비디오 워터마킹에 대한 프레임 공격 구현

이혜주°, 홍진우
한국전자통신연구원 무선방송연구소 방송미디어부

Implementation of Frame Attacks against Video Watermarking

Hyejoo Lee, Jinwoo Hong
Dept. of Broadcasting Media, Radio & Broadcasting Research Laboratory,
Electronics and Telecommunications Research Institute

요 약

비디오 워터마킹 시스템은 이미지 워터마킹과 비교할 때 프레임 공격이라는 새로운 공격 형태가 발생된다. 프레임 공격은 워터마크를 공격하기 위하여 비디오 프레임들을 적절하게 이용하는 방법으로, 지금까지 제시된 프레임 공격들을 살펴보면 프레임 제거, 프레임 재배치, 프레임 평균화, 프레임 공모(결탁) 공격들을 예로 들 수 있다. 그러나, 프레임 공격에 대한 인식은 있지만, 그 구현 방법에 대해서는 일부 문헌을 제외하고는 논의되고 있지 않다. 따라서, 본 논문에서는 각 프레임 공격들의 목적과 그 구현 방법을 기술한다. 본 논문에서 구현된 프레임 공격들은 설계한 비디오 워터마킹 시스템의 강인성 검사에 활용 가능할 것이다.

1. 서론

비디오 워터마킹 시스템은 이미지 워터마킹에 비하여 상대적으로 많은 연구가 이루어지지 않았다[1-7]. 그 이유는 비디오 데이터 자체가 프레임으로 이루어져 프레임을 하나의 정지영상으로 간주하여 이미지 워터마킹을 그대로 이용하는 것이 가능하기 때문이다. 그러나, 비디오는 시간적인 흐름을 가지고 있어 이미지 워터마킹과 다른 몇 가지 점들을 고려하여야 한다. 예를 들어, 프레임마다 독립적으로 삽입되는 워터마크에 의해 깜빡임 현상(flickering effect)을 초래할 수 있고, DFT(discrete fourier transform)와 같은 많은 계산량을 갖는 변환 처리를 이용하는 경우에는 실시간 처리시에 이용하기가 어렵다는 점이다. 또한, 공격의 측면에 있어서 비디오 워터마킹은 프레임들을 이용한 새로운 공격 형태가 가능하다. 본 논문에서는 이러한 것들 중에서 프레임을 이용한 프레임 공격(frame attack)에 대해서 논의하고자 한다. 먼저, 2장에서는 비디오 워터마킹 기술과 관련된 프레임 공격에 관한 연구들을 기술하고, 3장에서 각 프레임 공격들에 대한 구현 방법들을 기술한다. 그리고, 4장에서는 각 프레임 공격에 의한 비디오 재생 시의 영향에 대해서 기술한다. 마지막으로 5장에서는 향후 연구 과제로 결론을 맺는다.

2. 프레임 공격

정지 영상과 달리 비디오 프레임들은 연속된 프레임들간 높은 유사성(similarity)을 지니고 있다. 이러한 유사성을 시간적 중복성(temporal redundancy)이라고 하는데, 연속되는 비디오 프레임에서 임의의 프레임을 제거하거나 혹은 프레임의 순서를 약간 변경시키는 것은 비디오의 품질에 많은 영향을 주지 않는다. 이와 같은 프레임간의 유사성을 주목하여 워터마크가 삽입된 비디오 프레임들로부터 워터마크를 제거하거나 워터마크의 검출을 혼란시켜 워터마크 검출 시스템의 성능을 저하시키는 공격들을 프레임 공격(frame attack)이라고 한다.

지금까지 제시된 비디오 워터마킹 기술 중에서 Langelaar 방식과 같이 압축 비트 스트림에 워터마크를 삽입하거나[2], Hartung 방식과 같이 압축된 비트 스트림을 부분 디코딩하여 워터마크를 삽입하는 방법[3]은 프레임 공격들을 고려하지 않았기 때문에 프레임 공격에 대한 강인성을 확신할 수 없다. 반면에 Swanson의 방법은 프레임 공격을 고려하여 시간적 웨이블릿 변환(temporal wavelet transform)을 적용하여 프레임 평균화에 대한 강인성을 고려하고 있으나[4], 워터마킹에 대한 많은 계산량이 요구되기 때문에 실시간 응용에는 적합하지 않다.

비디오에 대한 프레임 공격과 비교하여, 정지영상에 대한 기하학적 공격(geometrical attack)과 같이 공간적(spatial)으로 워터마크의 동기를 없애버리는 동기화 공격(synchronization attack)에 대해 프레임 공격은 시간적(temporal)으로 워터마크의 동기를 없애버리는 특징을 가지게 된다. 이외에도 워터마크를 예측하여 워터마크를 제거하는 공격(watermark removal attack)도 가능하다.

이와 같은 공격의 형태에 대해 각 프레임 공격의 관계를 기술하고, 그 구현 방법들을 아래에 기술한다.

3. 프레임 공격의 구현

프레임 공격의 종류로써, 프레임 제거(frame removal), 프레임 재배치(frame replacement), 프레임 평균화(frame averaging), 프레임 공모(frame collusion) 공격들이 있다. 각 프레임 공격 방법들을 기술하기 앞서, 프레임 공격을 수행하는 공격자의 목적은 공격한 비디오 프레임의 가치를 가능한 떨어뜨리지 않는 범위

에서 워터마크의 검출이 불가능하게 하는 것이다. 그러나, 만일 워터마크의 검출이 불가능하더라도 비디오 프레임의 화질 저하가 발생한다면 이 경우에는 공격이 성공적으로 이루어졌다고 할 수 없다. 따라서, 공격자는 이를 위해 프레임 공격을 수행할 때 공격에 의해 발생되는 화질의 저하가 최소화될 수 있도록 공격 가능한 프레임을 선택하는 방법을 이용할 수 있을 것이다. 이러한 부분들을 계산하기 위한 방법으로 본 논문에서는 프레임간의 유사성을 이용하기 위하여 MSE(mean squared error)를 공격의 척도로써 이용한다.

3.1 프레임 제거 공격

프레임 제거 공격은 워터마킹 기법에 따라 효과가 달라지게 된다. 모든 프레임에 동일한 워터마크를 삽입하는 경우에는 어떤 한 프레임을 제거하더라도 다른 프레임으로부터 충분히 워터마크 검출이 가능하다. 그러나, 각 프레임에 랜덤한 성질을 갖는 독립적인 워터마크가 삽입되는 경우, 프레임 제거 공격은 동기화 공격의 형태가 되어 시간적으로 발생되는 워터마크과 동기가 일치하지 않기 때문에, 이 경우에는 워터마크 검출 성능이 공격의 영향을 받게 된다.

프레임 제거 공격은 가능한 화질 저하를 최소화하기 위하여 현재 프레임에 대하여 이후(혹은 이전) 프레임 간의 차이가 작은 프레임을 제거하는 것이 바람직할 것이다. 크기가 $M \times N$ 인 프레임들의 시퀀스에 대해 크기가 K 인 윈도우를 설정하고, 각 윈도우 내의 프레임들에 대해 2개의 프레임의 MSE를

$$d_k = \frac{1}{NM} \sum_{0 \le m < M, 0 \le n < N} | f_k(m,n) - f_{k+1}(m,n) |^2 \qquad (1)$$

와 같이 계산한다. 이때, $0 \le k < K-1$ 이다. 계산된 MSE d_k 를 이용하여

$$d_{\min} = \min_{0 \le k < K-1}(d_k) \qquad (2)$$

와 같이 최소 MSE를 가지는 2개의 프레임을 결정한다. 그리고 결정된 2개의 프레임 중 하나를 윈도우 내의 프레임 시퀀스로부터 제거하고 나머지의 프레임들을 출력하게 된다.

이후의 비디오 프레임들에 동일하게 윈도우를 중첩하지 않고, K 개의 프레임에 대하여 위 과정을 수행한다.

3.2 프레임 재배치 공격

프레임 제거 공격에서 기술한 바와 같이, 프레임 재배치 공격도 워터마크 삽입 방법에 따라 공격의 영향이 달라진다. 워터마크가 모든 프레임에 동일하게 삽입되는 경우, 프레임 재배치 공격의 영향은 받지 않는다. 그러나, 각 프레임에 독립적으로 랜덤한 워터마크가 삽입되는 경우에는 워터마크 재배치 공격은 시간적(temporal) 워터마크 동기화 공격이 되며, 워터마크 검출이 워터마크간의 동기가 일치되지 않기 때문에 공격의 영향을 받게 된다.

프레임 재배치 공격은 프레임 시퀀스에서 프레임의 순서를 변경하는 공격이다. 움직임이 있는 부분에서 재배치 공격이 이루어진다면 공격 이후, 비디오 재생시 움직임들이 시간적으로 앞뒤로 흔들리는 경우가 발생할 것이다. 따라서, 공격자들은 이러한 부분들에는 가능한 공격을 적용하지 않을 것이다.

기본적으로 프레임 재배치 공격은 프레임 제거 공격과 같이 크기가 K 인 서로 중첩되지 않는 윈도우를 설정하고, 윈도우 내에서 식(1)과 같이 d_k 를 계산한다. 계산된 d_k 로부터 최소 MSE $d_{\min}$ 을 갖는 2개의 프레임을 결정한다. 그러나, 결정한 두 프레임의 위치를 서로 변경함에 따라 발생되는 재생 시의 흔들림 현상이 심하지 않기 위하여, 최소 MSE $d_{\min}$ 과 미리 주어진 임계값 τ 에 대해 $d_{\min} \le \tau$ 을 만족하는 경우에만 두 프레임의 위치를 서로 교환하여 비디오 시퀀스를 출력한다.

이후의 비디오 프레임들에 대해서도 윈도우를 중첩하지 않고, K 개의 프레임에 대하여 위 과정을 수행한다.

3.3 프레임 평균화 공격

프레임 평균화 공격은 동일한 워터마크를 프레임에 삽입하는 경우에는 워터마크의 에너지를 감소시키는 효과를 주기 때문에 프레임 제거나 재배치 공격과 비교하면 어느 정도 공격의 영향을 받게 된다. 그러나, 각 프레임마다 서로 독립인 워터마크가 삽입되는 경우에는 통계적으로 서로 독립인 워터마크의 합은 0에 가깝기 때문에 프레임 평균화 공격은 워터마크의 검출 성능을 저하시키는데 효과적이라고 할 수 있다. 따라서, 프레임 평균화 공격은 흔히 통계적 성질을 이용한 워터마크 제거 공격의 형태라고 한다. 프레임 평균화 공격은 비디오 프레임들을 평균화하는 것으로 평균화 공격의 적용 시 비디오 프레임들을 시간적 저역통과(temporal low-pass) 필터링과 유사하게 흐려짐(blurring)을 초래한다. 특히, 움직임이 많은 프레임들에 이러한 프레임 평균화를 수행하면 화질의 저하는 더 심각하다. 따라서, 공격자는 프레임 간의 움직임이 적은 프레임들에 프레임 평균화를 수행하여야 하며, 이를 기반으로 프레임 평균화 공격에서도 MSE를 척도로써 이용하게 된다.

먼저, 크기가 K 인 윈도우를 설정하고 윈도우 내의 프레임 간 MSE d_k 를 식(1)을 이용하여 계산한다. 평균화 공격은 주어진 윈도우 내의 프레임들을 평균화하고 그 결과를 윈도우 내의 첫 프레임을 대신하여 출력하게 된다. 프레임 평균화를 수행하기 위해, 윈도우 내에서 계산된 d_k 에 대해서

$$d_{sum} = \sum_{0 \le k < K-1} d_k \qquad (3)$$

와 같이 MSE의 합을 계산한다. 윈도우 내의 프레임에 평균화를 수행할 것인가를 결정하기 위해, 미리 정의된 임계값 θ 에 대해서, $d_{sum} \le \theta$ 를 만족한다면,윈도우 내 모든 프레임에 대하여

$$f'(m,n) = \frac{1}{K} \sum_{0 \le k < K-1} f_k(m,n) \qquad (4)$$

와 같이 평균화된 프레임을 구하고, 이 프레임을 윈도위 내의 첫번째 프레임과 대체시킨다. 이때, 임계값 θ 는 MSE의 평균값으로 설정하거나, 임의의 값으로 설정할 수 있다. 프레임 평균화 공격은 앞의 두 공격과 달리 윈도우를 중첩하여 프레임 시퀀스에 순차적으로 평균화 공격을 적용하여 수행한다.

3.4 프레임 공모 공격

프레임 공모(결탁) 공격은 앞에서 기술한 공격들과 달리 각

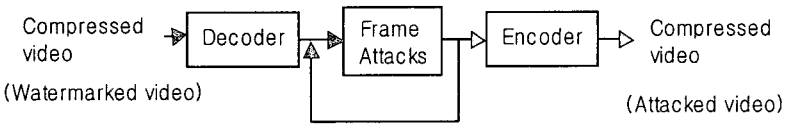

그림 1. 프레임 공격 과정

프레임에 서로 동일한 워터마크를 삽입하는 경우에 효과적임이 알려져 있는데, 프레임 공모 공격은 다수의 프레임으로부터 워터마크의 통계적 성질을 알아내어 프레임으로부터 워터마크를 제거하는 공격 방식이기 때문이다. 워터마크의 통계적 성질을 알아내기 위해 공모 공격에서는 예측 기법을 기본적으로 이용하게 된다. 워터마크를 예측하기 위하여 워터마크가 삽입된 각 프레임 f_t 에 대한 참조(예측) 프레임 f_t' 를 구성한다. 구성된 참조 프레임 f_t' 에 대해서 $f_t' - f_t = w_t'$ 를 계산한다. 각 프레임으로부터 계산된 w_t' 을 이용하여 워터마크의 분산 및 평균을 계산하고, 그 통계적 성질을 갖는 워터마크를 예측하게 된다. 따라서, 서로 동일한 워터마크가 모든 프레임에 삽입되는 경우 통계적 성질을 알아내기는 어렵지 않다.

본 논문에서는 참조 프레임을 구성하기 위하여 저역 통과 필터링이나 Wiener 필터링을 이용하지 않고 보다 간단하게 MPEG 비디오 압축 시 이용하는 움직임 예측을 워터마크가 삽입된 프레임에 대해 수행하여 움직임 보상 프레임(frame compensated frame)을 구성하고, 이를 참조 프레임으로 사용한다. 워터마크의 예측은 다음과 같이 수행한다. 먼저, 프레임 f_t 에 대해 MPEG 비디오 압축의 양방향 움직임 예측을 이용하여 참조 프레임 f_t' 를 구한다. 프레임 f_t 와 구성된 참조 프레임 f_t' 로부터

$$\hat{w}(m,n) = \frac{1}{N_f}\sum_{i=0}^{N_f} f_i'(m,n) - f_i(m,n) \qquad (5)$$

와 같이 계산한다. 이때, N_f 는 프레임의 개수이다. 예측된 워터마크는

$$\tilde{w}(m,n) = \frac{1}{\sigma_{\hat{w}}}\hat{w}(m,n) - \mu_{\hat{w}} \qquad (6)$$

와 같이 유도된다. 이때, $\mu_{\hat{w}}$ 와 $\sigma_{\hat{w}}$ 는 각각 $\hat{w}$ 의 평균 및 분산이다. 식(6)에 의해 예측된 워터마크는

$$\tilde{f}_t(m,n) = f_t(m,n) - \lambda\tilde{w}(m,n) \qquad (7)$$

에 의해 워터마크가 삽입된 프레임으로부터 제거하게 되고, 공격된 결과의 프레임 $\tilde{f}_t$ 를 출력하게 된다. 이때, λ 는 공격의 강도를 조절하는 인수이다.

4. 실험 및 결과

프레임 공격은 송신 측(워터마크 삽입자)에서 워터마크를 삽입되고 압축이 수행된 비디오 데이터에 수행되어진다. 공격자는 압축되어 있는 비디오를 복호하여 프레임을 구성한 후, 자신이 원하는 부분에 프레임 공격을 취하고 다시 재압축을 수행하게 된다. 그림1은 이 과정을 나타내고 있다. 따라서, 프레임 공격에 대한 강인성은 프레임 공격에 대한 강인성 뿐만 아니라 재압축에 대한 강인성을 모두 포함하여야 한다. 공격의 효율성은 공격에 대한 인지도와 워터마크 검출 성능으로 측정할 수 있다. 공격에 대한 인지도란 비디오 시퀀스를 재생할 때 공격이 있었는가를 인지할 수 있는가를 의미한다. 공격의 인지도가 낮으면서 워터마크 검출 성능이 떨어지면 공격의 효율성은 증가되며, 워터마크 검출 성능이 떨어지고 공격의 인지도가 높은 경우에는 공격의 효과는 없다고 판단된다. 따라서, 프레임 공격을 수행할 때, 두 요소의 관계를 적절하게 고려하여야 하며, 그림1의 피드백은 이 점을 의미한다.

본 논문에서 제안한 프레임 공격들에 대하여 실험용 비디오 시퀀스로써 "Football(크기:720×480. YUV(4:2:0))"을 이용하여 $N_f = 30$ 개의 비디오 프레임에 프레임 공격을 수행하였다. 비디오 데이터의 압축은 MPEG-2 압축을 이용하고, 압축된 비디오를 복호한 후, 앞에서 기술한 각 프레임 공격을 적용하였다. 여기서 사용하는 워터마크 방법은 Dequilaume 이나 Lin 등이 지적한 바와 같이[6,7] 평균화 공격이나 공모 공격에 대한 강인성을 제공하기 위하여 워터마크는 일정 길이를 갖는 구간의 프레임에는 동일한 워터마크를 삽입하고, 서로 다른 구간에서는 다른 워터마크를 삽입하는 방식으로 삽입하였다[8]. 각 공격은 윈도우 크기에 따라 공격의 효과가 달라진다. 제거 공격과 재배치 공격은 주어진 윈도의 크기가 증가하면 제거나 순서가 변경되는 프레임 수가 감소하기 때문에 공격의 인지도는 떨어진다. 그러나, 워터마크 검출 성능에는 많은 영향을 주지 못한다. 평균화 공격은 윈도우 크기 증가하면 많은 수의 프레임이 평균화되기 때문에 워터마크 검출 성능이 떨어지지만, 화질 저하가 많기 때문에 공격의 영향이 적다고 할 수 있다. 워터마크를 삽입하고 7Mbps의 bit rate로 MPEG-2 압축을 수행한다. 프레임 공격을 위해 MPEG-2 복호를 수행하여 프레임 공격을 적용한 후에 다시 6Mbps의 bit rate로 MPEG-2 재압축을 수행한다. 표1은 실험을 위해 공모 공격을 제외한 각 공격에 설정된 윈도우의 크기를 나타내고 있다.

표 1. 공격에 대해 설정된 윈도우 크기

종류	제거	재배치	평균화	공모
크기	10	6	5	-

각 프레임 공격에 대한 비디오 재생의 영향을 살펴보면, 프레임 제거 공격은 전체 재생 시간이 짧아지게 되고, 실제 재생시 움직임이 많은 비디오에서는 부자연스러운 움직임이 발생된다. 또한 프레임 재배치 공격은 프레임 순서가 앞뒤로 바뀜에 따라 흔들림을 감지할 수 있다. 프레임 평균화 공격은 전체 프레임들이 흐려짐을 알 수 있다. 프레임 공모 공격은 공격된 프레임

은 평균적으로 약 30dB의 PSNR(peak signal-to-noise ratio)을 나타났으며, 워터마크가 예측에 의한 예측 에러가 존재함에 따라 MPEG-2 압축에 의한 블록킹 현상이 두드러지게 나타난다. 그림2(a)-(c)는 워터마크가 삽입된 프레임과 프레임 제거 및 재배치 공격을 제외한 평균화 공격, 공모 공격에 의한 프레임을 나타내고 있다.

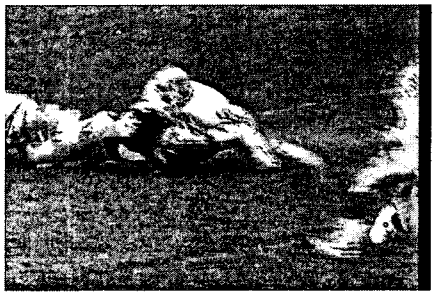

그림 2(a).워터마크가 삽입된 프레임

그림 2(b). 프레임 평균화 공격

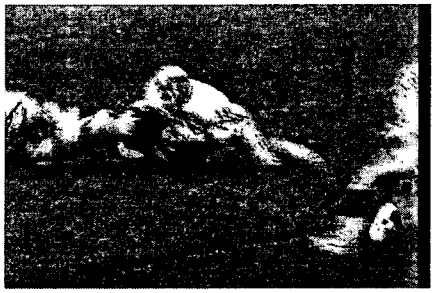

그림 2(c). 프레임 공모 공격

표2는 30개의 프레임에 삽입한 70비트의 워터마크 비트에 대해 프레임 공격 이후 검출된 워터마크 비트의 에러 비트 수를 나타내고 있다.

표 2. 워터마크 비트 에러 수

종류	제거	재배치	평균화	공모
크기	2	2	4	1

프레임 공격의 계산량은 제거 공격 < 재배치 공격 < 평균화 공격 < 공모 공격의 순으로 증가한다. 그러나 이러한 공격의 효과는 계산량의 순서와 꼭 일치하지 않고, 워터마크 시스템의 파라메터들, 예로써 워터마크의 구조 등에 의존한다. 이것은 표2의 결과에 의해서도 설명이 가능하다. 가령 프레임 제거 및 재배치 공격은 공모 공격보다 적은 계산량으로도 에러 비트의 수가 더 발생이 되는데, 이것은 제거 및 재배치 공격에 의해 발생되는 워터마크의 손실이 더 크기 때문이다. 또한 공모 공격은 참조 프레임의 예측 기법에 공격의 효과가 매우 의존된다. 따라서, 더 효과적인 공모 공격을 위해서는 보다 효율적인 예측 기법이 필요하다.

5. 결론

본 논문에서는 비디오 워터마킹에 대한 공격으로써 프레임 제거 및 재배치 공격, 평균화 공격 그리고 공모 공격과 같은 프레임 공격에 대해서 기술하고 각 구현 방법들을 제안하였다. 프레임 공격의 효과는 비디오 프레임들의 특징에 의존된다. 프레임 공격을 계획하고자 하는 경우에는 비디오의 특징을 적절하게 이용하는 방법을, 워터마킹 시스템을 설계하고자 하는 경우에는 그러한 프레임 공격을 고려하여야 한다. 본 논문에서 프레임 공격을 위한 척도로써 단순한 MSE를 이용하였으나, 비디오 워터마킹 시스템의 환경에 따라 다양한 방법이 가능하다. 향후 연구과제로 강인한 워터마킹을 설계하기 위해 다양한 환경에서 발생될 수 있는 효과적인 공격 방법에 대한 연구와 이러한 공격 방법에 대하여 강인한 워터마킹 시스템의 설계가 될 것이다.

[감사의 글]
본 논문은 정보통신부 지원 "디지털 컨텐츠 관리 기술 개발" 과제의 수행 결과의 일부로써 관계자분들에게 감사의 글을 전합니다.

[참고문헌]

[1] T. Y. C. et al., "Digital Watermarking for copyright protection of mpeg2 compressed video," IEEE Transaction on Consumer Electronics, 44(3), pp.895-901, 1998

[2] G. Langelaar and J. Biemond, "Real-time labelling of mpeg-2 compressed video," Journal of Visual Communication and Image Representation, 9(4), pp.256-270, 1998

[3] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," Signal Processing 66, pp.298-310, 1998

[4] B. Swanson and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," IEEE Journal on Selected Areas in Communications, 6(4), pp.256-270, 1998

[5] Z. Zhu and Y. Zhang, "Multiresolution watermarking for images and video," IEEE Transaction on Circuits and Systems for Video Technology, 9(4), pp.545-550

[6] E. T. Lin and E. J. Delp, "Temporal synchronization in video watermarking," Security and Watermarking of Multimedia Contents IV, SPIE Proceeding, 4675, 2002

[7] G. C. F. Dequilaume and T. Pun, Counter measure for unintentional and intentional video watermarking attacks," Security and Watermarking of Multimedia Contents II, SPIE Proceeding, 3971, 2000

[8] 이혜주, 홍진우, "인접 프레임의 특징을 이용한 워터마킹 삽입 영역의 구성 및 비디오 워터마킹", 한국멀티미디어 추계학술발표대회, 2002