

모션의 방향성 히스토그램을 이용한 내용 기반 비디오 복사 검출

(Content-Based Video Copy Detection using Motion Directional Histogram)

현 기 호 [†] 이 재 철 ^{**}
(Ki-Ho Hyun) (Jae-Cheol Lee)

요 약 내용기반 비디오 복사 검출(content-based video copy detection)은 기존의 워터마킹 방법과 반대의 개념으로서, 비디오 스트림에 독특한 패턴을 첨가하는 워터마킹에 비해, 비디오의 복사본을 검출하기 위해 패턴을 첨가하지 않고 원본 비디오의 내용 기반 특징(content-based signature)을 비교한다. 기존의 일반적인 내용 기반 복사 검출방법은 키 이미지를 선택 한 다음 이미지 정합을 사용하였으나, 본 논문은 비디오 복사검출을 위해 시간에 따른 영상의 변화를 나타내는 모션을 구한다. 이를 각 방향으로 양자화하여 제한한 방향성 히스토그램을 구하면 비디오의 클립은 1차원 그래프 형태로 변환된다. 제한한 알고리즘은 실시간 검색을 위한 인덱스 구성에 적합하고, 비디오 특징의 정합에 의해 움직임 변화가 많은 TV광고 노출 횟수 검색 등에 유리하다.

키워드 : 내용기반 비디오 복사 검출, 순서정합, 모션 방향성 히스토그램

Abstract Content-based video copy detection is a complementary approach to watermarking. As opposed to watermarking, which relies on inserting a distinct pattern into the video stream, video copy detection techniques match content-based signatures to detect copies of video. Existing typical content-based copy detection schemes have relied on image matching which is based on key frame detection. This paper proposes a motion directional histogram, which is quantized and accumulated the direction of motion, for video copy detection. The video clip is represented by a motion directional histogram as a 1-dimensional graph. This method is suitable for real time indexing and counting the TV CF verification that is high motion video clips.

Keywords : Video copy detection, Sequence Matching, Motion directional histogram

1. 서 론

인터넷을 통한 미디어의 교환이 활발히 이루어지고 있는 요즘에 미디어의 불법복제에 대한 저작권보호의 필요성이 증가하고 있다. 최근의 보호방법으로 독특한 패턴을 첨가하는 워터마킹등이 대안으로 연구되고 있으나 미디어 파일의 변환이나 이미 배포된 미디어에 대하여는 보호할 수 없는 단점이 있다.

내용기반 복사검출(content-based copy detection)의

기본적인 주제는 “미디어자체가 워터마크(the media itself is the watermark)”이다. 즉 미디어(비디오, 오디오, 이미지)는 복사본을 검출하기 위한 충분하고 독특한 정보를 보유하고 있다는 것이다. 내용기반 복사검출 방법은 먼저 원래의 미디어에서 특징(signature)을 추출한다. 같은 특징이 테스트 미디어 스트림으로부터 추출되고 테스트 스트림이 원본 미디어의 복사본을 포함하는지를 판단하기 위해 원본 미디어의 특징과 비교한다.

워터마킹에 비해 내용기반 복사검출의 중요한 장점은 “특징 추출을 미디어가 보급된 이후에도 할 수 있다”는 것이고 워터마킹을 제거할 목적의 미디어 파일 변환에도 적용할 수 있다. 예를 들면 내용기반 복사검출 방법으로 Star Wars무비의 마스터 테이프에서 특징들을 추출하고, 이 특징들은 인터넷상에서 불법 복제된 스

[†] 정 회 원 : 영산대학교 컴퓨터정보공학부 교수
khhyun@mail.yosu.ac.kr

^{**} 비 회 원 : 한국원자력연구소
jlee2@kaeri.re.kr

논문접수 : 2002년 3월 21일
심사완료 : 2003년 3월 7일

타워스 무비의 클립들을 찾는 데 사용될 수 있다. 또한 이 방법은 워터마크방법의 보완적인 백업처럼 활용할 수도 있다. 다른 응용분야로 실시간 검색을 구성하면 방송에서 노출되는 TV 광고의 횟수를 검출할 수도 있다. 최근에 여러 연구노력과 상업적인 회사가 내용기반 복사검출방법에 대한 연구를 진행하고 있다[1-8]. 기존의 내용기반방법은 급격한 장면변환이나 점진적인 장면변환인 프레임간의 키(key) 이미지들로부터 이미지 기반 특징을 추출하여 정합한다.

본 논문은 키 프레임으로 정합하지 않고 연속적인 비디오 프레임에서 시간에 따른 변화특성인 모션을 사용한 방향성 히스토그램 특징을 제안하였다. 제안한 방향성 히스토그램을 구하면 비디오의 클립은 1차원 그래프 형태로 변환된다. 원본 비디오와 테스트 비디오의 정합은 마치 두개의 1차원 그래프를 상충 적분하는 것과 같은 순서정합(sequence matching)을 사용하였다. 이는 실시간 검색을 위한 인덱스 구성에 적합하고, 비디오 특징의 정합에 의해 움직임 변화가 많은 비디오 즉 TV 광고 노출 횟수 검색 등에 유리하다. 실험은 약 30초 길이의 일반 TV광고를 연결한 디지털화한 파일에서 특정 광고를 찾아내는 방법으로 하였다. 모션의 방향성 히스토그램을 기준 특징(reference signature)으로 하여 테스트 미디어인 광고로부터 추출되고, 연속적인 TV 광고에서 테스트미디어의 존재와 위치를 검출하기 위해 기준 특징과 순서 정합한다.

2. 내용기반 비디오 복사검출

내용기반 비디오 복사검출은 원본 비디오로부터 시간 혹은 공간영역에서의 특징을 추출하고, 이들 특징을 실시간 비디오 스트림으로부터 추출된 특징과 정합하여 복사여부를 검출한다. 복사검출 단계는 크게 특징추출단계, 데이터베이스 인덱싱 단계, 정합단계로 나눌 수 있는데 전체적인 시스템의 구성도는 그림 1과 같다. 실제 시스템을 구현하기 위해서는 각 단계가 이상적으로 2초 미만의 짧은 클립을 정합해야 하고, 25시간 이상의 데이터베이스 인덱스를 구성할 수 있어야 하며 실시간 처리가 되어야 한다. 비디오의 특징추출은 칼라, 명암도 등의 공간적인 특징과 모션, optical flow 등의 시간적인 특징으로 나눌 수 있다. 현재까지는 키 프레임에 정한 후 이미지 특징을 이용한 정합이 내용기반 복사검출 분야에 적용되어 왔다.[1,2,3,4,5]

Linehart 등은 클립의 키프레임 들을 표현하기 위해 칼라 결함(color coherence)벡터를 이용하여 비디오 클립 정합을 위한 시스템에 적용했다.[3] Sanchez 등은 복

사검출을 위해 키프레임의 칼라 히스토그램에서 주성분(principal component)을 특징으로 사용했다.[6]

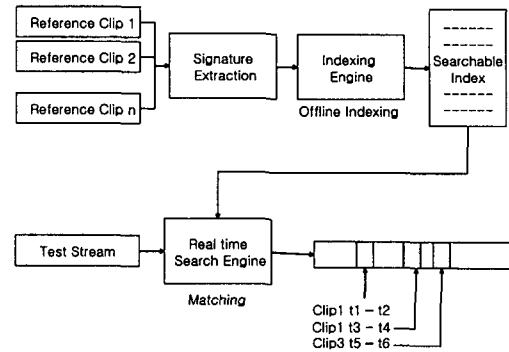


그림 1 비디오 복사검출 시스템 구성도

이들은 TV광고영역에 적용했으나 칼라 변화에 정합된 결과가 민감하였다. Hampapur 등이 제안한 방법은 에지특징을 사용하여 칼라변화에 강하게 하였다.[2] 그러나 이전의 방법들은 비디오 순서(video sequence) 고유의 중복성을 효과적으로 활용하지 못했다. 비디오의 시간적인(temporal) 성분을 연구한 Indyk 등은 비디오에서 장면변환(scene change)사이의 거리를 특징처럼 사용하였다.[4,6,7] 이는 실제로 적용하는데 제한이 있는 특징이다. Mohan은 비슷한 동작을 표현하는 비디오 검색을 위해 ordinal 특징을 사용하였다.[7] Naphade 등은 비디오클립을 정합하기 위해 효과적인 히스토그램 압축 방법을 제안하고 MPEG 비디오에서 DC순서(sequence)의 YUV 히스토그램의 교차부분을 사용했다.[4] 그들은 프레임 특징의 순서를 사용하였으나, 특징이 복사본에서 인코딩과 디지털화단계에서 칼라나 명암도 등의 변화를 가질 경우에 대한 실험을 하지 않았다.

3. 모션 방향성 히스토그램을 이용한 비디오 복사검출

비디오 복사검출을 위해 본 논문에서는 시간영역의 특성을 표현하는 모션특징에 대하여 기술하고자 한다. 모든 비디오 클립은 공간, 칼라와 시간적인 차원에 분포된 정보를 가지고 있다. 특징들은 이런 차원의 상대적인 중요도를 비교하여 선택된다.[8,9] 모션기반 특징은 비디오에서 변화하는 정보만을 대상으로 하고, 방향성 히스토그램은 클립을 1차원 그래프 형태로 변환하고, 프레임 순서 정합(frame sequence matching)을 사용하여 비디오의 시간성분을 반영하게 된다.

3.1 모션 방향(Motion DIRECTION)

모션 특징추출 과정의 전체적인 흐름을 설명하면 다음과 같다. 각 프레임은 $N = N_x \times N_y$ 개의 블록으로 나누고 각 블록에 대한 모션 벡터를 현재 프레임과 다음 프레임을 가지고 계산한다. 각 블록의 모션벡터는 Q개의 방향으로 양자화를 한다. 제안한 모션특징은 양자화된 모션방향을 한 프레임에 대해 통계를 구한, 즉 한 프레임의 특징은 각 Q 방향을 포함하는 모션벡터의 갯수이다. 만약 Q 방향이외에 움직임이 없는 블록이라면 모션벡터의 크기가 0이되고 블록은 0방향으로 정의된다. 예를 들면 프레임당 15×15 개의 블록으로 나누고 모션 벡터가 4방향으로 양자화 된다면 어느 t시간의 비디오 특징은 식(1) 및 (2)로 나타낼 수 있다.

$$S_m(t) = q_0(t), q_1(t), q_2(t), q_3(t), q_4(t) \quad (1)$$

$$q_i(t) = \{0, \dots, 225\}, i = 0, \dots, 4, \quad (2)$$

여기서 $q_i(t)$ 의 합은 255이다. 이는 비디오 특징을 프레임 당 5바이트로 표현할 수 있는 것으로 인덱스의 용량을 줄일 수 있다.

현재 프레임을 f_t , 후프레임을 f_{t+1} 이라고 하면 그림 2와 같이 방향성 모션추정과정을 표현할 수 있다. 프레임 f_t 의 블록 B에서 중심을 (x_t, y_t) 에 둔 블록크기(p_x, p_y)의 작은 명암도 조각을 P라고 둔다. 프레임 f_{t+1} 에 일치하는 블록 B에서 중심을 (x_{t+1}, y_{t+1}) 에 둔 추적영역(S_x, S_y)을 정의한다. 명암도 조각 P는 추적영역안의 모든 가능한 위치에 대해 절대화소차이(absolute pixel difference)의 합을 계산한다.

절대화소차이의 합은 조각의 유사도를 계산하기 위해 사용한다. 이들 중 최소값을 갖는 위치(Mx_{t+1}, My_{t+1})가 조각 P와 가장 유사하다고 가정하여, 국부적인 흐름의 변이벡터 $d_x = x_t - Mx_{t+1}$ 와 $d_y = y_t - My_{t+1}$ 를 구하고 방향 $\theta = \tan^{-1}(d_y / d_x)$ 를 계산한다. 이중 벡터의 방향만을 가지고 한 프레임에 대한 통계를 구한다. 차후 벡터의 크기를 양자화 하여 표현할 수도 있다. 벡터의 방향정보는 각 방향에 대한 한 클립을 움직임에 의한 1차원 그래프 모양으로 나타내는 것이다.

3.2 모션 방향성 히스토그램 정합

모션 히스토그램특징을 기준클립과 테스트클립 모두에서 추출하고, 특징을 정합하는 방법은 각 클립을 1차원 그래프 성분으로 표현하였으므로 1차원 그래프의 유사도를 비교하였다. 이는 테스트클립을 기준클립에 대하여 상충적분시킨 후 각 위치에서 정규화 상관계수(normalized correlation coefficient)를 구한다.[10] 이를 구한 후 최고의 상관계수 값을 갖는 위치가 최상정합이라고 판단한다. R(t)를 기준클립으로 T(t)를 테스트 클

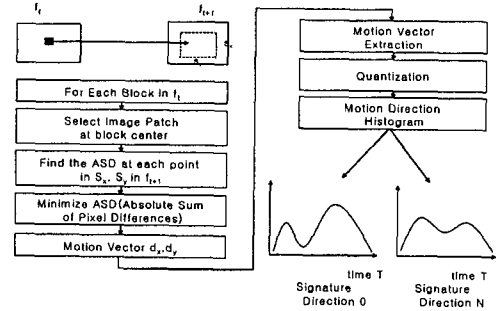


그림 2 모션특징 추출 순서도

립으로 두면, t주위의 테스트클립의 길이 L_T 윈도우에서 정규화상관계수 $C(t)$ 는 식(3)과 같다.

$$C(t) = \frac{E(R(t)T(t)) - E(R(t)) \times E(T(t))}{\sigma(R(t)) \times \sigma(T(t))} \quad (3)$$

여기서 확률 E는 다음 식(4)와 같다

$$E(S(t)) = \frac{1}{L_T} \sum_{i=t-L_T/2}^{t+L_T/2} [S(i)] \quad (4)$$

$\sigma(R(t))$ 와 $\sigma(T(t))$ 는 윈도우안에서 특징의 편차를 나타낸다. 그리고 $C(t)$ 에서 최고값은 t_{max} 이고 $R(t)$ 와 $T(t)$ 의 최적 정합이라고 판단한다.

4. 실험 결과 및 고찰

실험을 위해 비디오의 움직임 변화가 많은 29"의 TV 광고 3가지를 연결한 테스트 비디오 클립을 2' 09"의 길이와 240×180크기의 AVI파일로 디지털화하고, 각각의 TV광고는 352×240크기의 MPG파일로 기준 클립을 표 1과 같이 구성하였다. 이상적인 복사검출 특징은 초당 30프레임인 비디오 프레임율의 변화와 영상의 크기변화에 무관한 특성을 포함해야 하므로 두 비디오 영상의 해상도를 다르게 인코딩 하였다. 테스트 비디오클립에 관한 구조를 그림 3에 나타내었다. CF1은 테스트비디오의 88프레임부터 981프레임에 위치하고 있으며, CF2는 995부터 1882프레임, CF3는 1906부터 2797프레임에 각각 위치하고 있다. 이는 모션 방향성 히스토그램에 의해 순서 정합하므로써 각 CF의 위치를 확인하게 된다.

표 1 실험데이터

Video title	Resolution	Usage	Length	Test 클립안의 위치
CF 1	352×240	Reference 1	29"	88-981
CF 2	352×240	Reference 2	29"	995-1882
CF 3	352×240	Reference 3	29"	1906-2797
CF1+CF2+CF3	240×180	Test	2' 09"	-

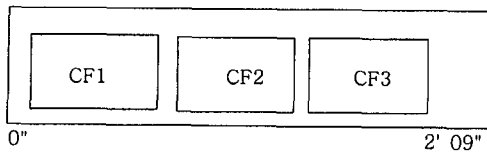


그림 3 테스트 비디오 클립의 구조

다음은 특징정합을 위한 실험 과정을 설명하였다.

1. 기준 비디오(CF1, CF2, CF3)로부터 모션방향성 특징을 추출한다.
2. 테스트 비디오(T)로부터 같은 특징을 추출한다.
3. 테스트 비디오 T에서 임의의 점(P)을 선택한다.
4. P주위로 기준 비디오 CF1 길이의 클립(C)을 선택한다.
5. T에 대해 C의 최적 정합위치 M_i 를 구하고 정합 스코어 M_s 를 계산한다.
6. 4-6단계를 CF2, CF3에 대해 반복한다.

그림 4는 테스트 비디오 클립의 4방향 모션 방향성 히스토그램을 구한 것이다. 가로축은 프레임번호이고, 세로축은 각 프레임에서 모션의 크기이다. 이 값이 클수록 프레임의 블록에서 각 방향의 움직임이 많은 것이고, 작으면 움직임이 없는 것이다.

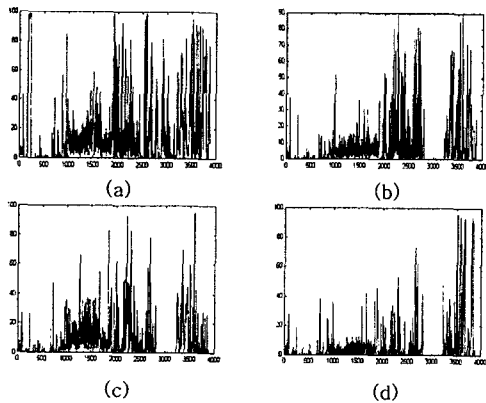


그림 4 테스트클립의 4방향 모션방향성 히스토그램

- (a) 모션 방향 0 (b) 모션 방향 1
(c) 모션 방향 2 (d) 모션 방향 3

결과에 따르면 테스트비디오 클립은 모션 방향성 히스토그램 특징에 의해 1차원 그래프 형태로 표현된 것이다. 그림 5는 CF1클립에 대한 모션 방향성 히스토그램을 나타내었다. 테스트 비디오클립의 결과에서 88프레임부터 981프레임의 1차원 그래프 형태는 비디오 해상도의 차이에도 불구하고 CF1결과와 유사한 모양을 나

타낸다. 그림 6은 테스트 클립결과와 CF1결과와 각 방향별 정규화 상관계수를 나타내었다. 계수의 최고치는 559번째 프레임으로 이는 CF1의 위치를 나타낸다. 그림 7은 각 CF의 방향별 정규화 상관계수를 더하여 구한 최종 정합 그래프이다. 그림7의 (b)는 CF2의 경우 1465번째 프레임에서 (c)는 CF3의 2366번째 프레임에서 각각 최고 정합결과를 얻었다.

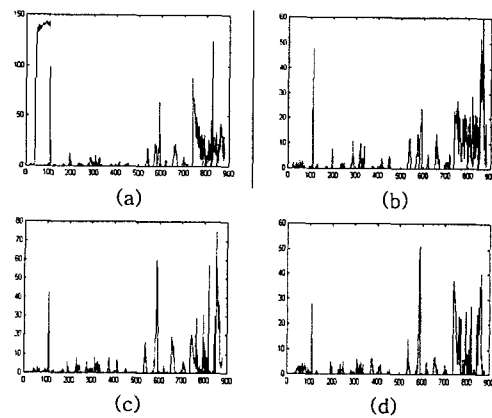


그림 5 CF1 클립의 4방향 모션방향성 히스토그램

- (a) 모션 방향 0 (b) 모션 방향 1
(c) 모션 방향 2 (d) 모션 방향 3

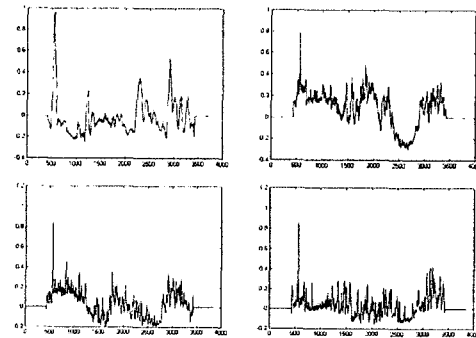


그림 6 테스트 클립결과와 CF1결과와 각 방향별 정규화 상관계수

최종 정합결과를 표 2에 나타내었다. 최대 정합 프레임과 각 CF클립의 중심과는 이상적인 경우 일치하지만, 약 30프레임의 오차가 발생한다. 이는 테스트 비디오에 각 CF를 삽입하는 편집과정에서 발생한 오차라 추정된다.

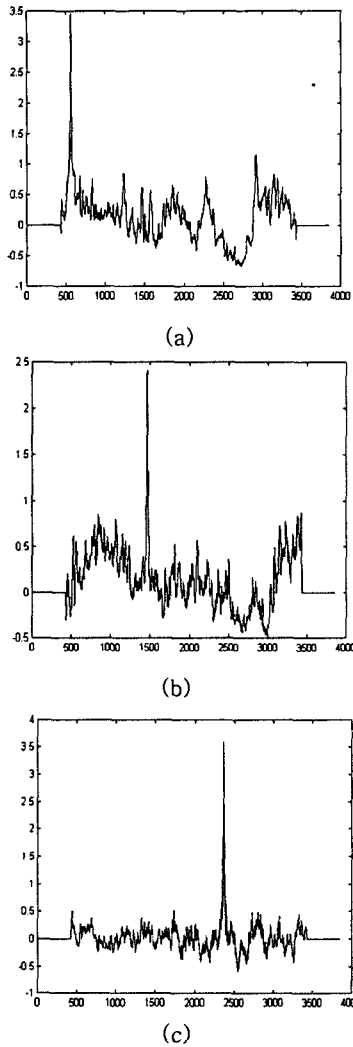


그림 7 각 CF와 테스트클립의 정규화 상관계수
 (a) 방향별 정규화 상관계수의 합 (CF1)
 (b) 방향별 정규화 상관계수의 합 (CF2)
 (c) 방향별 정규화 상관계수의 합 (CF3)

모션 방향성 히스토그램 특징의 성능은 ROC (Receiver operating characteristics) 곡선으로 나타내었으며, 곡선은 false positive(F_{pr})와 false negative(F_{nr})의 비를 나타낸다. N_T 가 실행한 정합테스트의 전체 개이고 τ 를 정합 임계치, F_n 은 false negative의 개수이고 F_p 는 false positive의 개수일 때 F_{pr} 과 F_{nr} 은 식(5)와 같다.

$$F_{pr}(\tau) = \frac{F_p}{N_T}, F_{nr}(\tau) = \frac{F_n}{N_T} \quad (5)$$

표 2 테스트 비디오와 순서 정합 결과

Video title	Test 클립안의 위치	Medium of clip	Maximum Match
CF 1	88 - 981	534-535	559
CF 2	995 - 1882	1438-1439	1465
CF 3	1906 - 2797	2351-2352	2366

ROC곡선은 τ 의 값을 최소값부터 최고값까지 5%씩 증가시키면서 구한 것이다. 좋은 ROC곡선일수록 축에 가깝게 된다. 즉 임의의 임계치에 대해 F_{pr} 과 F_{nr} 은 0에 가깝다. 이상적인 경우 $F_{pr} = F_{nr} = 0$ 이다.

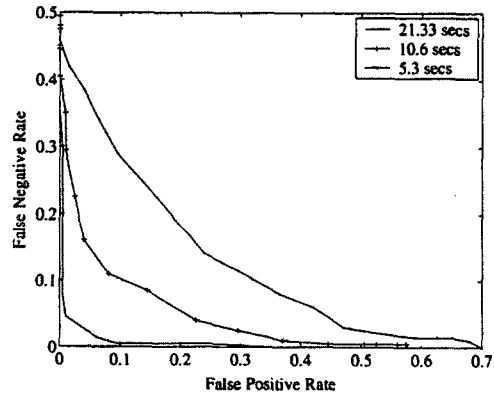


그림 8 모션특징 : false positive vs. false negative 비

4.1 고찰

그림 8은 모션기반 정합 결과를 나타낸다. 각 ROC곡선은 클립길이 L에 대한 결과이다. 기대한 대로, 정합성능은 클립길이가 증가할수록 개선된다. 표 3은 다른 클립길이에 대하여 100번을 수행한 결과 정합 에러율의 최소 값을 나타낸다. L이 20초의 경우 7%의 에러율을 나타내었다.

표 3 클립 길이변화에 대한 F_{pr} 과 F_{nr} 의 값

L	20.3s	10.6s	5.3s	2.6s	1.3s
Motion	6.1	8.11	24.14	-	-

모션 방향성 히스토그램 특징은 시간영역에서 명암도의 상대적인 변화를 포함하기 때문에 비디오의 엔코딩 시 발생하는 칼라변화에 강하다. 또한 비디오 클립의 길이가 20초정도의 TV 광고를 검출하는 목적에 적합하다. 그러나 공간영역의 정보를 가지지 못하므로 비디오 프레임의 중요한 부분들이 모션정보를 포함하지 않을 경

우 에러율이 증가할 수 있다. 시간 영역에서의 특성을 나타내는 모션 특징은 차후 공간영역의 특징과 결합한 새로운 특징에 대한 연구도 필요하다.

5. 결론

기존의 비디오 복사검출은 키 프레임에 의존한 이미지 정합방법을 사용하였으나, 본 논문은 비디오의 움직임 통계를 표현하여 클립을 나타내는 모션 히스토그램 특징을 제안하였으며, 연속적인 비디오 프레임의 특징을 포함한 순서정합(sequence matching)방법을 사용하였다. 제안한 알고리즘의 타당성을 보이기 위해 연속적으로 연결한 TV광고에서 각 광고를 정확하게 검출하였으며 ROC곡선 성능을 분석하였다.

분석결과 순서정합의 클립에서 모션 방향성 히스토그램은 실시간 처리를 위한 간단한 처리방법, 20초 정도의 짧은 클립에 대하여도 우수한 에러율, 다음단계인 인덱스를 위한 비교적 적은 저장용량 등을 보였다. 앞으로 다른 형태의 비디오에 대해 적용할 예정이고, 공간 특징과 결합하는 새로운 모션 특징의 변형에 대해서도 지속적인 연구가 필요하다. 또한 현재의 순서 정합방법보다 기존 스트림을 병렬로 구현할 수 있는 인덱스 기법을 개발해야 한다.

참고 문헌

- [1] C.L.E. Chang, J. Wang and G. Wiederhold, "Rime: A replicated image detector for the world wide web", in SPIE Multimedia Storage and Archiving Systems III, Nov. 1998.
- [2] A Hampapur and R. M. Bolle, "Feature based indexing for media tracking", in Proc. of Int. Conf. on Multimedia and Expo, pp. 67-70, Aug. 2000.
- [3] C.K.R Lienhart and W. Effelsberg, "On the detection and recognition of television commercials", in Proc. of the IEEE Conf. on Multimedia Computing and Systems, 1997.
- [4] M.Y.M. Napphade and B.-L.Yeo, "A novel scheme for fast and efficient video sequence matching using compact signatures", in Proc. SPIE, Storage and Retrieval for Media Database 2000, Vol. 3972, pp. 564-572, Jan. 2000.
- [5] J.V.J.M. Sanchez, X. Binefa and P. Radeva., "Local color analysis for scene break detection applied to tv commercials recognition.", in Proceedings of Visual 99, pp. 237-244, June 1999.
- [6] G.I.P Indyk and N. Shivakumar, "Finding Pirated video sequences on the internet". in Stanford

Infolab Technical Report, Feb. 1999.

- [7] R. Mohan, "Video sequence matching", in Proceedings of the International Conference on Audio, Speech and Signal Processing Society, 1998.
- [8] Z. Aghbari, K. Kaneko, and A. Makinouchi, "A Motion-Location Based Indexing Method for Retrieving MPEG Videos", ----
- [9] M. Ioka and M. Kurokawa, "A method for retrieving sequences of IMA", in Proc. SPIE, Storage and Retrieval for Media Database 1992, Vol. 1662, pp. 35-46, Feb. 1992.
- [10] D.Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, 1982.
- [11] S.-C. Cheung and A. Zakhor, "Estimation of web video multiplicity", in Proc. SPIE-Internet Image, Vol. 3964, pp. 34-6, 2000.
- [12] w. Contentwise Inc
- [13] D. Bhat and S.Nayar, "Ordinal measures for image correspondece", in IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 Issue: 4, pp. 415-423., April 1998.
- [14] M. Swain and D. Ballard, "Color indexing", in International Journal of Computer Vision, Vol. 7, No. 1, pp. 11-32, 1991.
- [15] A. Hampapur and R. M. Bolle, "Comparison of distance measures for video copy detection", in Proc. of Int. Conf. on Multimedia and Expo, Aug. 2001.



현 기 호

1986년 경북대학교 전자공학과(공학사).
1990년 경북대학교 전자공학과(공학석사).
1994년 경북대학교 전자공학과(공학박사).
1990년~1993년 대현정밀(주) 자동화개발 연구소 연구실장. 1993년~1997년 마산창 신대학 전자과 조교수. 2001년 IBM T.J. Watson Research Center 방문교수. 1997년~현재 영산대학교 컴퓨터정보공학부 조교수. 관심분야는 컴퓨터비전, 비디오 신호처리, 영상처리, 패턴인식 등



이 재 철

1986.2 경북대학교 전자공학과 졸업.
1988.2 경북대학교 전자공학과 석사과정 졸업. 1988.3~한국 원자력 연구소 연구원. (인공지능 및 로봇의 원자력 응용 분야 연구)