Length Coding Method

김천식\*, 김형중\*\*

Cheonshik Kim, Hyung-Joong Kim

### Abstract

In this paper, a novel method for data hiding into binary images is proposed. Binary image has two tones: black and white. For this reason, it is not easy to hide messages into binary images. PWLC (pair-wise logical computation) is a new method to hide message into binary images. However, it yields images of unacceptable quality and its decoding process is unclear. Therefore, in order to solve this problem, we adopt the run-length method to achieve advanced reversible data hiding. That is, we find a proper region to hide messages to reduce visual degradation. In this paper, we introduce an improved procedure to hide messages into binary images.

**Keywords :** Data hiding, Steganography, Binary Image, Run-length

### 1. 서 론

데이터 은폐 기술은 고대 그리스의 전쟁터에서 은밀한 통신을 하기 위한 수단을 그 기원으로 하고 있다. 현대 대부분은 사람들은 네트워크 기반의 메일, 메신저 등과 같은 통신기술을 일상적으로 사용하고 있다. 이들 데이터 통신은 암호화를 기반으로 하지 않기 때문에 오늘날의 첨단 기술로 얼마든지 해킹이 가능하다. 따라서, 비밀 통신을 위해서 이미지를 하나의 통신 채널로 사용하기도 한다. 또한, 오늘날 많은 디지털 이진이미지가 수없이 만들어지고 있다. 이들 이진이미지의 권리를 보호하기 위한 방법의 하나로서 데이터 은폐 기술이 하나의 보완 장치가 될 수 있다. 왜냐하면, 이진 디지털 이미지를 소유한 사람이 알 수 없는 형태로 데이터의 원 소유자 정보를 저장

하는 것이 가능하기 때문이다[1].

하지만 데이터 은폐기술 자체는 데이터를 이진 이미지에 아무도 모르게 데이터를 숨기는 데 목적이 있기 때문에 데이터를 어떤 방법으로 저장했는가를 알고 있는 제삼자가 있다면 언제든지 쉽게 그들에게 데이터가 쉽게 노출될 수 있다. 따라서, 이 기술은 암호화 처리를 동반할 때 보다 뛰어난 안전장치가 될 수 있다[2].

이진 이미지의 경우는 흑과 백의 두 가지 색을 이용하여 이미지를 표시하는 방법이기 때문에 이미지 처리기법이 매우 단순한 점 때문에 오히려 이진 이미지에 데이터를 은폐하는 것이 쉽지 않았다. 따라서, 많은 연구자들은 이진 이미지보다 그레이 이미지나 컬러 이미지에 데이터를 은폐하는 것에 관심을 갖고 있고, 또 그 쪽에 논문이 많이 발표되었다. PWLC (Pair-Wise Logical Computation)는 최근의 논문으로서 이진 이미지에 데이터를 은폐하는 방법으로 발표된 논문이다. 이 방법은 여러 단점을 갖고 있다. 가장 큰 문제점은 데이터의 저장을 누구나 알 수 있도록 이미지에 노이즈(noise)를 만든다는 것이다. 따라서,

접수일자 : 2009년 5월 04일

최종완료 : 2009년 6월 22일

\*김천식 : 안양대학교

교신저자, E-mail : [database.lab@gmail.com](mailto:database.lab@gmail.com)

\*\*김형중 : 고려대학교 정보경영공학부

본 논문에서 이러한 문제점을 개선하여 보다 향상된 이진 이미지 형태를 유지하면서 데이터를 은폐하는 방법을 제안하고자 한다.

### II. 관련 연구

PWLC[3,4]은 이진 이미지에 데이터를 은폐하고 원 영상을 복원할 수 있는 기법을 제안한 논문이다. 하지만, 이 방법은 데이터를 정확하게 추출하는데 약간의 문제점을 노출하였다. 그 결과 완전한 영상을 복원하는데 실패할 경우가 있다. PWLC는 어떤 압축 기법도 사용하지 않는데, 이 기법은 호스트 이미지에 데이터를 저장할 때, XOR 이진 연산을 사용한다. 이 방법은 단순히 연속된 "000000" 혹은 "111111"와 같은 이진 이미지의 경계 지점 즉, 백과 흑의 경계 지점을 찾는다.

이 경우 0을 삽입하고 싶다면, "000000"을 "001000"으로 변환하고 1을 삽입하고 싶다면 "000000"을 "001100"와 같이 변환 한다. 또, 이미지의 순서가 "111111"의 경우라면, "110111"혹은 "110011"과 같은 방법으로 0과 1을 은폐할 수 있다.

하지만 이 방법[3,4]은 이진 이미지에 기존의 "000100"과 데이터가 은폐된 "000100"을 구분하는 방법을 언급하지 않고 있다. 따라서, 제안된 이진 이미지로부터 정확한 데이터를 추출하는 것은 매우 제안된 조건임을 알 수 있다. Pixel-wise[5,6] 방법은 dispersed-dot 하프톤 이미지에 적합하다. 그러나, 시각적으로 salt-pepper 노이즈 현상을 볼 수 있다. Block-wise 방법[7]은 호스트 이미지를 적당한 크기로 분할한 다음 각블록의 특징을 변형하는 형태로 데이터를 숨기는 방법이다.

### III. 제안한 방법

본 논문에서 제안한 방법을 아래와 같이 은폐 알고리즘과 데이터 추출알고리즘으로 구성되며 상세한 설명은 각 절에서 설명한다.

#### 1. 데이터 은폐 알고리즘

단계 1: 이진 이미지를 3픽셀 단위로 읽고 이를

10진수로 만든다. 데이터를 읽을 때는 항상 3픽셀 단위로 데이터를 읽되, 중첩(overlap)해서 데이터를 읽는다. 즉, 지금 3픽셀을 읽었다면 다음 3픽셀을 읽을 때, 2픽셀이 중첩된다. 수식 (1)과 같이 처리를 한다.

$$f(g_1, g_2, g_3) = \left[ \sum_{i=1}^n (g \cdot i) \right] \tag{1}$$

where,  $i=1,2,4 \dots n$

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	1	0	0	1	0	1	0	0	1	0	1	0	0
1	1	0	0	0	1	1	1	0	0	0	1	1	0	1

그림 1 이진 이미지 일부

데이터의 은폐위치를 찾기 위해서, 10진수로 만들어진 데이터에 대해서 히스토그램을 구한 후 가장 큰 빈도를 갖는 값과 적은 값과의 쌍을 구하는데, 인접한 픽셀만이 한 쌍이 될 수 있다. 예를 들어서, (1,2), (2,3), (3,4), (4,5), (5,6), (6,7) 등이 쌍으로 구성될 수 있다. 이것 들 중에서 가장 큰 빈도수를 가지면서, 두 쌍의 숫자의 차이가 큰 값을 선택한다. 각 쌍 중에서 가장 큰 값을 L이라고 하고 작은 값을 S라고 놓는다.

단계 2:  $f$ 가 L과 같으면 데이터를 저장하기 적합한 곳이다. 그렇지 않고  $f$ 가 L을 제외한 다른 값일 경우 데이터를 은폐하기에 적합하지 않으므로 다음 3비트를 읽는다. 그림 2는 저장위치를 찾아서 회색 블록으로 표시하였다. 나머지는 저장할 수 없는 공간이다. 그림에서 데이터 저장위치 사이에 하나 혹은 2개 픽셀은 아무것도 저장되지 않은 픽셀이 있다. 이것은 저장위치의 겹침으로 인한 디코드의 문제점을 해결하기 위한 것이다.

이 때, L의 빈도수는 S의 빈도수보다 커야 데이터를 저장 가능함으로 스캔을 할 때는 이러한 조건을 분석하는 과정이 따라야 한다.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	1	0	0	1	0	1	0	0	1	0	1	0	0
1	1	0	0	0	1	1	1	0	0	0	1	1	0	1

그림 2 데이터 저장위치 검색

단계 3: 메시지가 0이고  $f$ 가 L이면  $f$ 는 수정이 필요 없다. 메시지가 1이고  $f$ 가 L이면  $f$ 를 S로 바꾼 다음 map에 이동된 비트를 기록한다. 기

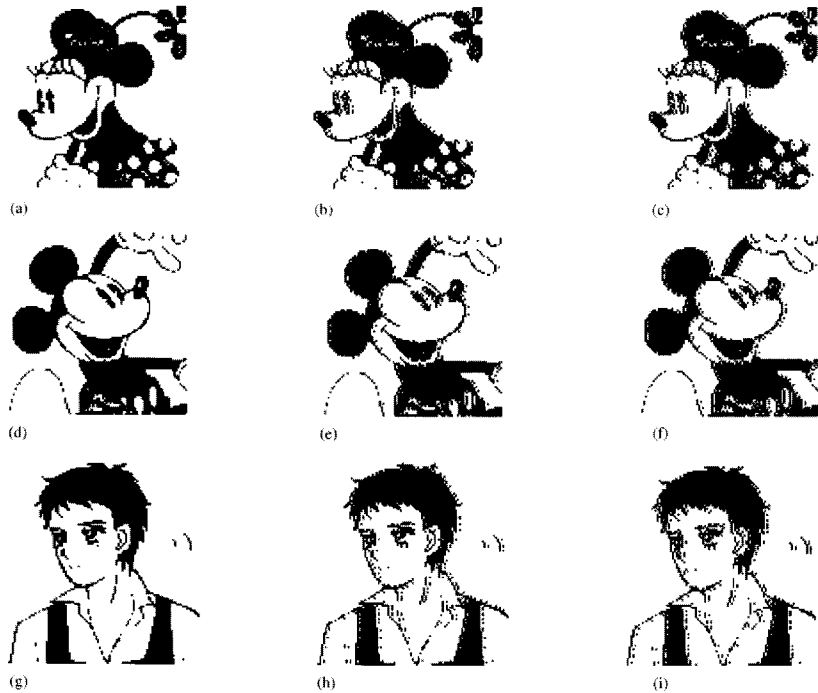


그림 4. PWLC의 실험결과

록하는 이유는 기존의 S와 L에서 S가 된 값을 구분할 수 있도록 하여 디코드 할 때 문제가 생기지 않도록 하기 위함이다.

이미지에 대해서 이와 같은 과정을 반복하여 적용한다.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	1	1	0	1	0	1	0	0	1	0	1	0	0
1	1	1	0	0	1	1	1	0	0	0	1	1	0	1

그림 3 메시지의 은폐 형태

그림 3은 단계 3의 방법에 따라서 데이터를 은폐한 그림으로 진한 회색으로 데이터를 은폐한 모습을 나타낸 것이다.

## 2. 데이터 추출 알고리즘

데이터가 은폐된 이진 Stego이미지로부터 데이터를 추출하는 과정은 데이터를 은폐하는 과정과 유사하다. 단계별로 살펴보면 다음과 같다.

단계 1: 그림 3과 같은 형태의 이진 이미지를 읽어서 역시, 데이터를 은폐할 때와 마찬가지로 연속한 3픽셀을 단위로 읽어서 10진수로 만든다.

단계 2:  $f$ 가 L이면 0이 은폐된 것으로 디코딩을 하고  $f$ 가 S와 같으면서 해당 위치의 맵의 값이 1이면 1이 은폐된 것으로 디코딩을 한다. 전체

단계 3: 맵에 표시해 두었던 곳의 위치를 모두 찾아서 S를 L로 변환 한다. 이와 같은 과정을 반복함으로써 그림 1의 원본이진 데이터를 복원할 수 있다.

## IV. 실험결과

제안한 방법의 기존의 방법보다 개선되었음을 증명하기 위해서 이진이미지를 대상으로 제안한 알고리즘을 다음의 실험환경에서 적용하였다.

본 논문에서는 실험을 위해서 그레이 이미지를 비트맵 이미지로 변환하여 사용하였다. (그림 5)의 a, c, e, g, i, k는 그레이 이미지에 대한 이진 이미지이다. (그림 5)의 o, q는 pdf 문서에 포함된 그림을 스크린 캡처를 통해서 얻어낸 이미지이다. 본 논문에서는 실험을 위해서 Matlab 6.5 버전을 사용하였고, 은폐에 사용한 데이터는 Matlab의 랜덤 함수를 사용하여 데이터를 자동 생성을 통해서 얻었다. 또한, 이미지의 질의 비교하기 위해서 PSNR을 사용하였다. PSNR은 이미지의 왜곡을 측정하는 방법으로 그레이 이미지의 질의 비교하



$$PSNR = 10 \times \log_{10} \frac{255^2}{MSE} \quad (2)$$

식 (2)에서 MSE는 평균 제곱에러를 의미하고, 다음과 같이 계산한다.

$$MSE = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (I(x, y) - I^{\wedge}(x, y))^2}{M \times N} \quad (3)$$

식 (3)에서 I (x, y)은 원본 호스트 이미지이고, I^{\wedge}(x, y)은 stego 이미지를 의미한다.

그림 4는 PWLC 방법에 의해 만들어진 이미지로서, 시각적으로도 이미지가 많이 훼손되었음을 알 수 있다.

표 1. PWLC 기반의 데이터 은폐 성능

이미지	크기	데이터은폐 량
Mickey	96×96	124bits
Minnie	96×96	133bits
Man	108×102	172bits
Mixed text	300×300	6624bits
English text	320×380	17808bits

표 1은 PWLC에 의해서 데이터를 얼마나 은폐 가능한가를 보인 실험결과이다. 본 논문에서 제안한 방법에 의하여 실험한 이미지의 모습이 그림 5와 같다. PWLC 방법보다는 시각적으로 보다 향상되었음을 알 수 있다. 텍스트 이미지의 경우는 텍스트 데이터의 이미지가 매우 민감하기 때문에 약간의 데이터 삽입만으로도 시각적으로 이미지가 훼손 되었음을 쉽게 인식할 수 있다.

표 2는 제안한 알고리즘의 데이터 은폐 능력과 PSNR을 나타내었다. 또한, 데이터를 은폐하기 위해서 부가적으로 사용되는 맵의 크기를 나타내었다. 본 논문에서 제안한 방법이 기존의 방법보다 이미지의 질적인 면에서 매우 향상되었음을 알 수 있다.

표 1과 직접 비교는 가능하지 않지만, 데이터 은폐 량이 상대적으로 크다. 왜냐하면 PWLC는 데이터를 저장하기 위해서 찾아내야하는 경계지점 때문에 미키마우스 이미지와 같이 경계점이 뚜렷하지 않으면 이미지를 삽입하기 어렵다. 그러나, 본 논문에서 제안한 방법은 Baboon과 같은 연속적인 1과 0이 아닌 경우라도 얼마든지 데이터를 삽입하는 것이 가능하다.

표 2. 제안한 방법의 실험결과

이미지	크기	데이터 은폐 량	map크기	PSNR
Lena	512×512	3649	1846	65.0563dB
Baboon	512×512	10000	4957	59.0754dB
Barbara	512×512	7478	3715	61.9086dB
Boat	512×512	6286	3104	62.5545dB
Bridge	512×512	9000	4447	60.5013dB
Mouse	748×600	3408	1734	65.6030dB
Text	629×720	8228	4077	60.0787dB
Text2	708×754	10000	4958	59.7153dB

## V. 결 론

본 논문에서는 PWLC 방법의 문제점인 이미지의 노이즈 문제와 데이터의 은폐 성능을 개선하기 위해서 런LENGTH 코딩기법을 제안하였다. 제안한 방법은 기존의 방법의 문제점인 이미지의 질 저하를 상당한 개선을 하였음을 알 수 있다. 또한, 데이터의 은폐 능력을 향상 시키는 방안으로서 빈도수가 높은 패턴과 빈도수가 적은 패턴을 찾아 데이터 은폐에 이용하였다. 이와 같은 과정을 통해서 데이터의 은폐성능 및 이미지의 질적인 측면에서 많은 개선을 보였다.

향후에는 제안한 방법을 하프톤 영상에 적용하여 제안한 방법의 성능을 및 가능성을 측정하고자 한다.

## [참 고 문 헌]

- [1] M. Steinder, S. Iren, P. D. Amer, "Progressively authenticated image transmission," in *Proceedings of the Military Communications Conference*, vol. 1, pp. 641-645, 1999.
- [2] R. Z. Wang, C. F. Lin, J. C. Lin, "Image hiding by optimal LSB substitution and genetic algorithm," *Pattern Recognition*, vol.34, no. 3, pp. 671-683, 2001.
- [3] C. L. Tsai, K. C. Fan, C. D. Chung, and T. C. Chung, "Data hiding of binary image using Pair-Wise logical computation mechanism," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 2., pp. 951-954, 2004.
- [4] C. L. Tsai, K. C. Fan, C. D. Chung, and T. C. Chuang, "Reversible and lossless data hiding with application in digital library," in *proceedings of the International*

- Carnahan Conference on Security Technology*, pp. 226-232, 2004.
- [5] M. S. fu and O. C. Au, "Data hiding in digital binary image," *IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 393-396, 2000
- [6] M. S. Fu and O. C. Au, "Data hiding by smart pair toggling for Halftone images," in *Proceedings of the Acoustics, Speech, and Signal on IEEE Conference*, vol. 4, pp. 2318-2321, 2000.
- [7] C. S. Pei and J. M. Guo, "Hybrid Pixel-Based data hiding and block based watermarking for Error-Diffused Halftone images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 867-884, 2003.

Biography



**김 천 식**

1997년 한국외국어대학교 컴퓨터및정보통신공학과(공학석사)  
 2003년 한국외국어대학교 컴퓨터및정보통신공학과(공학박사)  
 2000년~2003년 경동대학교 교수  
 2004년~현재 안양대학교 교수

2007년~현재 대한전자공학회 컴퓨터소사이어터 분과위원장  
 2006년~현재 인터넷 정보학회 학회편집위원  
 2006년~현재 대한교통학회 정회원  
 2005년~현재 한국데이터베이스학회 정회원  
 2008년~인터넷방송통신tv학회 상임이사  
 2008년~ICHIT 2008 committee member  
 2009년~ICHIT 2009 committee member  
 2010년~ACIIDS 2010 committee member  
 <관심분야> 데이터베이스, 데이터마이닝, 이미지처리, e-Learning, Agent system  
 <e-mail> database.lab@gmail.com



**김 형 중**

1978년 서울대학교 전기공학과 (공학사)  
 1986년 서울대학교 제어계측공학과(공학석사)  
 1989년 서울대학교 제어계측공학과(공학박사)  
 1992년~1993년 USC 방문교수  
 1989년~2006년 강원대학교 교수  
 2006년~현재 고려대학교 정보경영공학부 교수

<관심분야> 멀티미디어보안, 분산처리, 콘텐츠공학 등  
 <e-mail>khj@korea.ac.kr