한글 텍스트가 내장된 디지털 워터마크 생성 알고리즘

조 대 제† · 김 현 기†

요 약

현재 알려진 워터마킹 방법들은 주로 PN-수열과 같은 잡음 특성을 가진 이진 코드를 워터마크로 사용하고 있는데, 이것은 생성 과정에서 일정한 길이의 이진 코드가 주기적으로 반복되는 단점이 있다. 그러나 혼돈 수열(chaotic sequence)은 기존의 PN-수열을 대체 할 수 있는 새로운 개념의 수열로 여러 가지 장점을 가지고 있어서 워터마크로 사용하기에 적합하다. 본 논문에서는 우리 실정에 적합 할 뿐 아니라, 의미를 바로 알 수 있는 한글 텍스트를 혼돈수열로 변환하여 워터마크로 사용한다. 즉, 한글로 이루어진 저작권 정보를 워터마크로 만들어 영상에 내장하고, 내장된 워터마크를 추출하여 다시 한글로 변환하여 원 저작권자가 누구인지를 알 수 있도록 한다. 한글 텍스트를 워터마크로 사용하는 경우, 내장할 수 있는 텍스트의 길이는 한계가 있다. 그래서 본 논문에서는 한글 텍스트를 단일 실수로 함축하는 방법과 이것을 다시 복원하여 원래의 문장으로 재생하는 알고리즘을 사용하여 한글 텍스트의 길이가 길어지더라도 워터마크로 사용할 수 있도록 하였다.

Digital Watermark Generation Algorithm Embedding Hangul Text

Dae-Jea Cho† · Hyun-Ki Kim†

ABSTRACT

In this paper, we propose the possibility of introducing chaotic sequences into digital watermarking systems as potential substitutes to commonly used pseudo noise sequences. Chaotic sequences have several good properties including the availability of a great number of them, the ease of their generation, as well as their sensitive dependence on their initial conditions. And the quantization does not destroy the good property. So this paper proposes a method that transforms Hangul text to chaotic sequence. And we presents how the Hangul text is expressed by an implied data and the implied data is regenerated into the original text. In this paper, we use this implied Hangul text for watermarking.

키워드 : 워터마크(Watermark), P-N 수열(P-N sequence), 혼돈함수(Chaotic function), 인증(Authentication)

1. 서 론

최근 인터넷의 급속한 확산으로 전 세계가 하나의 컴퓨터 망으로 연결되고 있으며 이와 함께 많은 오디오, 비디오 데이터들이 디지털화되고 있다. 뿐만 아니라 전자 상거래의 발달로 디지털 오디오, 이미지, 비디오 등을 인터넷상에서 구매할 수도 있다. 이것은 디지털 신호의 특성상 손실 없는 복사가 가능하고, 아무리 큰 데이터라 하더라도 거리의 제약 없이 네트워크를 통하여 간단히 전송되기 때문이다. 하지만 디지털 데이터의 속성 때문에 인터넷과 같은 열린 공간에서는 누구나 쉽게 불법적으로 복사하거나, 변형, 재분배 등의 작업을 하는 것이 가능해졌다. 따라서 이러한 행위로부터 원본 데이터를 보호하고, 원본 데이터의 소유권과 저작권을 인정하기 위한 다양한 기술들이 연구되고 있다[1].

인터넷과 같은 네트워크 상에 존재하는 모든 이용자들이 누구나 자유롭게 정보를 이용할 수 있도록 하면서 정보 제공자가 차후에 정보 이용에 대한 권리를 주장할 수 있는 방법으로 워터마킹(watermarking) 기술이 제안되었다. 디지털 워터마킹은 멀티미디어 데이터 안에 다른 사람은 알지 못하게 정보를 넣는 모든 기술을 포함한다. 그리고 여기에 들어가는 정보를 워터마크(watermark)라고 한다. 워터마크의 사용범위는 매우 다양하여 정지 영상과 동영상뿐만 아니라, 전자책(E-book), DVD(Digital Versatile Disk) 등에서는 상용화 단계에 있으며, 현재 불법으로 유통되면서 저작권 문제가 심각하게 대두되고 있는 MP3 형태의 음악 파일에도 적용할 수 있다.

현재 알려진 워터마킹 방법들이 주로 사용하는 워터마크 데이터는 도장 또는 서명과 같은 이진 영상, ASCII 코드, PN-수열(pseudo noise sequence) 등이다. 그러나 이진 영상으로 만들어진 워터마크는 길이가 너무 길어서 원 영상의 해상도를 떨어뜨리는 단점이 있고 ASCII 코드는 너무 단순하고 사용 방법이 많이 알려져 있어서 이용하기 힘들

※ 이 논문은 2003학년도 안동대학교 학술연구 조성비에 의하여 연구되었음.
† 정 회 원 : 안동대학교 전자정보산업학부 교수
논문접수 : 2003년 4월 19일, 심사완료 : 2003년 8월 21일

다. 그래서 PN-수열과 같은 잡음 특성을 가진 이진 코드를 워터마크로 주로 사용하고 있는데, 이것은 생성 과정에서 일정한 길이의 이진 코드가 주기적으로 반복되는 단점이 있다[2]. 그러나 혼돈 수열(chaotic sequence)은 기존의 PN-수열을 대체 할 수 있는 새로운 개념의 수열로 다음과 같은 장점을 가지고 있어서 워터마크로 사용하기에 적합하다[3]. 첫째, 간단한 혼돈 함수의 사용으로 계산이 쉽고 초기 값만 알고 있으면 얼마든지 만들 수 있다. 둘째, 초기 값에 민감하다. 즉, 초기 값을 조금만 변경시켜도 전혀 다른 값을 만들어 낼 수 있으며 길이가 아무리 길어져도 반복되지 않는다. 셋째, 잡음과 같은 무작위(random) 특성을 가지고 있다. 텍스트를 워터마크 데이터로 사용하는 예로는 저작권 정보나 복사 제어 신호를 담은 일련 번호(serial number)만으로 구성하는 방법과 영문 텍스트를 ASCII 코드로 고쳐서 워터마크로 사용하는 경우가 있다. 하지만 본 논문에서는 우리 실정에 맞고 의미를 바로 알 수 있는 한글 텍스트를 혼돈수열로 변환하여 워터마크로 사용한다. 즉, 한글로 이루어진 저작권 정보를 워터마크로 만들어 영상에 내장하고, 내장된 워터마크를 추출하여 다시 한글로 변환하여 원 저작권자가 누구인지를 알 수 있도록 한다. 한글 텍스트를 워터마크로 사용하는 경우, 내장할 수 있는 텍스트의 길이는 한계가 있다. 그래서 본 논문에서는 한글 텍스트를 단일 실수로 함축하는 방법과 이것을 다시 복원하여 원래의 문장으로 재생하는 알고리즘을 사용하여 한글 텍스트의 길이가 길어지더라도 워터마크로 사용할 수 있도록 하였다.

2. 디지털 워터마킹 기법

영상 데이터에 워터마크를 삽입하는 다양한 방법들이 연구되어 왔으며, 이들을 분류하는 기준도 다양하다. 먼저 워터마크를 시각적으로 보이지 않게 영상에 삽입하는 방법과 보이게 워터마킹 하는 방법으로 나눌 수 있다. 시각적으로 보이는 워터마킹 방법은 내장된 워터마크의 제거가 어렵도록 설계되어야 하며 워터마크 자체가 원 영상을 심하게 훼손시키지 않아야 한다. 또한 워터마크를 삽입하는 영역에 따라 분류하는데, 공간 영역에서의 워터마킹(spatial domain based watermarking)과 주파수 영역에서의 워터마킹(frequency domain based watermarking)으로 나눌 수 있다.

공간영역에 대한 워터마크는 초기에 주로 연구되었으며, 대표적인 방법으로는 영상의 화질에 영향이 적은 픽셀의 LSB (Least Significant Bit)에 워터마크를 대체하여 넣는 방법이다. 하지만 이 방법은 압축, 필터링 등 일반적인 영상처리에도 쉽게 지워지고 특히 워터마크가 첨가된 위치를 알고 있으므로 다른 비트 열을 대신 넣을 경우, 워터마크가 바뀌어진다는 단점이 있다. 이런 단점을 제거하기 위해서 여러 다양한 방법들이 연구되었다.

Wong은 암호화 기법을 이용해서 LSB에 워터마크를 첨가하는 방법으로 영상자료에 대한 저작권의 인증과 워터마크의 훼손여부 및 영상의 크기 변경까지 검출하여, 영상이 훼손되지 않았다는 무결성 증명에 이용하였다[4].

일반적으로 공간영역의 방법들은 영상처리기법들이 간단하고 영상에 대한 손실압축, 변형 등에 워터마크가 손상되어 워터마크를 검출하거나 추출하기 어려워서 인위적인 공격에 취약한 단점을 공통적으로 지니고 있으며, 워터마킹된 영상의 시각적인 품질을 보장할 수 없는 경우가 많다. 이러한 공간영역의 방법의 단점을 보완한 방법으로 주파수 영역에서의 워터마킹 방법이 연구되어 왔다. 주파수 영역에 워터마킹하는 방법들은 다시 이산 여현 변환(DCT)를 사용한 방법, 웨이브릿 변환(DWT)을 사용한 방법, 퓨리에 변환(FFT)을 이용하는 방법으로 나눌 수 있다. 기본적인 방법은 영상을 DCT 또는 웨이브릿 변환을 이용하여 주파수 계수를 구하고, 그 계수 값을 변환시켜서 워터마크를 첨가하는 방식이다. Kunder는 DCT 대신에 웨이브릿 변환을 이용해서 블록화 현상이 생기지 않도록 워터마크를 영상에 첨가하는 방법을 제안했다[5].

주파수 영역에서의 워터마킹 방법은 일반적으로 영상을 주파수 변환한 후, 중간대역의 특정 위치의 주파수 계수를 뽑아내어 긴 수열을 만들고, 첨가하는 워터마크에 의해서 각 주파수 계수를 공식에 맞추어 변화시킨 후 원래의 위치에 다시 넣어 역 변환하는 방법을 쓰고 있다. 최근에는 사람이 감지할 수 없는 곳에 워터마크를 내장하기 위하여 인간 시각 시스템(HVS)[6]을 이용하기도 한다. 인간의 시각 시스템은 고주파 성분에 둔감하나, 저주파 성분에 민감한 특성을 나타낸다. 하지만 저주파 성분에 워터마크를 삽입하면 화질의 저하가 생기고, 고주파 성분에 삽입하면 압축과정에 의해 워터마크 신호가 손실된다. 따라서 워터마크 신호가 삽입될 적절한 주파수 영역의 선택은 매우 중요하다. 최근 주파수 영역에서 워터마크를 삽입하는 경우, 저주파와 고주파 사이의 중간 대역에 워터마크 정보를 분산하여 삽입하는 방법이 주로 사용되고 있다.

3. 확정적 혼돈함수를 이용한 워터마크 생성

기존의 워터마킹 방법에서 주로 사용하는 워터마크 데이터는 PN-수열이다. 이것은 공격자가 워터마크의 위치에 대한 정보를 알 수 없도록 하며, 어느 정도 알고 있더라도 정확한 워터마크를 확인하기 어렵기 때문에 제거하기가 힘들다. 뿐만 아니라 워터마킹을 한 영상에다 다시 워터마킹을 하는 다중 워터마킹 공격에도 견고하다는 장점이 있다. 하지만 PN-수열은 이미 널리 사용되었고 사용 방법 또한 노출되었다. 그리고 일정한 길이를 가진 수열이 반복적으로 생

성되는 특징이 있다. 그래서 이것을 대신할 수 있는 잡음 특성을 가지면서도 초기 값을 알면 얼마든지 만들어 낼 수 있는 새로운 함수를 이용하려는 연구가 진행되고 있다. 특히 Xiang[7] 등은 여러 가지 혼돈함수를 이용하여 혼돈 수열을 만들어, 디지털 워터마크로 사용하였다.

본 논문에서는 위에서 설명한 불규칙이고 비 주기적인 성질을 유지하면서도 예측 가능한 확정적 혼돈 함수(deterministic chaos function)를 이용하여 워터마크를 생성한다.

본 논문에서 사용하려고 하는 혼돈 함수는 기본형이 식 (1)과 같으며, 로지스틱 맵(logistic map)이라고 불리는 확정적 혼돈 함수이다.

$$X_{c+1} = \mu X_c(1 - X_c) \tag{1}$$

식 (1)은 μ의 범위가 3.57 보다 크고 4와 같거나 작으면 혼돈 영역(chaotic state)으로 몰입하는 특성을 가진다. 이때 혼돈 예측 함수에서 발생되는 값이 잡음 특성을 가지므로 PN-수열과 매우 유사한 정보 값을 만들 수 있다.

3.1 한글정보의 함축 알고리즘

<표 1>에서 나타낸 변환 표에 의해서 한글 문장을 이산 신호 코드로 변환한다. 그리고 본 논문에서 제시한 알고리즘에 적용하기 좋도록 혼돈 함수의 영역으로 끌어들이기 위하여 식 (2)를 사용하여 0과 1 사이의 실수로 변환한다.

$$\begin{aligned} H &= \{h_1,\ h_2,\ h_3, \cdots,\ h_n\} \ : \text{한글코드집합} \\ G &= T_c(H) \end{aligned} \tag{2}$$

이런 과정을 통해 변환된 G는 g_1, g_2, …, g_n으로 구성된 0과 1사이의 실수들의 집합이다. 물론 이 실수는 함축하려고 하는 한글 문장에 대한 자음과 모음에 대응되는 값이다. 이것을 한 개의 실수로 함축하는 방법은 다음과 같다.

1. The Hangul text is converted to the discrete signal code using Hangul code converting table.
2. The discrete signal code is converted to chaotic area using following function:

$$mhh = h_{max} + h_{min} + 1.0$$
$$g_i = (h_i + h_{min})/mhh$$

Where, h_{max} and h_{min} are maximum and minimum values in the discrete signal code set. And h_i is the i-th discrete signal code.

3. Generate new chaotic sequences using equation 1 with seed value CX_0.
4. Find the closest value CX_1 to g_1 in the chaotic sequences.
5. Compute new seed value by subtracting index value from CX_1.
6. The procedures (3..5) are repeated to all chaotic sequences until final value CX_n is computed.
7. Finally, convert CX_n to binary code.

(그림 1) 데이터 함축 알고리즘

(그림 1)에 나타낸 바와 같이 먼저 식 (1)을 이용하여 임의의 초기 값 CX_0를 입력해서 k번 반복 계산하여 새로운 혼돈(chaos) 값을 만들어 낸다. 이러한 과정을 통하여 만들어진 혼돈 값 중에서 g_1과 같거나 가장 가까운 값을 찾는다. 여기서 반복 횟수를 무한대로 하여 g_1과 가장 유사한 값을 찾을 수도 있으나, 실제 계산에서는 불가능하므로 k번으로 제한한다. 여러 개의 예측 값 중에서 g_1과 같거나 가장 가까운 값이 C_{1n}이라고 가정하면, C_{11}에서 C_{1n}까지 순서 인덱스를 부여하고 C_{1n} 값에서 순서 인덱스를 감산한 값이 새로운 초기 값 CX_1이 된다. 여기서 구해진 초기 값 CX_1을 다시 식 (1)에 입력하여 g_2와 가장 가까운 값을 찾을 때까지 계산을 반복한다. 이런 과정을 통하여 g_2와 같거나 가장 가까운 값 C_{2n}을 구하여 앞에서 계산한 C_{1n}의 순서 인덱스를 포함하여 C_{21}에서 C_{2n}까지 순서 인덱스를 부여하고, C_{2n}에서 순서 인덱스를 감산하여 새로운 초기 값 CX_2를 구한다. 이러한 과정을 n번 반복하여 CX_n을 구하며 이 값이 n개의 데이터를 함축한 하나의 실수가 된다.

즉, G의 각 계층에서의 초기 값 CX_l는 식 (3)과 같으며, 이것을 식 (1)에 n번 적용시켜 최종 값 CX_n을 구하는 것이다.

$$CX_i = C_{ij} - I_l \quad \begin{cases} i = 1, 2, 3, \cdots, n \\ j = 1, 2, 3, \cdots, k \\ l = 1, 2, \cdots, k \times n - 1,\ k \times n \end{cases} \tag{3}$$

이때 최대 인덱스 값이 I_{max}이라면 G의 영역은 예측 값이 발산하거나 특정 값으로 수렴되지 않도록 $I_{max} < G$가 되어야 한다. 여기서 반복 횟수 k의 값이 크면 서로 다른 데이터에 대한 함축 값이 같게 나올 수 있으므로 원치 않는 값을 복원 할 가능성이 있다. 이를 방지하기 위하여 식 (4)를 이용하여 확인 값 P를 구하여 복원 할 때 참조하도록 한다.

$$P = \frac{\sum_{i=1}^{n} g_i}{n} \tag{4}$$

3.2 한글정보의 복원 알고리즘

여러 개의 데이터를 하나의 실수로 함축하는 과정에서 식 (1)을 사용하였지만 복원 과정에서 식 (1)의 역함수를 사용해서는 결과가 나오지 않는다. 그래서 함축된 데이터를 원래의 정보로 복원하기 위하여 함축 과정과 거의 유사한 과정을 다시 수행한다. 즉, (그림 2)에 나타낸 바와 같이 초기 값 CX_0에서 S_{11}을 예측하고 S_{11}에 순서 인덱스를 지정하여

S_{11}에서 순서 인덱스를 감산한 값 SX_1을 구한다. 그리고 SX_1을 초기 값으로 하여 S_{21}을 구하고, S_{21}에 S_{11}의 순서 인덱스를 포함한 S_{21}까지의 순서 인덱스를 지정하여 다음 SX_2을 구한다. 이러한 과정을 n번 반복하여 SX_n을 구하여 함축 값 CX_n과 일치하면 계산 과정에서 나온 예측 값 S_{11}, S_{21}, …, S_{n1} 들의 평균을 구한 후, 확인 값 P와 비교하여 일치하면 식 (2)의 역함수인 식 (5)를 이용하여 변환하면 복원 정보가 된다.

$$h_n = g_n \times (h_{\max} + h_{\min} + 1.0) - h_{\min} \tag{5}$$

복원 정보를 구성하는 값을 구하기 위한 초기 값 SX_i는 식 (6)과 같으며 이것을 식 (1)에 n번 적용시켜 SX_n과 CX_n이 일치하고 확인 값 P가 같아지면, 숨어있던 예측 값을 찾아 일반 함수 영역으로 변환한다.

$$SX_i = S_{ij} - I_l \quad \begin{cases} i = 1,2,3,\cdots,n \\ j = 1,2,3,\cdots,k \\ l = 1,2,\cdots,k\times n-1,\ k\times n \end{cases} \tag{6}$$

만일, 앞의 수행 과정에서 SX_n과 CX_n이 일치하는 값을 찾지 못했다면 $S_{(n-1)1}$에서 S_{n2}, S_{n3}, …, S_{nk}까지 다시 비교한다. 또 일치하지 않으면 $S_{(n-1)2}$에서 구하여진 $SX_{(n-1)}$을 이용하여 일치하는 값을 찾을 때까지 반복한다. (그림 2)의 데이터 복원 알고리즘 가운데 P_{ave}는 각 계층에서 찾아낸 숨은 혼돈 예측 값의 평균을 의미한다. 확인계수 P_{ave}를 사용하는 이유는 수많은 복원 값 중에서 함축 값과 일치하는 값을 찾았지만 실제 값이 아닌 경우를 찾아내기 위함이다.

1. First, extracted watermark data(binary code) is converted to CX_n.
2. Generate chaotic sequence s_{11} using equation 1 with seed value CX_0.
3. Compute SX_1 by subtracting index value from s_{11}.
4. Generate chaotic sequence s_{21} using equation 1 with new seed value SX_1.
5. Compute SX_2 by subtracting index value from s_{21}.
6. The procedure(3..5) are repeated until find SX_n that is equal to CX_n and P is equal to P_{ave}.
7. If SX_n is equal to the implied value CX_n, find prediction values s_{11}, s_{21}, s_{n1},
8. Restore to the discrete signal code using following inverse function :

$$h_n = g_n(h_{\max} + h_{\min} + 1.0) - h_{\min}$$

(그림 2) 데이터 복원 알고리즘

4. 실험 결과 및 고찰

본 논문에서 제시한 알고리즘을 한글 처리에 응용하기 위해서 한글을 구성하는 각 자음과 모음의 변환을 이산신호 코드로 변환하는 변환 코드를 임의로 설정하였다. 변환코드는 'ㄱ'을 10으로 하고 사전 순으로 10씩 증가 시켰으며, 자음은 10에서 190까지 모음은 210부터 340까지 이며 200은 중성 또는 종성이 없는 글자를 표현하기 위한 것이다. 그리고 혼돈 함수로는 식 (1)을 사용하였으며 초기 값은 0.0001로 하였다.

한글로 된 워터마크를 생성하기 위해 실험에서 사용된 단어는 '저작권', '복사방지' 등이며, 이 단어들을 제안된 알고리즘에 적용하기 위하여 변환한 신호코드는 <표 1>과 같다. <표 2>은 위의 문장들을 하나의 함축 값으로 만든 것이다. 표에서 k는 반복 횟수인데, k값이 클수록 원래의 한글 코드와 유사한 값을 찾을 수 있는 확률이 높다.

〈표 1〉 문장 1, 문장 2의 코드 표

문장 1	신호 코드	문장 2	신호 코드
저	90	복	60
	200		250
	230		200
	200		10
작	90	사	70
	200		200
	210		210
	10		200
권	10	방	60
	270		200
	230		210
	20		80
		지	90
			200
			300
			200

〈표 2〉 k값의 변화에 따른 문장 1, 문장 2의 함축 값

문장	함 축 값						
	k = 30	k = 50	k = 70	k = 100	k = 150	k = 200	k = 250
문장 1	0.83552	0.82139	0.79150	0.79641	0.80601	0.80145	0.80464
문장 2	0.64619	0.67359	0.66118	0.67465	0.68129	0.67527	0.67647

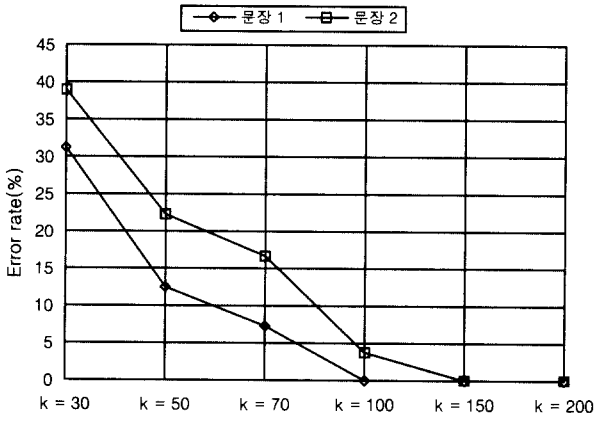
<표 2>에 제시된 두 개의 문장에 대한 함축 값과 중간 예측 값의 평균 값(P)을 가지고, 복원 알고리즘을 이용하여 한글로 복원한 결과는 <표 3>과 같다. 표에서 보는 바와 같이 반복 횟수에 따라 복원 값의 정확도는 차이가 있다. 표에서 'x'는 잘못된 값이 복원된 경우이다.

<표 3>에서 각각의 재생된 글자들 중에서 반복 횟수 k가 30으로 함축한 후 재생하였을 때는 정확한 정보가 하나도 없지만, k가 100일 때는 완전한 정보로 재생되었음을 알 수 있다. (그림 3)에는 문장 1, 2에 대한 에러율을 나타내었다. 그래프에 나타난 바와 같이 원래의 신호코드에 대한 재생된 신호코드의 오차 변화는 반복 횟수 k가 50일 때까지는

빠르게 감소하다가 k가 100, 150에서는 오차가 없는 정확한 정보가 재생된다. 그리고 문장의 길이가 길면 길수록 완전한 문장을 복원하는데 걸리는 시간은 길어진다.

〈표 3〉 함축 값에서 재생된 한글 코드 및 복원된 글자

문장 1	원래 코드	재생신호 코드 및 재생단어											
		k = 30	오류	단어	k = 50	오류	단어	k = 70	오류	단어	k = 100	오류	단어
저	90	90			90			90		자	90		
	200	210	×	자	210	×		210			200		저
	230	230			230		자	230	×		230		
	200	200		ㅓ	200			200		ㅓ	200		
							ㄱ						
작	90	90		ㅈ	90			90			90		
	200	210	×	ㅏ	200			200		작	200		작
	210	210		ㅏ	210			210			210		
	10	30	×	ㄷ	20	×	잔	10			10		
권	10	10			10			10			10		
	270	270			270			270		권	270		권
	230	210	×	관	230		권	230			230		
	20	20			20			20			20		



(그림 3) 문장 1, 문장 2에 대한 반복횟수에 따른 에러율의 변화

이상의 실험 과정에서 만들어진 "저작권"이라는 문장의 함축 값, "0.79641"과 확인 계수, "0.76257"을 <표 4>와 같이 40비트의 이진수로 바꾸어서 워터마크 데이터로 사용한다.

〈표 4〉 한글 텍스트와 변환된 이진 워터마크

내장할 한글 텍스트	함축값 및 확인 계수	이진수로 변환된 워터마크
저 작 권	0.79641 0.76257	0111100101101000000 1 0111011000100101011 1

<표 5>는 Langelaar의 워터마킹 알고리즘[8]에 <표 4>의 워터마크를 내장한 후, 다양한 압축률로 JPEG 압축한 후 추출한 워터마크이다. 품질 계수가 30 이상이면 내장된 한글 문장이 완전하게 재생되는 것을 볼 수 있다.

〈표 5〉 JPEG 압축 후 추출한 워터마크

Quality factor	검 출 결 과	검출율
90, 80, 70 60, 50, 40, 30	0111100101101000000 1 0111011000100101011 1	100%
25	0111101101101000000 1 0111011000110101011 1	95.0%
20	0111000101101100000 1 0111011010100101011 1	92.5%
15	0101100101001000100 01 0111011010101101011 1	87.5%
10	0101100100001011000 1 0111011100100001010 1	80%

5.결 론

본 논문에서는 우리 실정에 맞고 의미를 바로 알 수 있는 한글 텍스트를 혼돈수열로 변환하여 워터마크로 사용하였다. 즉, 한글로 이루어진 저작권 정보를 워터마크로 만들어 영상에 내장하고, 내장된 워터마크를 추출하여 다시 한글로 변환하여 원 저작권자가 누구인지를 알 수 있도록 하였다. 내장되는 워터마크의 크기를 크게 하면 많은 정보를 담을 수 있으나, 화질 저하의 요인이 될 뿐 아니라, 원 영상의 크기에 따라 내장 할 워터마크의 크기에도 한계가 있다. 그래서 본 논문에서는 혼돈함수의 예측성을 이용하여 한글로 이루어진 문장을 단일 실수로 함축하여 워터마크로 사용하는 방법을 제시하고, 함축된 데이터를 다시 원래의 한글 정보로 복원하는 알고리즘을 제안하였다. 그리고 여러 가지 한글 문장에 대한 다양한 실험을 통하여 함축 및 복원 작업을 수행하였으며, 워터마크 데이터로의 사용 가능성을 제시하였다. 또한 제안된 방법으로 만들어진 워터마크를 널리 알려진 Langelaar의 워터마킹 알고리즘에 적용하여 실험 해본 결과, 다양한 압축 공격에도 살아남는 것을 확인하였다.

향후 과제는 본 연구를 바탕으로 하여, 잡음과 기하학적 변환에 강한 워터마킹 방법을 개발하고, MPEG 형태의 동영상과 디지털 오디오에 적합한 워터마크를 생성하는 방안을 연구하는 것이다. MP3 파일과 같은 디지털 오디오가 크게 활성화되고 있는 상황에서 이러한 연구의 필요성이 절실히 요구되고 있다.

참 고 문 헌

[1] I. J. Cox, J. Kilian. T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Images, Audio and Video," *Proc. of IEEE Int. Con. on Image Processing 1996*, Lausanne, Switzerland, Vol.3, pp.243-246, September, 1996.

[2] B. Zhu, M. D. Swanson and A. H. Tewfik, "Transparent Robust Authentication and Distortion Measurement Tech-

nique for Images," *Proc. of IEEE Int. Con. on Image Processing 1996*, Lausanne, Switzerland, Vol.3, pp.211-214, September, 1996.

[3] I. Pitas, "Image Authentication Using Chaotic Mixing Systems," *Proceedings of the 2000 IEEE International Symposium on Circuits and Systems*, Vol.1, pp.216-219, May, 2000.

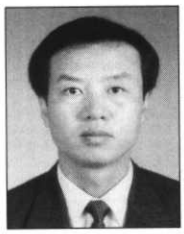
[4] P. W. Wong, "A Public key Watermark for Image verification and Authentication," *Proc. of IEEE Int. Con. on Image Processing 1998*, Chicago Illinois, Vol.1, pp.455-459, October, 1998.

[5] D. Kunder and D. Hatzinakos, "A Robust Digital Image Watermarking Method using Wavelet-Based Fusion," *Proc. of IEEE Int. Con. on Image Processing 1997*, Santa Barbara, CA., Vol.1, pp.544-547, October, 1997.

[6] C. I. Podilchuk, R. Wolfgang and E. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE (USA)*, Vol.87, No.7, pp.1108-1126, July, 1999.

[7] H. Xiang, L. Wang, H. Lin, J. Shi, "Digital Watermarking Systems with Chaotic Sequences," *Proceedings of the Security and Watermarking of Multimedia Contents*, pp. 449-457, January, 1999.

[8] G. C. Langelaar, J. C. A. van der Lubbe and J. Biemond, "Copy Protection for Multimedia Data based on Labeling Techniques," *Proc. of 17'th Symposium on Information Theory in the Benelux*, pp.33-39, May, 1996.



조 대 제

e-mail : djcho@andong.ac.kr

1984년 경북대학교 전자공학과(공학사)

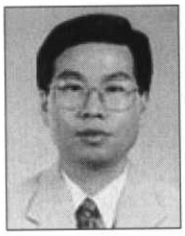
1986년 경북대학교 대학원 전자계산 전공(공학석사)

2001년 경북대학교 대학원 컴퓨터공학과(공학박사)

1989년~2002년 우송공업대학 부교수

2002년~현재 안동대학교 전자정보산업학부 멀티미디어공학 전공 조교수

관심분야 : 멀티미디어응용, 보안, 원격교육 등



김 현 기

e-mail : hkkim@andong.ac.kr

1986년 경북대학교 전자공학과(공학사)

1988년 경북대학교 대학원 전자공학과(공학석사)

2000년 경북대학교 대학원 전자공학과(공학박사)

1988년~1995년 한국전자통신연구원 멀티미디어연구부 선임연구원

1995년~2001년 경남정보대학 전자정보학부 조교수

2002년~현재 안동대학교 전자정보산업학부 멀티미디어공학 전공 조교수

관심분야 : 멀티미디어 시스템, 원격교육, 멀티미디어응용 등