오류 역전파 학습 알고리즘을 이용한 디지털 워터마킹에 대한 소유권 인증

o
최은주, 서정희, 차의영
부산대학교 전자계산학과 신경회로망 및 실세계응용 연구실

Copyright Authentication for Digital Watermarking using Error Backpropagation

Eun-Ju CHOI, Jung-Hee SEO and Eui-Young CHA
Dept. of Computer Science, Pusan National University

요 약

인터넷의 보급으로 인하여 디지털 데이터의 복제가 확산됨에 따라 멀티미디어 데이터에 대한 소유권 보호와 인증에 대한 문제가 대두되고 있는 실정이다. 본 논문에서는 디지털 영상을 다중해상도 표현이 가능한 웨이브릿 변환(Wavelet Transform)을 통하여 저주파수 영역에 인간 시각으로 지각 할 수 없는 워터마크를 삽입하고, 삽입된 워터미크의 영상을 인증하기 위한 방법으로 오류 역전파 학습 알고리즘(Error Backpropagation)을 이용한 신경회로망적 접근방법을 제안한다 워터마크를 추출하기 위해서는 원영상이 필요하고, 내장된 워터마크가 손실 압축과 필터링 등의 일반적인 영상 처리에 강인함을 실험 결과를 통하여 증명하고, 제안한 신경회로망적 접근방법이 좋은 결과를 나타냄을 실험을 통하여 증명하였다

1. 서 론

멀티미디어 저장과 전달 기술의 발달로 증가하고 있는 많은 양의 정보를 디지털 형식으로 저장하고 전송하도록 허용되었고, World Wide Web의 출현으로 확장되었다. 데이터의 디지털화와 멀티미디어의 발달, 그리고 인터넷의 보급으로 인하여 디지털 데이터의 복제가 확산됨에 따라 여러 가지 멀티미디어 데이터에 대한 소유권 문제와 이를 효율적으로 보호할 수 있는 기술이 요구되고 있다

이런 문제점들의 해결 방안의 하나인 디지털 워터마크는 영상의 픽셀에 작은 변형을 가함으로써, 인간 시각으로는 지각할 수 없는 소유권 정보를 영상에 삽입하는 과정으로 비소유권자의 불법적인 조작을 막고 소유권 인증을 위한 방법을 제공해 준다. 워터마킹의 요구조건은 시각적으로 보이지 않아야 하고, 영상의 소유권자가 워터마크를 쉽게 추출할 수 있어야 하며, 고의든 고의가 아니든 signal 변형에 강해야 한다[1].

본 논문에서는 직교 웨이브릿 변환(Forward Wavelet Transform)을 이용하여 저주파수 영역 상에 시각적으로 인식 할 수 없는 워터마크를 내장함으로써 손실 압축이나 필터링, 잡음에 강인한 워터마크를 삽입하는 알고리즘을 제안하고, 삽입된 워터마크 영상의 인증을 위한 방법으로, 신경회로망적 접근방법으로 비선형 판별 문제를 해결할 수 있는 오류 역전파 학습 알고리즘(Error Backpropagation)을 사용한다[2].

신경망을 영상 인식 시스템으로서 패턴 분류기로 사용하는 근본적인 이유는 인간의 학습 패턴을 모방한 것으로 외부로 들어오는 입력 벡터를 보다 효율적으로 사용 할 수 있다는 특성 때문이다.

패턴 분류기에는 통계학적인 모델과 구문론적인 모델, 그리고 신경망 모델로 구분할 수 있는데, 이중에서 신경망 모델은 기존의 인공지능 분야에서 비교적 해결하기 힘든 잡음이나 정보 처리에 매우 유용하며, 동적인 응용분야에서 널리 사용되고 있다.

2. 워터마크 삽입 알고리즘

공간 영역의 디지털 워터마크 기법은 인간 시각 시스템(Human Visual System'HVS)을 이용한 방법으로 인간 시각은 영상의 윤곽선을 잘 추출하는 반면 윤곽선 밝기값의 변화에 민감하지 않다는 것을 이용하여 윤곽선의 밝기값을 변화시킨다 이 방법은 인간 시각으로 지각할 수 없다는 것을 전제로 한 것이지만 일반적인 영상 변형에 강인하지 못하다[3]

주파수 영역의 디지털 워터마크는 인간 시각으로 감지할 수 없는 고주파수 영역에 워터마크를 내장한다 이 방법도 고의적인 영상의 변형, 손실 압축 등과 같은 영상의 왜곡에 워터마크가 손실 될 수 있다[1,4,5]

따라서 본 논문에서는 웨이브릿 변환을 이용하여 정보의 중요한 성분인 저주파수 성분에 워터미크를 삽입하는 방법으로 손실 압축, 고의적인 영상의 변형에 강인한 워터마크를 내장하는 방법을 제안한다.

웨이브릿 변환(Wavelet Transform)은 영상을 피라미드 구조로

표현하는 기존의 방식이 다해상도 분석(multiresolution analysis)과 대역분할 부호화를 하나로 통합한 변환방식으로 이들 부영상들은 각 대역의 평균 신호와 각 방향을 세부적인 신호로 표현한다. 즉, 저주파수 성분, 수평성분, 수직성분, 대각성 성분으로 이루어지며, 고주파 대역에서는 시간 분해능(resolution)을 높이고, 저주파 대역에서는 주파수 분해능을 높이는 옥타브 대역 분할을 통해 시간과 주파수에 대한 국부성(locality)을 가지고 신호를 표현하므로 저주파 성분이 많은 일반 영상신호의 분석에 유리하다[5].

웨이브릿(Wavelet)은 Ψ_x로 정의되고, mother wavelet을 변이시키고 확대, 축소 시킴으로써 얻어지는 함수들의 집합이다 웨이브릿 기저함수는 식(1)로 정의되고, 식(2)는 직교 웨이브릿 변환 함수이다.

$$_{(s,t)}(x) = 2^{-s/2}\psi(2^{-s}x^{-t}) \qquad \text{식(1)}$$

여기서 s는 웨이브릿 기저의 크기이고, t는 웨이브릿 기저의 위치이다.

$$WT = \int_{-\infty}^{\infty} f(x)\Psi_{(s,t)}(x)dx \qquad \text{식(2)}$$

본 논문의 워터마크 내장 알고리즘은 다음과 같다

단계 1 Serial Number(SN)를 생성한다. 즉 소유권자의 정보와 시간 등의 내용이 포함된 워터마크를 말한다.

단계 2. 워터마크를 인간 시각으로 인식 할 수 없게 영상에 내장하기 위해서 임계치(Threshold:T)를 설정한다.

단계 3. 원영상을 FWT에 의해 주파수 영역으로 변환한다.

단계 4. 단계 2에서 설정된 임계치를 적용하여 저주파계수에 워터마크를 내장시킨다. 이 때, 삽입되는 워터마크의 전체 길이는 테스트 이미지에 따라 다른 가변길이(variable lenght)를 가진다. $Y_{(i,j)}$는 워터마크가 내장된 이미지의 계수값, $X_{(i,j)}$는 원영상의 주파수 영역에서의 계수값, W_k는 워터마크 시퀀스(watermark sequence), T는 임계치이다

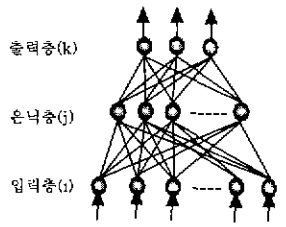
$$Y_{(i,j)}=X_{(i,j)}+W_k \ , \ X_{(i,j)} > T$$

$$X_{(i,j)}, \text{ Otherwise}$$

단계 5. 다시 주파수 영역에서의 이미지를 공간 영역으로 웨이브릿 역변환(Invert Forward Wavelet Transform)을 수행하여 워터마크가 내장된 영상을 생성한다

3. 워터마크 인증

삽입된 워터마크의 영상을 인증하기 위해서 신경망인 오류 역전파 학습 알고리즘이 사용된다. 오류 역전파 학습 알고리즘은 주어진 입력에 대해 목표값을 제시하는 지도학습(supervised learning)으로, 학습시켜야 할 모든 벡터에 대해서 각각에 대한 기대되는 출력(target output)과 실제로 신경망이 산출해 낸 실제 출력(actual output)과의 차이를 학습하기 위하여 연결 가중치(connection weight)를 조절하는 알고리즘이다. 기본 원리를 보면, 입력층의 각 노드에 입력벡터를 제시하면, 이 신호는 각 노드에서 변환되어 은닉층에 전달되고 최후에 출력층에서 신호를 출력하게 된다. 이 출력값과 기대값을 비교하여 차이를 줄여나가는 방향으로 연결강도를 조절하고, 상위층에서 역전파하여 하위층에서는 이를 근거로 다시 자기층의 연결강도를 조정해나간다[6]. 오류 역전파 학습 모델의 구조는 <그림1>과 같다.



<그림1> 오류 역전파 학습모델의 구조

워터마크 영상의 인증을 위한 오류 역전파 학습 알고리즘은 다음과 같다.

단계 1. Network을 구성, 오류한계(error criteria) 및 Weight와 biases를 설정한다

단계 2. 입력 벡터와 목표값(target value)을 설정한다

단계 3. 각 노드에 대한 출력값을 계산한다. 실제 출력을 산출하기 위한 netinput으로는 식(3)과 같다.

$$F(net_i) = F(\sum_j w_{ij} o_j + \theta_i) \qquad \text{식(3)}$$

단, θ_i는 node$_i$의 bias

비선형 활성함수($F(net_i)$)로는 미분 가능하고 단조증가 함수인 시그모이드 함수를 사용한다.

단계 4. biases와 Weight를 조절한다.

출력층과 은닉층의 노드에 대한 오차신호 δ_i는 출력층의 경우 $\delta_i = (t_i - o_i)o_i(1-o_i)$이고, 은닉층의 경우 $\delta_i = o_i(1-o_i)\sum \delta_k w_{ki}$가 되어 은닉층의 노드에 대해서는 재귀적 연산이 이루어지게 된다 결과적으로 가중치의 변화량은 식(4)와 같다.

$$w(n+1) = \eta\delta_i o_i + \alpha\Delta w_{ij}(n) \qquad \text{식(4)}$$

단, n은 학습시간, η는 학습률,

α는 모멘텀, δ_i는 노드의 오차

단계 5. 학습은 총오류자승합(TSS:Total Sum of Square)이 오류한계(error criteria)를 만족할 때까지 단계 2로 가서 반복학습하고 그렇지 않으면 학습을 끝내고 Weight와 biases를 저장한다.

본 논문에서는 학습시 입력 값으로 그레이 영상을 웨이브릿 변환하여 16*16 크기의 원영상과 처리영상들과의 차로 구한 값들을 4개의 부분벡터들로 분할, 정규화(normalization) 시켜서 64개의 계수값을 하나의 입력벡터로 구성하였다. 은닉층의 노드수는 너무 많은 노드수를 정할 경우 overfitting이, 반대의 경우는 underfitting을 일으킬 수 있으므로 오류 역전파 학습모델의 구조에 많은 영향을 미친다. 본 논문에서는 실험을 통하여 은닉층의 노드수를 32개로 설정하였다. 출력층의 노드수는 4개로 설정하였고, Target output의 값은 각각 1000, 0100, 0010, 0001으로 하였다 학습률은 0.5, 초기 가중치의 범위는 [-1,1]이고, 오류 한계값은 0.1로 정하였다. 각 학습 단계에서 특정 노드의 가중치가 다른 노드들의 가중치들과 상치되어 더 이상 학습이 진전되지 않는 경쟁단계로 인하여 지역 최소화 문제가 발생하고, 오류 역전파 학습 구조가 완전연결망(fully connected Network)이기 때문에 정체 현상이 발생한다. 따라서 이러한 문제점을 위와같이 값을 설정함

으로써 네트워크의 구조를 개선시켰다.

4. 실험환경 및 결과

본 논문은 실험환경 Pentium-200Mhz 프로세스와 Visual C++ 5.0/Matlab 5.0으로 시뮬레이션 프로그램을 작성하고 256*256 그레이 영상을 사용하여 실험하였다.

먼저, 256*256 크기의 그레이 영상에 웨이브릿 변환을 이용하여 저주파 성분 중 인간 시각으로 지각할 수 없는 위치에 워터마크를 삽입하였고, 삽입된 워터마크가 손실 압축이나 잡음, 필터링에 강한 것을 확인하였다. <그림2>는 원영상, <그림3>은 워터마크영상, <그림4>는 압축된 영상을 나타낸다

<그림2> 원영상

<그림3> 워터마크 영상

<그림4> 워터마킹 후 95% 압축된 영상

그리고 워터마크가 삽입된 영상의 인증을 위하여 오류 역전파 학습 알고리즘을 이용하였다 오류 역전파 알고리즘에 적용하기 위한 학습 데이터로는 웨이브릿 변환을 통하여 16*16으로 레졸루션하여 워터마크가 내장된 픽셀 영상의 압축이나 필터링이 가해진 영상과 원영상을 이용하여 워터마크를 추출한 다음 정규화하여 학습시켰다 실험 결과 워터마크가 들어있는 영상에 일반적인 영상처리를 가했더라도 워터마크 벡터를 검출하였다

<표1> 워터마크가 들어있는 Lena 영상

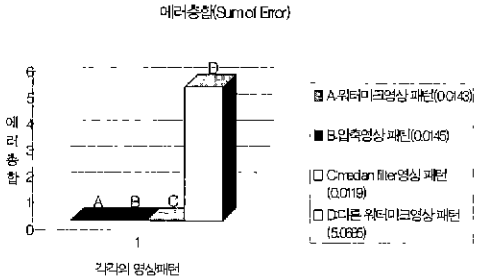
16*16의 부분벡터	target output값				워터마킹 후 압축된 영상				워터마킹 후 filter된 영상			
S1	1	0	0	0	.9980	0011	0003	.0007	.9990	.0005	.0007	.0008
S2	0	1	0	0	0016	9958	0003	.0011	.0003	.9984	.0004	.0012
S3	0	0	1	0	0016	0007	9997	0	.0009	0009	.9998	0
S4	0	0	0	1	0003	0018	0003	.9993	.0015	.0008	0	9997

<표2> 다른 워터마크가 들어있는 Lena 영상

16*16의 부분벡터	target output 값				다른 워터마크가 들어있는 경우			
S1	1	0	0	0	0543	4654	0029	.0030
S2	0	1	0	0	.8706	.7048	0013	.0010
S3	0	0	1	0	.1263	.1520	0390	0002
S4	0	0	0	1	0555	0097	0805	0013

<표1>과 <표2>는 Lena 영상에 대한 결과를 보인 것이다. 워터마킹 후 압축된 영상이나 필터링이 가해진 영상에서도 워터마크 벡터가 정확하게 검출되는 결과를 얻었고, 다른 워터미크가 들어있는 경우 워터마크 벡터는 검출되지 않았다 따라서 워터마크된 영상의 소유권을 인증할 수 있다.

<그림5>는 <표1>과 <표2> 등에서 얻은 출력값들의 에러 총합을 나타낸 것이다 워터미크가 들어있는 경우 에러 총합이 0에 가까운 값을 나타내고, 다른 워터마크가 들어있는 경우 대체로 5이상인 큰값을 나타낸다.



<그림5> lena영상의 패턴 영상에 대한 에러총합들

5. 결론 및 향후연구과제

본 논문에서는 주파수 영역의 다해상도에서 시각적으로 보이지 않는 부분에 워터마크를 삽입하고, 손실 압축이나 필터링 등 일반적인 영상 처리 및 잡음에도 강인함을 확인하였다. 그리고 삽입된 워터마크를 인증하기 위한 방법으로 신경회로망적 접근방법을 사용하였다

지금까지의 대부분의 연구논문에서는 영상의 인증이나 워터마크 추출을 위하여 통계학적인 방법[1,7]을 사용하였으나, 본 논문에서는 신경망을 적용하여 영상의 소유권을 인증하였고, 실험을 통하여 소유권을 주장할 수 있는 개선된 결과를 얻었다. 따라서 영상의 소유권 인증뿐만 아니라 워터마크를 추출도 통계학적인 방법보다 신경망을 이용하면 개선된 결과를 얻을 수 있을 것으로 기대된다. 따라서 향후 과제로는 원영상 없이도 워터마크 추출과 워터마크된 영상을 인증 할 수 있는 새로운 신경회로망적 접근 방법의 연구가 필요하다.

참고문헌

1 M D Swanson, Bin Zhu, A H Tewfik, *"Transparent Robust Image Watermarking,"* Proc ICEE ICIP, Vol 3, Sep, pp 211-214, 1996

2 오칭석, *뉴로 컴퓨터*, 기성출판사, pp 171-229, 1996

3 R B Woltgang, E J Delp, *"A Watermark for Digital Image,"* Proc IEEE ICIP, Vol 3, pp 219-222, 1996

4 D Kundur, D Hatzinakos, *"A Robust Digital Image Watermarking Method using Wavelet-Based Fusion,"* Proc. IEEE ICIP, Santa Barbara, California, Vol 1, pp 544-547, Oct, 1997

5 X Xia, C G Boncelet, G R Arce, *"A Multiresolution Watermark for Digital Images,"* proc IEEE ICIP, Vol 3, pp 548-551, 1997

6 Michael Chester, *Neural Network*, Prentice-Hall, pp 50 65, 1993

7. B. Tao, B Dickinson, *"Adaptive Watermarking in the DCT Domain,"* IEEE ICASSP, Vol 4, Apr, pp 2985-2988, 1997

8 김덕령, 박성한, *"디지털 영상의 스유권 보호를 위한 적응 워터마킹 기법,"* 제10회 신호처리합동학술대회 논문집 제10권 1호 pp 1133-1136, 1997