RGB 변환에 의한 워터마크 공격 모델

정회원 권 용 광*, 이 시 몽*

Watermark Attack Model by RGB Conversion

Yong-kwang Kwon*, Si-mong Lee* ***Regular Members***

요 약

동영상 캡쳐 툴의 발전과 함께 워터마크가 삽입된 동영상에 대한 캡쳐링 공격이 가능하다. 동영상에 대한 캡쳐링 공격은 다양한 영상처리의 복합공격이 될 수 있으며, 따라서 이에 대한 분석이 용의하지 않다. 다만 RGB 색 공간으로 변환되어 캡쳐링 되는 상황에서 색 공간의 변환은 캡쳐링 공격의 필수 요소이다. 본 논문에서는 색 모델 변환 중 RGB와 YCbCr 변환이 삽입된 워터마크에 어떤 영향을 주는 가에 대해 실험적 및 이론적인 근거를 제시한다. 실험 결과에 의하면 기존 대역 확산 기반의 워터마킹 방법도 색 변환 공격에 취약하며 색 모델 변환식으로부터 그 해결책을 제시한다.

Key Words : Color Model, RGB, YCbCr, Watermark Attack

ABSTRACT

With the development of video capturing tools, it is possible to attack the watermarks embedded in the video data by the capturing. Since the capturing attack normally includes various image processing attacks, it is not so easy to deal with. However, we found that color space conversion based on the RGB color space is the major element of the capturing attack. In this paper, we present experimental and theoretical results on the effect of the color conversion between RGB and YCbCr to embedded watermarks. According to our experimental results, the previous watermarking method based on the spread spectrum scheme is vulnerable to the color conversion attack. A solution to this problem is devised from the formula of the color conversion.

Ⅰ. 서 론

기존의 워터마크는 강인함(Robustness)을 검증하기 위하여 벤치마킹 툴(Benchmarking tool)인 StirMark를 사용하였다. StirMark는 간단한 영상처리들과 함께 크로핑(Cropping), 전단(Shear), 회전(Rotation), 스케일 변환(Scale change), 라인 제거(Line removal) 등의 공격에 대한 워터마크의 내성을 측정한다[1]. 그 중에서도 영상을 회전하거나 늘리고 이동시키는 기하학적 변형인 RST(Rotation, Scaling, Translation) 공격[2]을 비롯하여, 워터마크를 검출하기 전에 먼저 워터마크가 삽입된 신호를 워터마크가 검출되지 않을 정도로 작게 조각낸 후 디텍터(Detector)를 지난 후에 다시 조각을 맞추는 Mosaic 공격[3], 그리고 타인의 워터마크가 삽입된 이미지에 자신의 알고리즘으로 워터마크를 다시 삽입하여 각자의 디텍터에서 모두 워터마크가 검출되어 소유권 분쟁을 일으킬 수 있는 SWICO(Single watermarked image counterfeit original) 공격[4] 등에 대한 연구가 많이 진행되었다.

이와 같이 이미 알려진 공격 외에도 영상처리 기술이 발달함에 따라 새로운 공격 모델이 생성되고 있으며, YCbCr과 RGB 간의 색모델 변환이 대표적

* 동국대학교 전자공학과 디지털영상처리연구실(fifthave@dongguk.ac.kr)
논문번호 : 08062-0929, 접수일자 : 2008년 9월 29일

인 예이다. 비디오를 이용한 영상처리는 YCbCr 색 모델을 기반으로 사용한다. 그러나 공간 영역과 주파수 영역에 관계없이 실제로 모니터를 통해 본 비디오 영상은 RGB 색 성분으로 변환된 영상을 보는 것이다. 즉, YCbCr은 절대 색공간이 아니며 RGB 정보를 인코딩하는 방식의 하나로, 실제로 보여지는 이미지의 색은 신호를 디스플레이 하기 위해 사용된 원본 RGB 정보에 의존한다[5]. 동영상 데이터를 이용하여 영상처리를 하는 것과 모니터를 이용하여 눈으로 보는 것은 YCbCr과 RGB의 차이이고 두 가지 영상 포맷의 변환은 두 색 모델의 사용하는 환경이 겹치지 않기 때문에 워터마크에 대한 공격으로 고려되지 않았다. 하지만 예전에는 없었던 모니터에 보여지는 화면을 그대로 캡쳐 할 수 있는 이전에 존재하지 않았던 도구들[6]이 발전되었고, 이로 인해 유료로 제공되는 스트리밍 방식의 영화나 UCC(User created contents)같은 개인의 창작물이 임의대로 저장이 가능해졌다. 즉, YCbCr 색 모델 기반의 동영상이 스트리밍 방식으로 전송되어 개인의 컴퓨터에서 RGB 색 모델로 표현되고, 캡쳐 툴에 의해 YCbCr로 다시 변환하여 저장되는 시나리오가 생겼다.

이와 같은 새로운 공격 모델에 대한 특징을 파악하고 기존 워터마크가 이러한 진화된 공격 모델에 강인한가에 대한 고찰이 필요하다.

Ⅱ. 공격 모델로서의 RGB 변환

2.1 RBG 변환수식

색 모델 간에는 사용되는 목적에 따라 표현의 차이가 있고 변환 수식 사이에 오차가 발생한다[7]. 이것은 저작권 보호를 위해 데이터를 삽입한 워터마크가 색 모델 변환에 의해 영향을 받을 수 있음을 시사한다.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.344 & -0.714 \\ 1 & 1.772 & 0 \end{bmatrix} \begin{bmatrix} Y \\ Cb-128 \\ Cr-128 \end{bmatrix} \quad (2)$$

식 (1)과 (2)는 각각 YCbCr과 RGB 색 모델 간의 변환식이다[8]. 두 식의 3x3 변환 행렬을 A와 B로 가정했을 때 $A \cdot B^T \neq I$ 이라는 것을 알 수 있다. 즉 변환이 정규 직교(Othogonal normal)하지 않는다. 이것은 변환 행렬에 의해 값의 범위가 바뀔 수 있다는 것을 의미한다. 또한 식 (2)의 경우 0과 255 사이로 클리핑 되기 때문에 오차가 더욱 커질 수 있다. 식 (1)과 (2)에 의한 오차는 식 (2)의 RGB 식을 (1)의 Y 식에 대입 후 정리하면 색 모델 변환 이전과 이후의 관계식을 다음과 같이 얻을 수 있다.

$$Y' \approx 0.8588Y + 16 \quad (3)$$

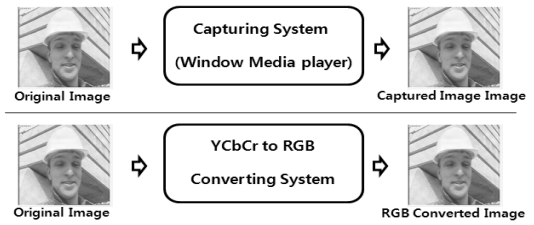
식 (3)은 대입식에서 일부 미약한 항을 무시하여 정리한 식이다. Y'는 변환 후의 밝기 값이고 Y는 변환 전의 밝기 값이다. 식 (3)을 보면 Y 밝기 값이 변환에 의해 0.8588의 일정한 비율로 감소한 후 16의 크기만큼 증가한 것을 알 수 있다.

변환식에 의한 밝기 성분의 선형적 변형은 주파수 영역의 계수에도 유사한 변형을 발생시킨다. 즉, 푸리에 변환은 선형변환이므로 공간 영역에서 비례적인 변화는 주파수 영역에서 같은 비율로 상대적인 변화를 보인다[9]. 즉, RGB 변환은 공간 영역과 주파수 영역에서 0.8588의 상대적 비율로 모든 밝기 값과 계수 크기를 스케일 변화 시킨 후 일정 크기만큼 증가하는 변환인 것을 알 수 있다. 따라서 선형 변환인 푸리에 변환 후의 변환 계수도 동일하게 영향을 받으며 주파수 계수에 삽입된 워터마크도 영향을 받을 수 있다.

2.2 RGB 변환과 캡쳐의 비교

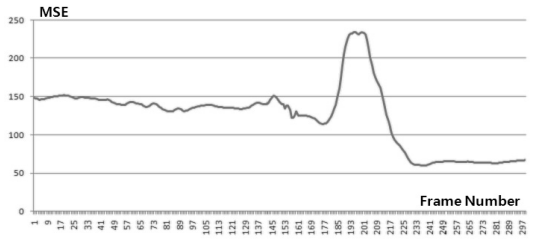
RGB 변환에 의한 공격 모델은 눈으로 보여지는 컴퓨터 화면의 캡쳐가 가능해지면서 새롭게 만들어진 워터마크에 대한 공격 시나리오다. YCbCr 기반의 동영상 파일이 모니터에서 동영상 플레이어를 통해 RGB로 표현되기 때문에 색 모델간의 변환이 이루어진 것이다. 캡쳐 공격과 RGB 변환이 같은 결과를 갖는지 확인함으로서 식 (3)에서 언급한 변환 오차에 대한 수식이 입증될 수 있다.

그림 1(a)는 'Foreman' 영상을 수식적으로 RGB 변환한 결과와 캡쳐로 취득한 영상의 비교 과정을 보여주며, 이때 캡쳐에 쓰이는 동영상 플레이어는 윈도우 미디어 플레이어(Windows Media Player, 이하 WMP)로 하였다. 그래프 (b)~(d)는 비교 과정을 거친 두 영상 사이에 MSE(Mean Square Error)를 구한 그래프이다. 가로와 세로축은 각각 'Foreman' 프레임 순서와 MSE 값이다. 총 300프레임의 평균 MSE는 원영상과 캡쳐 영상의 경우 124.5879, 원영상과 RGB 색 모델 변환 영상의 경우 128.6844, 그리고 캡쳐 영상과 RGB 색 모델 변환 영상의 평균 MSE 값은 0.1927이다. 이와 같이 원영상과 RGB 변환된 영상 및 원영상과 캡쳐 영상의 MSE값이 각

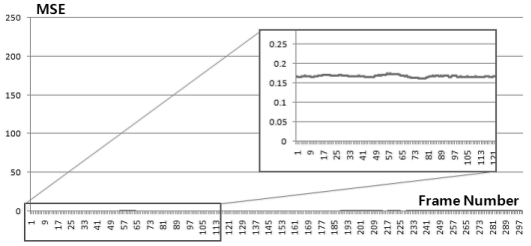


(a) Capturing and RGB conversion attacks to original image

(b) MSE between original and RGB converted image



(c) MSE between original and captured image



(d) MSE between captured and converted image

그림 1. RGB 변환과 플레이어 캡쳐 결과의 비교과정과 MSE

프레임에 대해 유사한 값을 보이는 반면 RGB 변환된 영상과의 MSE 값은 아주 작은 값을 보이는 것은 캡쳐에 의한 공격이 대부분 식 (3)의 색 변환에 기인한다는 사실을 입증한다.

2.3 기존 워터마크 알고리즘에 대한 강인성

Cox등[10]은 대역 확산 기반으로 개발된 워터마킹 방법의 강인성을 여러 영상처리 공격에 대해 입증하였다. 이미지 스케일링, 재 압축, 크로핑, 디더링, 공모 공격 등의 많은 강력한 공격들에 대해 강인한 면을 보였다. 이 논문은 이후 많은 알고리즘에 영향을 주었다.

$$v' = v(1+\alpha x) \tag{4}$$

$$x' = \frac{v'/v - 1}{\alpha} \tag{5}$$

$$\mathrm{sim}(x, x') = \frac{x \bullet x'}{\sqrt{x \bullet x'}} \tag{6}$$

식 (4)와 (5)는 Cox 알고리즘의 워터마크의 삽입 및 검출 수식이다. 식 (4)~(6)의 v는 DCT의 계수이고 v'는 워터마크가 삽입된 DCT의 계수를 나타낸다. x는 삽입할 워터마크 비트이고 x'는 검출한 워터마크 비트이다. 식 (5)를 보면 워터마크 비트 x'를 검출 할 때 원영상의 밝기 값인 v가 필요하다. α는 워터마크의 강도를 결정하는 스케일링 파라미터이며 클수록 워터마크는 강하게 삽입되지만 영상의 왜곡이 심해진다. α와 워터마크 비트의 길이는 워터마크의 강인성에 영향을 미친다. Cox의 결과[10]와 비교를 위해, 본 연구에서는 α를 0.1, 워터마크 비트 길이를 1,000으로 정하여 실험하였다.

그림 2는 'Foreman' 영상에 워터마크를 삽입하고 RGB 변환을 한 후에 계산한 상관계수(Correlation Coefficient)이다. 식 (6)의 상관관계식은 절대적인 비교 값을 나타내며, 식 (6)을 0에서 1로 정규화하여 상대적인 비교를 할 수 있는 상관 계수를 이용한 실험결과를 나타내었다[11].

RGB 변환 공격을 받은 후에는 상관계수가 낮은 값(0.2~0.3)을 보인다. 최대값 1과 비교해서 대략 20~30%의 검출율을 보인다는 것은 Cox 알고리즘이 RGB 색 모델 변환 공격에 의해 검출율이 현저히 떨어진 것을 시사한다. Cox 알고리즘은 DC 계수를 비롯하여 저주파 영역의 주요 계수를 조작하였고 강인성

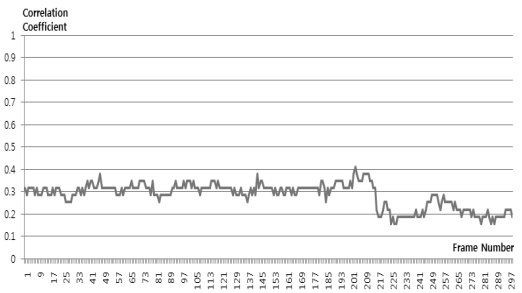


그림 2. 캡쳐 공격 후의 상관 계수

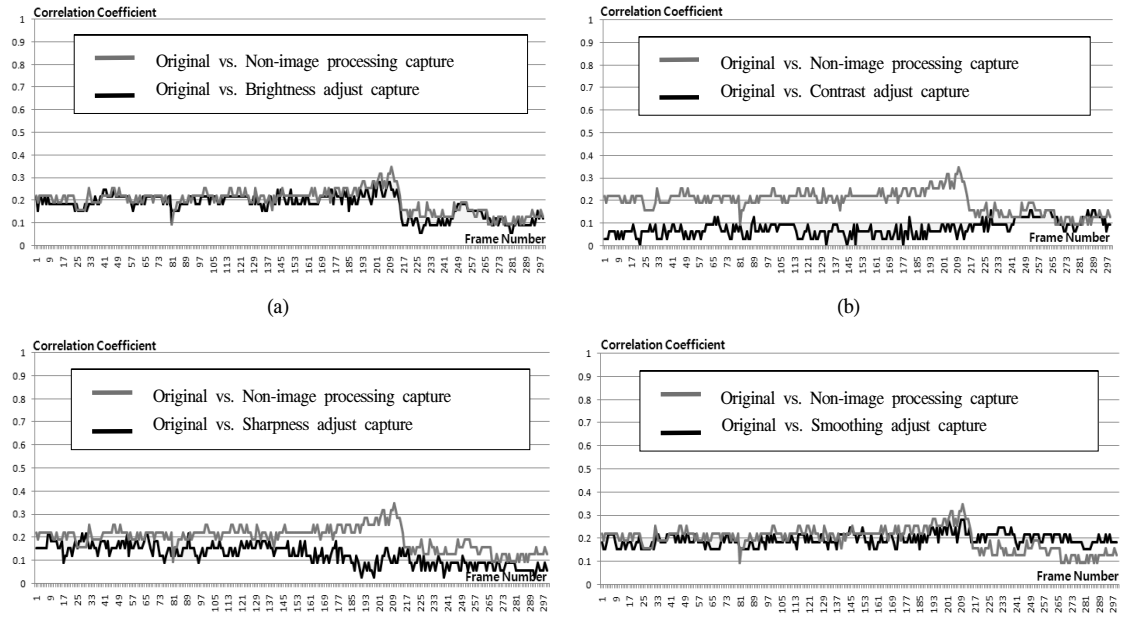


그림 3. 다양한 영상처리 후 캡쳐 공격에 대한 상관계수

을 테스트 했던 공격 모델들은 대체적으로 고주파 영역에 대해 영향을 주었기 때문에 워터마크 검출에 크게 영향을 미치지 못하였다. 하지만 RGB 색 모델 변환은 공간 영역 및 주파수 영역의 모든 계수를 비례적으로 변화시키므로 DC 계수나 저주파 영역의 계수들에도 큰 영향을 주어 대역 확산 기반 알고리즘의 상관계수를 20~30% 이하로 떨어뜨린다는 것을 을 알 수 있다.

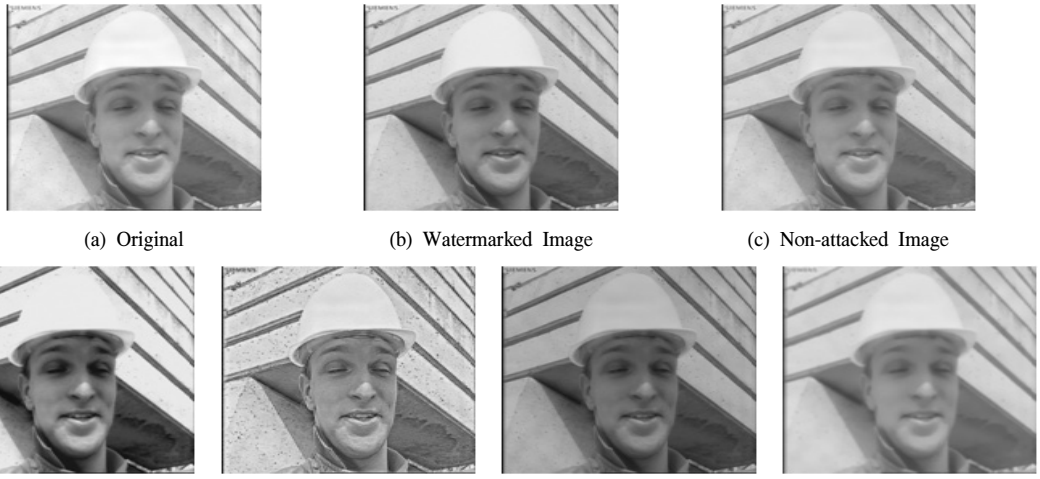
Cox의 방법은 검출할 때 원영상이 필요하다. 워터마크를 검출하고자 하는 영상의 Y성분의 밝기 값이 원영상보다 $(1+\alpha)$만큼 커야 워터마크 비트를 검출 할 수 있다. 위 실험의 경우 α를 0.1로 설정했으므로 원영상보다 1.1배의 밝기 값을 가져야 한다. RGB 변환은 0.8588의 비율로 밝기 값을 감소시키므로 워터마크가 삽입된 후 RGB 변환을 하면 1.1배 증가, 0.8588배 감소의 결과로 원영상의 0.94배의 밝기 값을 가지게 된다. 즉 원영상보다 밝기 값이 1.1배 커야하지만 RGB 변환 공격을 받은 후에는 0.94배의 크기밖에 되지 않기 때문에 그림 2의 결과와 같이 검출율은 현저히 감소한다.

2.4 캡쳐 공격의 다양성

동영상 플레이어는 화질 개선을 목적으로 여러 임의의 영상처리가 추가되며[12], 따라서 캡쳐는 다양한 영상처리를 포함한 복합적 공격 모델이 된다. 결국 캡쳐 결과물의 종류는 동영상 플레이어의 종류, 개인의 영상처리 설정 등에 따라 다양해진다.

그림 3은 동영상 플레이어의 각 영상처리를 거친 후의 캡쳐 결과물들이다. 본 논문에서는 임의의 영상처리를 하지 않은 캡쳐에 대해 실험을 하였고, 이는 RGB 변환 공격과 매우 유사한 결과를 나타내었다. 또한 대역확산 기반의 알고리즘이 임의의 영상처리를 하지 않은 캡쳐 또는 RGB 변환 공격으로 인해 워터마크의 검출률이 현저히 떨어진다는 것을 확인하였다. 그러나 다른 동영상 플레이어를 선택하거나 개별적 영상처리를 거친 캡쳐 결과물은 위의 실험과 결과가 다를 수 있다.

그림 4는 그림 3에서 동영상 플레이어에 개별적 영상처리를 추가한 후 캡쳐 영상에 대해 상관계수를 나타낸 그래프이다. 동영상 플레이어에 대한 캡쳐 공격은 기본 설정에서 이미 20~30%로 검출률을 감소시켰다. 이에 더해 그림 4의 실험결과는 영상처리를 추가한 후, 캡쳐 영상에 대해 워터마크를 검출할 경우 기본설정에 비해 상관계수가 더욱 감소한다는 것을 보여준다. 즉, 캡쳐 공격은 개인의 영상처리 설정으로 인해 많은 경우의 수를 가지며 기본 설정에 비해 검출률은 더욱 낮아진다. 그림 4는 동영상 플레이어에 임의의 영상처리를 추가한 경우 워터마크 검출률이 낮아진다는 것을 보여준다.



(a) Original (b) Watermarked Image (c) Non-attacked Image

(d) Brightness adjusted image (e) Sharpness adjusted image (f) Contrast adjust image (g) Smoothing adjust image

그림 4. 임의의 영상처리 후, 캡쳐 영상들의 비교

2.5 RGB 변환 공격에 대한 해결 대처 방법

서론에서 말했듯이 RGB 변환 공격은 플레이어로 재생중인 영상의 캡쳐로 발생되는 시나리오이다. 2.2절에서 캡쳐는 여러 영상처리가 합쳐진 복합적인 공격이며 사용자가 임의로 영상처리를 추가 할 수도 있다. 하지만 그림 1의 (d)에서 볼 수 있듯이 임의의 영상처리를 하지 않고 WMP로 재생한 동영상 캡쳐 결과는 RGB 변환 공격을 거친 영상의 결과와 매우 유사하였다. 즉, 0.8588의 비율로 영상의 Y성분 밝기 값의 크기를 줄이는 것이며 캡쳐와 RGB 변환 과정은 모두 동일하다고 볼 수 있다.

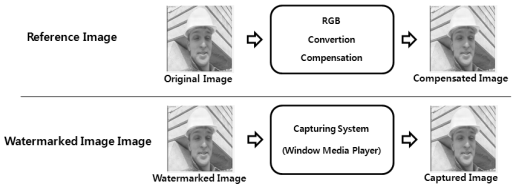
대역 확산 기반의 알고리즘의 경우 워터마크 삽입 후 영상의 밝기 값이 원영상을 기준으로 1.1의 비율로 증가한다. 캡쳐나 RGB 변환 공격은 영상의 밝기 값을 0.8588의 비율로 감소시킨다. 워터마크가 삽입된 영상에 대한 캡쳐 또는 RGB 변환 공격은 1.1의 비율로 증가한 밝기 값을 0.8588의 비율로 다시 감소시켜 결국 원영상을 기준으로 0.94배의 밝기 값 비율의 결과 값을 나타내어 워터마크 검출을 불가능하게 한다.

캡쳐 또는 RGB 변환 공격은 0.8588의 비율로 전체 픽셀의 밝기 값을 감소시키므로, 이를 해결하기 위해 워터마크 검출 시의 기준 원영상을 RGB 변환 보상(Compensation)하는 것이다. RGB 변환 보상은 기준이 되는 원 영상을 0.8588의 비율로 미리 감소시킴으로서 워터마크 삽입 후 캡쳐 공격을 받은 영상과 워터마크 밝기 값의 비율을 1.1배 보정시켜 주는 것이다.

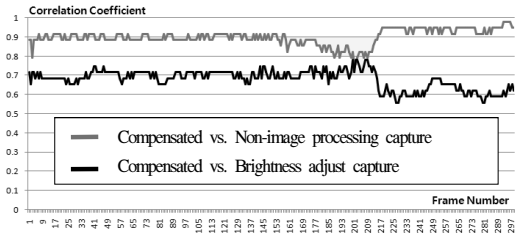
그림 5는 캡쳐 공격을 가한 후 워터마크 검출기에서 원영상과 RGB 변환 보상을 거친 원 영상을 이용하여 상관 계수를 비교한 그래프이다. 앞서 말한 것과 같이 원 영상을 RGB 변환 보상을 함으로서 워터마크를 검출할 영상과 Y성분의 밝기 값의 비율을 캡쳐 공격 이전의 비율로 조정해 주었다. 따라서 RGB 변환 보상을 하지 않은 원 영상을 이용하여 검출할 때 보다 더욱 향상된 검출 성능을 보였다.

Ⅲ. 결 론

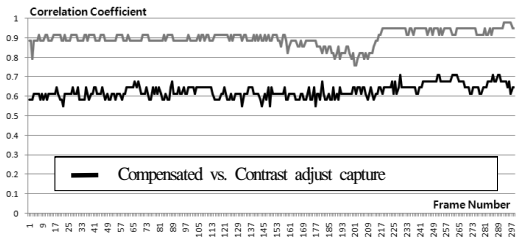
색 모델은 각각의 목적에 따라 만들어 졌고 다른 색 모델로 변환하는데 손실이 생긴다. 이전에는 색 모델을 변경할 시나리오가 존재하지 않았기 때문에 워터마크 알고리즘의 강인성 테스트에 있어서 색 모델 변환은 고려대상이 아니였다. 하지만 모니터 화면을 그대로 저장할 수 있는 캡쳐 툴이 발전하였고 누구나 손쉽게 자신이 보고 있는 컴퓨터 화면을 저장할 수 있게 되었다. 저작권 보호에 대한 관심이 높아지는 시점에서 개인이 만든 UCC나 유료로 스트리밍 서비스 되는 영화에 삽입 된 워터마크는 캡쳐에 무방비가 된다. 또한 본 논문에서 캡쳐 툴을 이용하여 RGB와 YCbCR의 변환이 워터마크에 영향을 줄 수 있다는 것을 확인하였다. 캡쳐 툴을 이용하여 RGB 색 모델을 변환하게 되면 공간 영역과 주파수 영역 전체가 0.8588 이라는 비율로 줄어들고 일정 크기만큼 증가한다. 어느 영역, 어느 위치에 워터마크를 삽입해도 RGB 변환에 의해 영



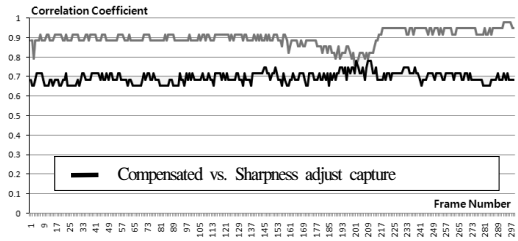
(a) Block Diagram



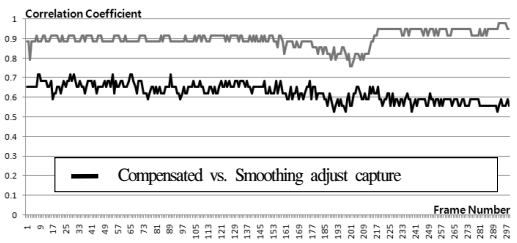
(b)



(c)



(d)



(e)

그림 5. RGB 변환 보상 방법과 보상 후 상관계수 비교

향을 받는다. 자신이 목적으로 선택한 공격 모델이 영향을 주지 않는 곳에 워터마크를 삽입하는 알고리즘을 만들어 왔지만 RGB 변환까지 고려한다면 계수들의 상대적 변화까지 견딜 수 있도록 알고리즘을 개선해야 할 것이다.

참 고 문 헌

〔1〕 M. Kutter, F. Petitcolas, "A fair benchmark for image watermarking systems," SPIE, Electronic Imaging, 99 Security and Wawtermarking of Multimedia Contents, Vol.3657, San Jose, Jan, 1999.

〔2〕 M. Kutter, "Watermarking resisting to translation, rotation, and scaling," In Proc. of SPIE Int. Sym. on Voice, Video, and Data Comm., Nov., 1998.

〔3〕 F. Petitcolas, R. Anderson, M. G. Kuhn, "Information Hiding: A survey," Proc. of IEEE, vol.87, no. 7, pp.1062-1077, Jul., 1999.

〔4〕 S. Craver et al., "On the invertibility of invisible watermarking techniques," Int. Conf. on Image Processing, vol.1, pp.540~543, Oct., 1997.

〔5〕 조맹섭, 디지털 컬러 프로세싱. 국제, 2006

〔6〕 http://www.techsmith.com/

〔7〕 Z.N. Li, M.S. Drew, *Fundamentals of multimedia*, Prentice Hall, 2003

〔8〕 ITU-R BT. 601 CCITT

〔9〕 J. McClellan, *Signal Processing First*, Prentice Hall, 2003

〔10〕 I.J. Cox et al., "secure spread spectrum watermarking for multimedia," IEEE Trans. on Image Processing, vol.6, no.12, Dec., 1997

〔11〕 M. George, J.Y. Chouinard, N. Georganas, "Spread spectrum spatial and spectral watermarking for images and video," IEEE CWIT99, 1999.

〔12〕 http://www.dscaler.com/

권 용 광 (Yong-kwang Kwon) 종신회원

2002년 2월 동국대학교 전자공학과 공학석사

2008년 2월 동국대학교 전자공학과 공학박사

<관심분야> H.264/AVC, 영상처리, 디지털 통신

이 시 몽 (Si-mong Lee) 정회원

2007년 2월 동국대학교 전자공학과 공학사

~ 현재 동국대학교 전자공학과석사과정

<관심분야> H.264, Watermark- ing, 영상처리