

복제 비디오 검출에서 비디오 지문의 강인함과 분별력 분석

김세민[†], 노용만^{**}

요 약

무분별한 복제 비디오를 막기 위하여 비디오 지문을 개발연구가 진행되고 있다. 이러한 비디오 지문들은 복제 비디오에서 발생하는 다양한 변화에 강인해야 하며 정확하게 구별될 수 있는 높은 분별력을 지녀야 한다. 일반적으로 비디오 지문들은 luminance(밝기), gradient(기울기), 그리고 DCT(주파수) 공간 등에서 주로 추출이 되고 있다. 그러나 아직 각 공간과 비디오 지문 사이에 실질적인 성능이 차이에 대한 연구가 부족하다. 따라서 본 논문에서는 각 공간에 따른 복제 비디오 검출 성능을 비교하기 위하여 강인함과 분별력에 기반한 복제 비디오 검출 실험을 진행하고 분석 하였다. 본 논문에서 동일한 패턴으로 각 공간에서 비디오 지문을 추출하고 각각 강인함과 분별력을 비교 한 후 최종적으로 복제 비디오 검출 실험을 진행하였다. 본 실험에서 DCT 공간에서 추출된 비디오 지문이 다른 공간보다 좀더 우수한 성능을 보여 주었는데 이는 해당 공간이 다른 비디오 지문들과 분별력이 가장 높았기 때문이다.

Analysis of the Robustness and Discrimination for Video Fingerprints in Video Copy Detection

Semin Kim[†], Yong Man Ro^{**}

ABSTRACT

In order to prevent illegal video copies, many video fingerprints have been developed. Video fingerprints should be robust from various video transformations and have high discriminative powers. In general, video fingerprints are generated from three feature spaces such as luminance, gradient, and DCT coefficients. However, there is a few study for the robustness and discrimination according to feature spaces. Thus, we analyzed the property of each feature space by video copy detection task with the robustness and the discrimination of video fingerprints. We generated three video fingerprints from these feature spaces using a same algorithm. In our test, a video fingerprint, based on DCT coefficient outperformed others because the discrimination of it was higher.

Key words: Video Fingerprint(비디오 복제 검출), Video Copy Detection(비디오 지문)

※ 교신저자(Corresponding Author) : 노용만, 주소 : 대전시 구성동 유성구 대학로 291 한국과학기술원 N1-김병호 IT융합센터 414호(305-701), 전화 : 042) 350-3494, FAX : 042) 350-7619, E-mail : ymro@ee.kaist.ac.kr
접수일 : 2013년 8월 28일, 수정일 : 2013년 9월 25일
완료일 : 2013년 9월 30일

[†] 준회원, 한국과학기술원 정보통신공학과
(E-mail : resemim@kaist.ac.kr)

^{**} 정회원, 한국과학기술원 전기및전자공학과
※ 본 연구는 미래부가 지원한 2013년 정보통신·방송(ICT) 연구개발사업의 연구결과로 수행되었음.

1. 서 론

IT 기기들의 급속한 발전과 멀티미디어 저작 도구의 배포로 인하여 누구나 쉽게 비디오 콘텐츠를 개발할 수 있는 시대가 도래되었다. 그러나 이와 함께 비디오 콘텐츠를 저작권자의 허락 없이 자신의 목적에 맞게 편집 및 재배포하여 수많은 복제 비디오들이 만들어 졌다. 이러한 복제 비디오는 저작권 침해와 네트워크 리소스 낭비 등의 결과를 초래하고 있기 때문에 이를 극복할 복제 비디오 검출을 위한 연구가 많이 진행되고 있다[1-11].

비디오는 데이터의 크기가 방대하기 때문에 직접적으로 비디오를 비교하여 복제된 비디오를 검출하기는 힘들다. 따라서 비디오를 작은 크기의 공간으로 표현할 수 있는 비디오 지문으로 변환하고 이를 비교하여 비디오의 복제 여부를 판별하게 된다. 그러나 복제된 비디오는 변형과정에 화질, 압축율, 그림 삽입, 색상, 밝기 등 다양한 변형이 발생되곤 한다. 그림 1은 원본 비디오 (a)와 5가지의 변형된 비디오들을 보여준다. 비록 변형된 비디오는 원본 비디오와 시각적 차이를 보여주고 있지만 여전히 복제 된 비디오라 인지된다. 따라서 복제 비디오 검출의 주된 연구 방향은 다양한 변형에서도 원본을 인지할 수 있는 강력하고 분별력 높은 비디오 지문 개발이라 할 수 있다.

비디오 지문은 비디오의 키프레임을 구성하는 색

상이나 질감 등의 특징들로 구성된다. 이러한 특징들을 추출하는 방법은 다양하나 일반적으로 luminance(밝기), gradient(기울기), 그리고 DCT(주파수) 공간 등에서 추출된다. Luminance 공간기반 지문은 비디오의 키프레임을 구성하는 화소에서 직접적으로 특징을 추출하며 대표적으로 밝기(luminance) 값이나 회색 (gray-scale)값 등을 이용하는 방법이다[1-5]. Gradient 공간기반 방법은 키프레임을 미분하고 에지(edge)를 추출한 다음 이를 이용하여 비디오 지문을 만든다[6,7,9]. 이런 gradient 기반 특징은 주로 키포인트 매칭을 통한 복제 비디오 검출[8]로 주로 이용되는데 대표적으로 SIFT [12]라는 특징이 주로 사용된다. 마지막으로 DCT 공간기반의 지문은 키프레임을 DCT 등의 공간으로 변형하고 AC 계수들을 이용하여 지문을 생성한다[10,15].

본 논문에서는 이 3가지 (luminance, gradient, DCT) 공간들에 대하여 비디오 지문으로서의 적합성을 비교하고자 한다. 비디오 지문은 변형이 되더라도 원본으로 인식 될 수 있어야 하는데 이를 강인함(robustness)이라 한다. 또 비디오 지문은 다른 비디오 지문들과 쉽게 구별 될 수 있어야 하는데 이를 분별력(discrimination)이라 한다. 따라서 본 논문에서는 강인함과 분별력에 기반하여 복제 비디오 검출 실험을 진행하였다. 그러나 기존의 비디오 지문 비교 연구들은 주로 다양한 알고리즘에 초점을 두었기 때



그림 1. 복제 비디오 예제. (a) 원본, (b) 흐림 효과, (c) 강한 재인코딩, (d) 그림 삽입, (e) 색상 변경, (f) 감마 변경

문에 특징 공간에 따른 성능을 알기 힘들었다. 따라서 본 논문에서는 동일하고 간소한 추출 알고리즘[8]을 참고하여 각 공간들이 복제 비디오 검출 성능에 미치는 영향에 초점을 두었다. 본 논문의 실험 결과에서는 DCT 공간의 비디오 지문들이 가장 높은 성능을 보여 주었다. 본 논문에서 실험에서 DCT 공간의 비디오 지문들은 강인함이 떨어지지만 높은 분별력을 지니고 있어 비디오의 다양한 변화에도 가장 높은 복제 비디오 검출 성능을 보여주었다.

본 논문의 구성은 다음과 같다. 2장에서는 비디오 지문을 생성하기 위한 방법을 서술하며 3장에서는 이에 기반한 복제 비디오 검출 실험을 진행한다. 마지막으로 4장에는 본 논문의 결론을 보여준다.

2. 비디오 지문 검출

본 장에서는 luminance, gradient, DCT 공간 등에서 비디오 지문을 추출 한다. 공간 특성에 대한 비디오 지문 적합성을 살펴보기 위하여 각 공간에 대하여 동일한 추출 방법을 하는데 Esmaeili [10] 방법을 참고하였다. Esmaeili 는 비디오 지문을 추출하기 위하여 영상의 여러 개의 블록으로 나누고 각 블록에서 특징을 추출하여 이를 해쉬(hash)화하여 이진화 패턴을 생성하였다.

먼저 비디오 영상의 크기는 다양하기 때문에 똑같은 크기로 변경을 해준다. 그림 2와 같이 키프레임 영상을 그레이(gray) 이미지로 변환하고 가우시안 필터를 씌워 작은 노이즈들을 제거한다. 본 논문에서

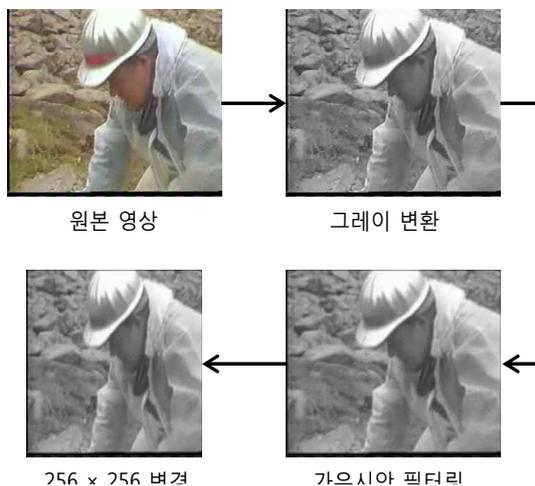


그림 2. 비디오 지문을 동일한 크기로 추출하기 위하여 영상 크기를 256×256으로 변경하는 과정.

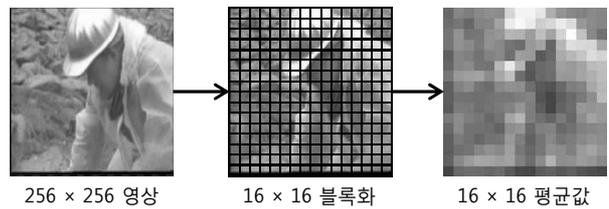


그림 3. Luminance 공간의 비디오 특징 추출

는 가우시안 필터 크기를 5×5 로 하고 시그마 값을 2로 사용하였다. 가우시안 필터를 씌우고 나서 영상을 256×256 크기로 변경한다.

변경된 키프레임으로부터 luminance, gradient, DCT 공간에 대하여 각각 256차원의 비디오 지문을 추출한다. 먼저 luminance 공간으로부터 지문 추출은 그림 3과 같이 이루어 지는데 영상을 16×16 개의 블록으로 나눈 후 각 블록의 평균 luminance 값으로 각 블록을 표현한다. 즉 256×256 영상은 각 블록의 대표값으로 16×16 크기의 행렬이 되고, 이 행렬을 1차 행렬로 표현하면 256 차원의 열벡터가 생성된다.

Gradient 공간에서의 비디오 지문 추출은 [13] 방법을 참고하여 그림 4와 같이 진행된다. 먼저 가로 방향과 세로 방향으로 각각 영상을 미분하고 두 미분 영상의 합(magnitude)을 구한다. 그리고 luminance 공간과 마찬가지로 256 차원의 열벡터를 생성한다.

마지막으로 DCT 공간에서의 비디오 지문 추출은 [10] 방법을 참고하여 그림 5와 같이 진행된다. 먼저 영상을 16×16 개의 블록으로 나누면 각 블록은 16×16 개의 픽셀을 가진다. 이때 각 블록에 대하여 2차원 DCT 변환을 한 후 그림 5처럼 DC 의 대각선으로 붙어있는 AC 를 선택한다. 왜냐하면 해당 AC가 가로와 세로의 DCT 값을 동시에 고려하고 있는 가장 낮은 저주파 영역이기 때문이다. 따라서 각 블록에서

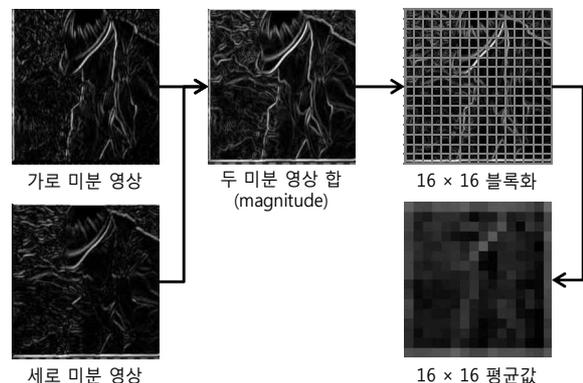


그림 4. Gradient 공간의 비디오 특징 추출

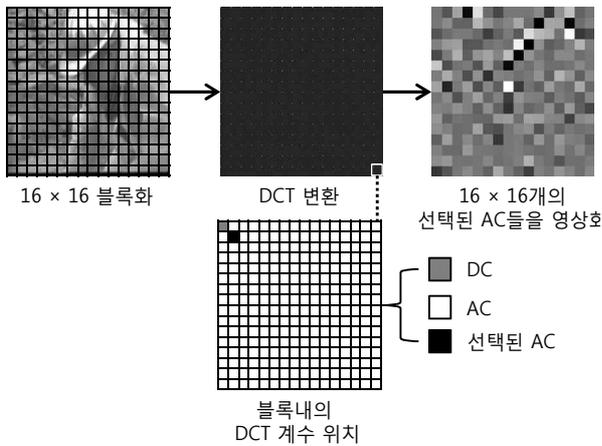


그림 5. DCT 공간의 비디오 특징 추출

해당 AC를 선택하면 16 × 16의 행렬이 완성되고 다시 256 차원의 열벡터로 변환한다.

이제 각각의 공간에서 추출된 열벡터들을 이진화 패턴으로 만드는 과정을 진행한다. 먼저 luminance, gradient, DCT 열벡터들을 $L = \{l_1, l_2, \dots, l_{256}\}$, $G = \{g_1, g_2, \dots, g_{256}\}$, $D = \{d_1, d_2, \dots, d_{256}\}$ 라 각각 정의 한다. 이때 각 열벡터 중 중간 값을 하나 선택하여 각각 l_m, g_m, d_m 이라 정의 한다. 마지막으로 각 공간들의 열벡터들은 각각 $b_L = f(L, l_m)$, $b_G = f(G, g_m)$, 그리고 $b_D = f(D, d_m)$ 으로 이진화가 되고 비디오 지문이 최종적으로 생성된다. 이때 $f(\cdot)$ 는 다음과 같이 정의 된다.

$$f(\mathbf{X}, x_m) = \{b_1, b_2, \dots, b_{256}\}, \quad (1)$$

여기서 $\mathbf{X} = \{x_1, x_2, \dots, x_{256}\}$, x_m 은 \mathbf{X} 내의 중간 원소이며 b_i 는 다음과 같이 계산된다.

$$b_i = \begin{cases} 1, & \text{if } x_i \geq x_m \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

3. 실험 결과

각 공간에서 추출된 지문들의 강인함과 분별력을 분석하기 위해서 복제 비디오 검출에서 주로 사용되는 MUSCLE-VCD 2007 비디오[14]를 사용하였다. 실험을 위하여 전체 비디오 중 40개의 비디오를 선택하고 이중 10개를 질의(query) 비디오로 선택하였다. 각 비디오는 샷(shot)으로 분할되고, 각 샷에서 하나의 키프레임이 추출된다. 이때 키프레임의 중복을 없애기 위하여 타이틀이나 블랙 스크린 등을 삭제한다. 본 실험에서 40개의 비디오로부터 40개의 비디오에

서 추출된 4,523 키프레임들을 참조 셋으로 이용하고 10개의 비디오 에서 추출된 1,494 키프레임들을 질의 셋으로 사용 하였다.

먼저 질의 셋에 대하여 그림 1과 같이 복제 비디오에서 발생하는 비디오의 변화를 표 1과 같이 적용하였다. 우리는 5가지 변형에 대하여 각 4개의 변수를 설정하고 각 질의 키프레임들을 변형 시킨다. 다음 변형된 키프레임들과 원본 키프레임들에 대하여 각각 luminance, gradient, DCT 공간으로부터 특징들을 추출하고 식(1)을 활용해 비디오 지문들을 생성한다.

본 실험에서 비디오 지문의 강인함과 분별력(다음과 같이 정의 된다. 먼저 강인함이란 변형된 키프레임으로부터 추출된 비디오 지문이 원본 지문과 어느 정도 일치하는지를 나타낸다. 예를 들어 \mathbf{I} 를 키프레임이라 하고 여기에서 추출된 비디오 지문을 \mathbf{i} 라고 하자. 그리고 \mathbf{I}' 를 변형된 키프레임이라 하고 추출된 비디오 지문을 \mathbf{i}' 라고 정의한다. 이때 강인함은 \mathbf{i} 와 \mathbf{i}' 사이의 해밍 거리(Hamming Distance)로 구해진다. 따라서 해밍 거리가 작을수록 지문의 강인함은 증가된다. 분별력은 서로 다른 키프레임들 간에 해밍 거리로 구해진다. 예를 들어 두 키프레임 \mathbf{I} 와 \mathbf{J} 에서 추출된 지문 \mathbf{i} 와 \mathbf{j} 간의 해밍 거리가 커질수록 분별력이 크다는 것을 의미한다.

따라서 비디오 지문이 강인하면 해밍 거리가 낮게 나타나야 하고, 비디오 지문의 분별력이 높으면 해밍 거리는 높게 나타나야 한다. 먼저 강인함을 측정하기 위하여 변형된 키프레임에서 추출된 비디오 지문들과 원본들을 해밍 거리로 확률분포를 그림 6과 같이 그려보았다. (a)에서 보이는 것처럼 luminance 공간

표 1. 실험에 사용된 비디오 변형 예시

비디오 변형	설 명
Smoothing	Smoothing 마스크 크기 변경 크기: 5×5, 10×10, 15×15, 20×20
Re-encoding	JPEG Quality 변경 Quality: 5, 10, 20, 40
Picture in Picture	키프레임에 다른 키프레임 삽입 비중: 10, 15, 20, 25
Hue-Correction	키프레임의 색상 변경 Hue Angle: -90, -45, 45, 90
Gamma-Correction	키프레임의 밝기 변경 Gamma: 0.2, 0.6, 1.5, 2.0

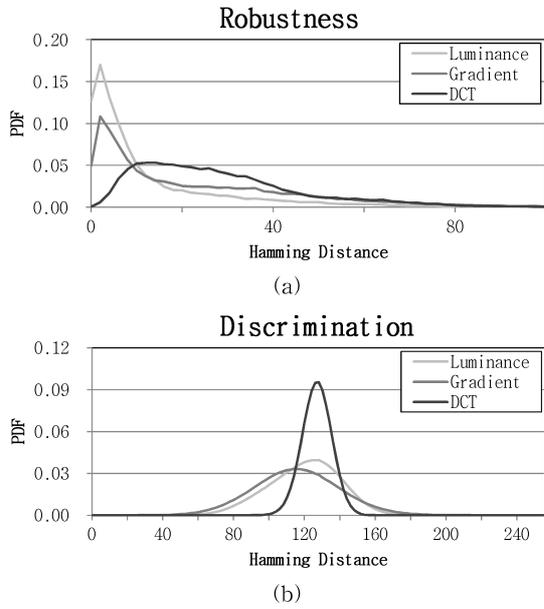


그림 6. 각 지문들에 대하여 강인함과 분별력을 확률밀도 함수로 비교. (a)강인함, (b) 분별력.

의 비디오 지문이 가장 낮은 해밍거리를 보이고 있는데 이는 비디오 변형에 다른 공간들보다 비디오 지문들의 변화가 작다는 것을 의미한다. 즉, luminance 공간의 비디오 지문이 가장 높은 강인함을 보여주었다. 반면에 분별력 분포는 (b)에서 보이는 것처럼 DCT 공간의 해밍 거리가 높게 나오는 것을 볼 수 있다. 즉, DCT 공간의 비디오 지문이 가장 높은 분별력을 보여주고 있다. 각 확률 밀도를 해밍 거리의 평균으로 구해보면 표 2와 같다.

마지막으로 실제적인 복제 비디오 검출에 대하여 성능을 분석하여 본다. 복제 비디오의 판별은 강인함과 분별력 그래프에 대하여 동시에 그래프로 그려보았을 때 중첩되는 부피에 대한 비교로 가능하다. 따라서 각 비디오 지문들에 대하여 강인함과 분별력에 대한 중첩부분을 검출 에러로 정의 하고 이를 그려보면 그림 7과 같다.

그림 7에서 DCT 공간기반의 비디오 지문이 가장 중첩 영역이 작고 Gradient 공간의 비디오 지문이 중첩 영역이 가장 컸다. 따라서 본 논문의 실험환경에서 진행된 복제 비디오 검출에 관하여 DCT 공간

표 2. 비디오 지문들에 대한 강인함과 분별력 평균

종 류	Luminance	Gradient	DCT
강인함	13.6	23.1	28.5
분별력	120.0	115.8	126.9

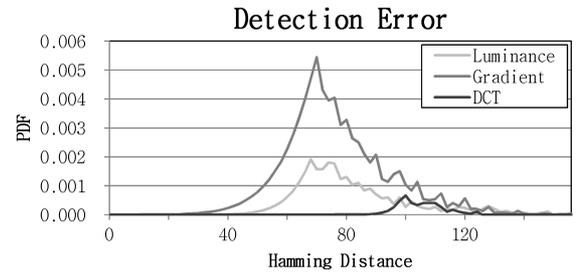


그림 7. 각 비디오 지문들에 대하여 중첩 영역에 대한 확률 분포. 각 그래프의 면적은 검출 에러의 크기를 나타낸다.

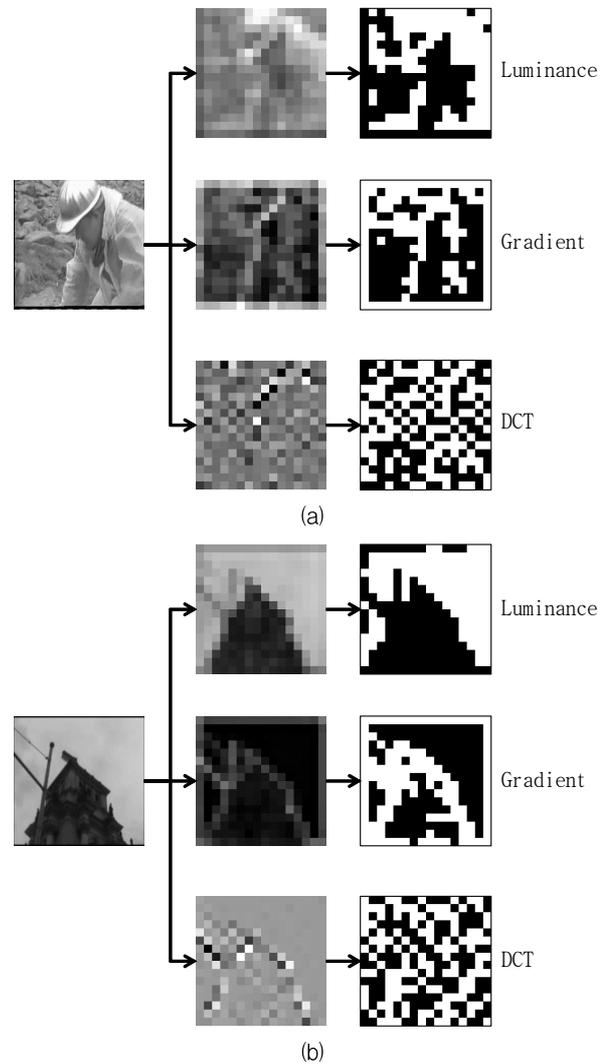


그림 8. 각 영상을 세 가지 공간으로부터 추출하여 이진화한 예제. Luminance와 gradient 기반의 이진화 영상들은 1(흰색)과 0(검은색)이 몰려있는 반면 DCT 기반의 영상은 다양하게 1과 0이 분포된다.

의 비디오 지문이 가장 우수하다고 결론지을 수 있었다.

그림 8은 위와 같이 DCT기반의 방법이 가장 높은 성능을 보여준 이유를 좀더 쉽게 설명해 주고 있다.

각 그림들은 원본 이미지들로부터 각각의 지문들을 추출하여 이를 이진화하여 이미지로 표현한 그림들인데 흰색은 1을 나타내고 검은색은 0을 나타낸다. Luminance 기반의 비디오 지문은 영상의 밝은 부분에 1이 몰려 있는 것을 알 수 있다. 그리고 gradient 기반의 비디오 지문은 영상의 에지영역에 1이 몰려 있는 것을 알 수 있다. 이에 반하여 DCT 기반의 비디오 지문은 불규칙하게 0과 1이 분포하고 있는 것을 볼 수 있다. 이는 비디오 지문의 분별력을 높일 수 있게 해주기 때문에 다른 지문들 보다 좀더 분별력이 높은 비디오 지문을 생성할 수 있었다.

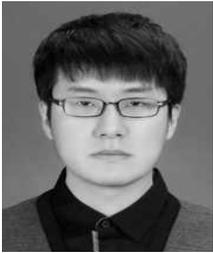
4. 결 론

본 논문에서는 복제 비디오 검출에 대하여 강인하고 분별력이 높은 비디오 지문을 개발하기 위한 특징 공간에 대하여 분석 하였다. 본 논문에서 luminance, gradient, DCT 공간에 대하여 비디오 지문을 추출하고 복제 비디오 검출에 대한 실험을 진행하였다. 진행된 실험에서 DCT 공간의 비디오 지문이 가장 낮은 강인함을 보여주었지만 비디오 지문들간에 가장 높은 분별력을 보여주어 실질적인 복제 비디오 검출에 가장 우수한 성능을 보여주었다. 따라서 본 논문의 결론처럼 복제 비디오 검출을 위한 특징 추출에서 DCT 공간을 활용하면 타 공간보다 높은 성능을 예상할 수 있다.

참 고 문 헌

- [1] C. Kim and B. Vasudev, "Spatio-temporal Sequence Matching for Efficient Video Copy Detection," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 15, No. 1, pp. 127-132, 2005.
- [2] L. Chen and F.W.M. Stentiford, "Video Sequence Matching Based on Temporal Ordinal Measurement," *Pattern Recognition Letters*, Vol. 29, No. 13, pp. 1824-1831, 2008.
- [3] A. Hampapur and R. Bolle. "Comparison of Sequence Matching Techniques for Video Copy Detection," *Conf. Storage and Retrieval for Media Databases*, pp. 194-201, 2002.
- [4] 정관민, 김정엽, 현기호, 하영호, "비디오 정합을 위한 오디널 특징의 유일성 및 강건성 분석," 한국멀티미디어학회논문지, 제9권, 제5호, pp. 576-584, 2006년.
- [5] 현기호, "저작권보호를 위한 내용기반 비디오 복사검출의 비디오 정합 알고리즘," 한국멀티미디어학회논문지, 제11권, 제3호, pp. 315-322, 2008년.
- [6] G. Leon, H. Kalva, and B. Furht, "Video Identification Using Video Tomography," *IEEE Int. Conf. Multimedia and Expo*, pp. 1030-1033, 2009.
- [7] O. Kucktunca, M. Basttanb, U. Gudukbayb, and O. Ulusoyb, "Video Copy Detection Using Multiple Visual Cues and MPEG-7 Descriptors," *Journal of Visual Communication and Image Representation*, Vol. 21, No. 8, pp. 838- 849, 2010.
- [8] S. Wei, Y. Zhao, C. Zhu, and C. Xu, "Frame Fusion for Video Copy Detection," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 21, No. 1, pp. 15-28, 2011.
- [9] J. Law-To, Gout-Branet, O. Buisson, and N. Boujemaa, "Local Behaviours Labelling for Content Based Video Copy Detection," *Int. Conf. Pattern Recognition*, pp. 232-235, 2006.
- [10] M.M. Esmaeili, M. Fatourechi, and K.R. Ward, "A Robust and Fast Video Copy Detection System Using Content-Based Fingerprinting," *IEEE Trans. Information Forensics and Security*, Vol. 6 No. 1, pp. 213-226, 2011.
- [11] B. Coskun, B. Sankur, and N. Memon, "Spatio-Temporal Transform Based Video Hashing," *IEEE Trans. Multimedia*, Vol. 8, No. 6, pp. 1190-1208, 2006.
- [12] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.
- [13] N. Dalal and B. Triggs, "Histograms of Oriented Gradients For Human Detection," *IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 886-893, 2005.

- [14] Muscle-VCD-2007: A Live Benchmark for Video Copy Detection, <https://www.rocq.inria.fr/imedia/civr-bench/>, 2007.
- [15] C. Kim, "Content-based Image Copy Detection," *Signal Processing: Image Communication*, Vol. 18, No. 3, pp. 169-184, 2003.



김 세 민

2006년 2월 안동대학교 멀티미디어공학과 공학사
 2008년 2월 경북대학교 컴퓨터공학과 공학사
 2008년 1월~6월 한국산업기술재단 연구원

2008년 8월~현재 한국과학기술원 정보통신공학과 박사과정
 관심분야: 영상 패턴 추출, 이미지/비디오 분류 및 검색



노 용 만

1985년 2월 연세대학교 전자공학과 공학사
 1987년 2월 한국과학기술원 전기 및 전자공학과 공학석사
 1992년 8월 한국과학기술원 전기 및 전자공학과 공학박사

1987년~1988년 Columbia University 연구원
 1996년~1997년 University of California Berkeley 연구원
 2006년~2007년 University of Toronto 방문교수
 1997년 2월~현재 한국과학기술원 전기 및 전자공학과 정교수
 관심분야: 영상 및 비디오 신호처리, 패턴인식, MPEG-7, Image and Video indexing, Spectral analysis of image signal