논문 2011-48SP-4-12

움직임과 영상 패턴 서술자를 이용한 중복 동영상 검출

(Detecting near-duplication Video Using Motion and Image Pattern Descriptor)

진 주 경*, 나 상 일**, 정 동 석***

(Jukyong Jin, Sangil Na, and Dong-Seok Jenong)

요 약

본 논문은 대용량 동영상을 관리하기 위한 빠르고 효율적인 내용기반 중복 동영상 검출 알고리즘을 제안한다. 효율적인 중복 동영상 검출을 위해 대용량의 동영상을 처리하기 쉬운 작은 단위로 나누는 동영상 장면 전환 기반 분할 기술을 적용하였다. 동영상 서비스 및 저작권 보호 관련 사업모델의 경우, 필요한 기술은 아주 작은 구간의 동영상이나 한 장의 영상 을 검색하기보다는 상당한 길이 이상 일치하는 동영상을 파악하는 기술이 필요하다. 이러한 중복 동영상 검출을 위해 본 논문에서 동영상을 장면 전환을 기준으로 분할하여, 나누어진 장면 내에서 움직임 분포 서술자와 대표 프레임을 선택하여 프레임 서술자를 추출한다. 움직임 분포 서술자는 동영상 디코딩 과정에서 얻어지는 매크로 블록의 움직임 벡터를 이용한 장면 내 움직임 분포 히스토그램을 구성하였다. 움직임 분포 서술자는 정합시 고속 정합이 가능하도록 필터링 역할을 한다. 반면 움직임 정보만는 낮은 변별력을 가진다. 이를 높이기 위해 움직임 분포 서술자를 이용하여 정합된 장면간에 선택된 대표 프레임의 패턴 서술자를 이용하여 동영상의 중복 여부를 최종 판단한다. 제안된 방법은 실제 동영상 서비스 환경에서 우수한 인식률과 낮은 오인식률을 가질 뿐만아니라 실제 적용이 가능할 정도의 빠른 정합 속도를 얻을 수 있었다.

Abstract

In this paper, we proposed fast and efficient algorithm for detecting near-duplication based on content based retrieval in large scale video database. For handling large amounts of video easily, we split the video into small segment using scene change detection. In case of video services and copyright related business models, it is need to technology that detect near-duplicates, that longer matched video than to search video containing short part or a frame of original. To detect near-duplicate video, we proposed motion distribution and frame descriptor in a video segment. The motion distribution descriptor is constructed by obtaining motion vector from macro blocks during the video decoding process. When matching between descriptors, we use the motion distribution descriptor as filtering to improving matching speed. However, motion distribution has low discriminability. To improve discrimination, we decide to identification using frame descriptor extracted from selected representative frames within a scene segmentation. The proposed algorithm shows high success rate and low false alarm rate. In addition, the matching speed of this descriptor is very fast, we confirm this algorithm can be useful to practical application.

Keywords: content-based video retrieval, motion descriptor, video matching, near-duplicate retrieval

Ⅰ. 서 론

컴퓨터와 영상 장비 및 통신 기술의 발달로 동영상 데이터의 사용이 급증하여 이에 맞추어 주문형 동영상, 디지털 라이브러리 등의 다양한 형태의 동영상 서비스 방식이 네트워크상에서 발달되고 있다. 그러나 동영상 데이터는 시간적, 대용양적 특성을 지니고 있어 효율적

* 학생회원, *** 정회원, 인하대학교 전자공학과
(Dept. of Electronic Engineering, Inha University)
** 학생회원, 한국전자통신연구원 콘텐츠연구본부
(Contents Research Division, ETRI)
※ 본 연구는 지식경제부, 문화체육관광부 및 한국 산업 기술평가관리원의 IT산업원천기술개발사업의 일환으로 수행하였음. [Rich UCC 기술 개발, 2010]
접수일자: 2010년10월28일, 수정완료일: 2011년5월9일

인 접근과 이용이 어렵다. 그로인해 동영상 내용기반을 통하여 다양한 방법의 동영상 처리 기술이 제안되고 있다. 동영상 응용분야 중 검색은 기존의 키워드나 동영상 자체의 메타데이터를 이용한 방법에는 성능의 한계로 인하여 동영상 자체의 내용을 기반으로 하는 검색 기술이 대두되고 있다. 대용량 동영상의 전체 혹은 선택된 프레임의 색상(color), 에지(edge) 및 질감(texture) 등에 관한 정보를 이용한 서술자를 만드는 접근법[1~3]이 주류를 이루고 있다. 이는 정지 영상에 대한 내용기반 검색 시스템에서 사용되었던 여러 가지 특징들을 동영상에 적용한 기술들이다. 이러한 기술들과 더불어 동영상만이 갖는 시공간 차원에서 움직임 정보를 사용하는 기술들이 활발히 이루어지고 있다[4~6]. 그러나 움직임 정보는 다른 기술에 비해 추출 및 정합 어려울 뿐만 아니라 변별력(discriminability)과 정확도(accuracy)의 균형을 맞추기 어렵다. 즉, 움직임 정보만을 이용한 동영상 내용 기반 검색에는 어려움이 있다. 예를 들어, 정적인 장면들의 연속인 동영상의 경우 움직임 정보만으로는 검색하는데 한계가 있다. 그로 인해 움직임을 이용한 서술자보다 세밀하게 구분할 수 있는 다른 정보의 내용 기반 서술자도 필요하다. 모든 프레임에서 내용기반 서술자를 추출하는 방법은 성능의 정확성에서는 좋은 결과를 가져 올 수 있으나 현실적으로는 추출과 정합 시 과도한 연산을 요구한다. 이러한 문제를 해결하기 위해 본 논문에서 다양한 동영상 변형에서 반복 재현성이 높은 프레임을 선택하여 선택된 극히 일부의 동영상을 대표하는 프레임에서만 내용기반 서술자를 추출하였다.

지금까지 제안된 내용기반 동영상 검색을 위한 서술자들은 큰 서술자 크기, 추출시 과도한 연산 및 낮은 정합 속도 때문에 동영상관련 서비스에서 쉽게 동영상 검색 기술을 적용하는데 어려움이 있었다. 동영상 검색 기술을 실제 적용하기 위해서는 작은 용량의 서술자이며 빠른 정합 및 추출 속도의 동영상 서술자가 필요하다. 이를 위해 동영상의 시공간 상관관계를 이용하여 대용량 동영상 데이터를 다루기에 편리하도록 시간적으로 분할하여 사용한다. 동영상 분할은 장면이 전환되는 장면전환점을 기준으로 분할한다. 일반적으로 동영상 장면전환 검출의 가장 기본 단위로 샷(Shot)을 사용한다. 본 논문에서 내용기반 장면전환 검출의 다양한 방법 중 프레임간의 히스토그램을 차이 값을 이용한 방법을[10] 이용하였다.

기존에 제시된 중복 동영상 검출 및 제거 방법으로 모든 또는 일정한 간격마다의 프레임에서 변형에 강한 작은 데이터를 추출하는 방법[8, 12]과 대표프레임을 선택하여 특징점(key point)을 찾고 특징점 주변에 특징 벡터를 이용하는 방법이 있다[9]. 이러한 방법은 주로 정합하는 동영상의 길이가 작을 때는 만족할만한 정합 속도 및 성능을 얻을 수 있으나 긴 동영상에서 현저히 성능에 비하여 정합 속도가 낮아진다. 또한 기존의 서술자를 이용한 정합 방법들은 질의 동영상이 원본 영상에 정확히 일부분에 해당된다는 가정이 들어가 있다. 이는 실제 동영상 검색에서 많은 문제를 제기한다. 예를 들어, 방송에서 캡쳐된 동영상에 경우 여러 가지 편집에 의해 시작과 끝이 다르다. 앞, 뒤 및 중간의 다른 콘텐츠가 삽입된 경우가 많다. 결과적으로 서술자 정합시에 질의 동영상이 원본 동영상의 정확히 연속된 일부분이라는 가정은 타당하지 않다. 또한 정합시 동영상의 프레임율(frame rate)이 동일하다고 가정하지만 실제로 같은 30 FPS(Frame Per Second)를 갖는 동영상일지라도 변형 과정을 통하여 29, 31 fps로 왜곡되어 몇 분마다 1, 2초 이상의 위치 차이를 보이는 경우가 많기 때문에 이러한 시간 정보의 허용오차를 감안한 정합 방법이 필요하다.

본 논문에서는 장면분할을 이용하여 동영상을 샷으로 분할한 후 각 장면 내에서 동영상 디코딩(decoding) 시 얻어지는 매크로블록(macro-block)의 움직임 벡터와 프레임을 이용하여 강인하고 빠른 중복 동영상 검출 알고리즘을 제안한다. 또한 프레임 서술자는 제안하는 16가지 저주파 패턴 마스크를 이용하여 구성한다. 다음 Ⅱ장에서 제시한 움직임과 프레임기반 서술자들에 대한 내용이고, Ⅲ장은 새롭게 제안하는 서술자 정합 방법이다. Ⅳ, Ⅴ장은 실험결과 및 결론으로 구성되었다.

Ⅱ. 움직임 분포, 패턴 서술자

1. 움직임 분포 서술자 (motion distribution descriptor)

내용 기반 동영상 검색분야에서 객체 및 카메라 움직임을 이용한 연구는[3, 5] 활발히 진행되고 있다. 이러한 움직임 내용 기반 접근법은 동영상 콘텐츠사용에서 발생하는 다양한 동영상 변형 대하여 강인한 특성이다. 그러나 움직임을 이용한 동영상 검색의 문제점은 크게 두 가지이다. 첫째, 움직임이 적거나 매우 복잡한 동영

상의 경우 유사도를 판단하는데 많은 어려움이 있다. 둘째, 동영상 프레임 크기에 따라 움직임 벡터의 크기가 달라짐에 정합시에 직접 비교할 수가 없다.

동영상에서 이웃하는 프레임사이에서 움직임정보를 얻기 위한 다양한 방법이 연구되었다[4~6]. 그러나 이러한 방법들은 객체 및 카메라 움직임을 얻기 위해 너무 복잡한 과정을 필요함에 실제 사용하기에 어려움이 있다. 그래서 이러한 복잡한 과정을 최소화하기 위해서 본 논문에서는 동영상을 디코딩시 얻어지는 매크로블록마다의 움직임 벡터만을 이용하여 움직임을 이용한 동영상 서술자를 구현한다. 움직임을 가지는 B, P 프레임 형태에서 오직 전방예측 움직임 벡터만을 이용하였다.

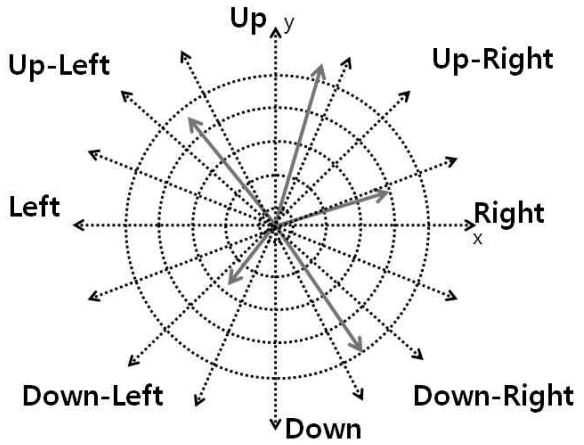


그림 1. 움직임 기반 장면내 방향, 크기 누적 히스토그램

Fig. 1. Cumulative histogram of motion based on direction and magnitude in a scene.

표 1. 양자화 와 히스토그램의 빈의 크기의 따른 에러율

Table 1. Error rate according to quantization and histogram bins.

Quantization	Bins	True	False	Minimum Error	threshold value
256	16	78.45	4.62	26.17	0.131
256	8	83.84	3.98	22.15	0.132
256	4	87.23	4.66	17.43	0.099
64	16	95.30	1.83	6.53	0.125
64	8	94.42	2.95	8.53	0.133
64	4	89.43	3.40	13.97	0.092
32	16	97.81	4.63	6.82	0.145
32	8	95.41	4.12	8.70	0.143
32	4	88.63	3.04	14.41	0.092
16	16	95.61	2.27	6.67	0.131
16	8	95.41	4.59	9.18	0.147
16	4	91.02	5.09	15.06	0.108

제안하는 움직임 분포 서술자는 기본단위로 장면 전환 검출방법을 사용하여 얻어진 각각 동영상의 분할 구간마다 서술자를 구성한다. 그림 1.과 같이 움직임 벡터의 방향성과 크기에 따라 장면내 누적 히스토그램을 구성한다. 구해진 히스토그램에서의 최대값을 갖는 빈(bin)을 기준으로 나머지 빈을 정규화한다. 이를 통해 서로 다른 프레임 크기를 갖는 동영상에 움직임 히스토그램을 비교할 수 있다. 양자화 단계의 크기가 256인 경우 히스토그램의 최댓값을 가지는 빈이 255가 되고 나머지 빈은 최댓값에 비례하는 값으로 표시한다. 움직임 분포 서술자 i와 j의 차이값 $d(i,j)$ 는 식 (1)과 같다. bin은 움직임 방향 도수, M_i는 0~255로 정규화된 히스토그램의 크기를 나타낸다.

표 1.은 각 움직임 히스토그램의 빈의 개수와 양자화 크기에 따른 정확도와 오류 정도를 나타낸다. 동영상에서 30초 정도의 구간을 떼어서 실제 동영상 콘텐츠에서 자주 발생되는 동영상 변형을 취하여 움직임 분포 서술자를 추출하여 비교한 결과이다. 에러가 최소화 되는 위치를 움직임 서술자의 문턱값으로 설정하였다. 변형 종류와 정도는 Ⅲ장 실험 결과에서 보인다.

$$d(i,j)=\sum_{k}^{bin}|M_i(k)-M_j(k)| \quad (1)$$

그림 2.는 양자화 16, 움직임 히스토그램 빈이 16개에 대한 움직임 분포 히스토그램 차이값 분포도를 나타낸다. 좌측 점선은 동영상 원본과 원본을 변형한 중복 동영상간 움직임 히스토그램 차이값 분포를 나타낸다. 우측의 실선은 서로 다른 동영상간의 움직임 히스토그램의 차이값을 나타낸다. 변형비교에 경우 문턱값을 기준

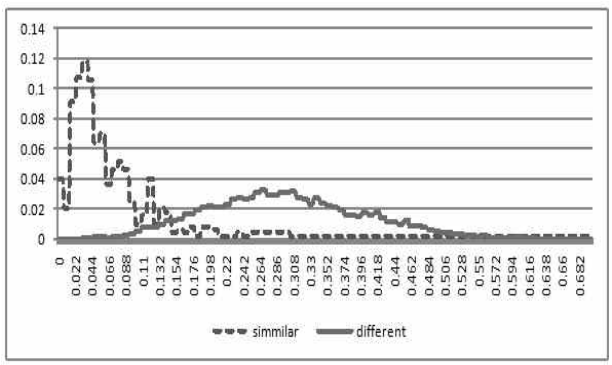


그림 2. 중복 및 이종 동영상 움직임 히스토그램 차이값 분포 (Quantization: 16, Direction :16)

Fig. 2. Distribution of distance between near-duplicate and different video using motion histogram.

으로 우측은 에러가 되고 이종 동영상 비교에서 좌측이 에러가 된다. 이 두 가지 에러율을 이용하여 가장 작은 위치를 움직임 히스토그램의 문턱값으로 결정하였다.

표 1.에서 보이는 바와 같이 최소 에러를 가지는 3가지 조합을 어두운 명암으로 표시하였다. 본 논문에서는 그중 최종 양자화 16, 히스토그램 빈 16개를 선택하였다. 성능뿐만 아니라 양자화 크기가 16이면 한 바이트에 2개에 빈의 값을 넣을 수 있어 한 동영상 분할 영역마다 16개 빈을 가지더라도 8 바이트만의 작은 서술자를 가진다. 매칭 속도의 향상을 위해 4바이트만 사용하는 양자화 16에 방향성 8도 사용 할 수 있다.

2. 외곽 박스(outer box) 삽입 검출

로컬 특징 기반 동영상 검색과는 다르게 프레임 영상 전체를 사용하여 서술자를 만드는 전역 특징 알고리즘에서는 대부분 필러(pillar), 레터(letter)박스 같은 외곽 박스(outer box) 삽입에 대한 고려가 필요하다. 순차측정(ordinal) 프레임 서술자[13] 경우 외곽 박스를 왜곡을 해소하기위해 영상을 2x2로 나누어 4영역의 평균에 대한 랭킹만을 서술자로 이용하였다. 그러나 영상을 너무 간략화 하기 때문에 프레임 서술자의 변별력이 떨어지는 문제가 있다.

외곽 박스 삽입 변형에 대한 문제를 해결하기위해 그림 3.(a)와 같이 수직, 수평 라인상의 화소만을 이용하여 시공간 슬라이스를 동영상 분할 구간마다 구성한다. 시공간 슬라이스를 분석하면 3개의 장면으로 구성됨을 알 수 있다. 그림 3.(b)는 각 장면마다의 대표영상이다. 좌, 우에 4:3 화면비율을 16:9로 구성하기위해 인위적으로 필러박스를 삽입한 경우이다. 시공간 슬라이스를 통하여 동영상에 외곽 박스가 삽입된 여부를 직관적으로 판단할 수 있다. 그림 3.(a)로부터 수직, 수평 단위에 라

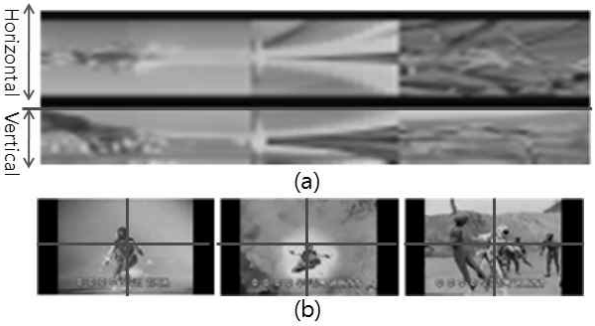


그림 3. 외곽 박스 삽입 시 시공간 슬라이스
Fig. 3. Spatio-temporal slice in case of outer box insertion.

인별 평균, 표준편차 값을 구한다. 식 (2)를 이용하여 라인 화소 평균값과 표준 편차 값이 정해진 문턱값 이하면 외부 박스 삽입으로 간주하여서 외곽 삽입 박스를 제거하였다. $width$는 시공간 슬라이스에서 가로 화소 수를 나타내며 장면내 프레임 수에 해당된다. S_h는 장면 분할 구간의 슬라이스 영상, $m_h(i)$는 i번째 라인의 평균 화소 값을 나타낸다. σ_h는 라인의 표준 편차를 나타낸다.

$$m_h(i) = \sum_{j}^{width} S_h(j)/width$$
$$m_h(i) < th_m, \sigma_h(i) < th_\sigma \quad (2)$$

실험을 통하여 라인 평균 임계값(th_m)은 25, 표준편차(th_σ)는 5로 설정하였다. 수직 시공간 슬라이스 영상에 대해서도 동일하게 적용하여 외곽 박스를 제거한다.

3. 프레임 패턴 서술자(motion descriptor)

움직임 서술자를 보완하기 위해 장면내 선택된 프레임에서 그림 4.와 같이 제안한 16개의 패턴 마스크를 이용하여 프레임 패턴 서술자를 구성한다. 동영상 서비스과정에서 발생하는 변형에는 영상의 일부만 발취하는 크롭(crop)이나 시차의(parallax)변화는 거의 없다. 본

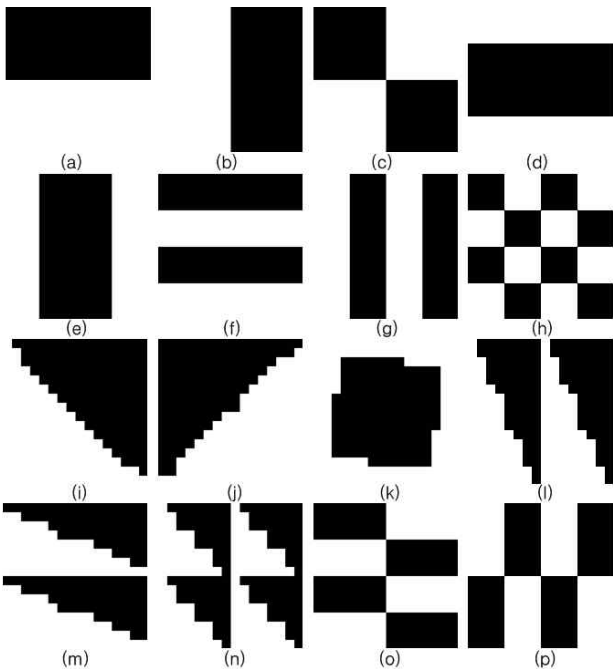


그림 4. 16개의 마스크 패턴 이미지
Fig. 4. 16 pattern mask image.

논문에서 중복 검출 및 검색 시나리오에서는 프레임 서술자로 영상 전역을 이용할 수 있다.

$$P(i)=\sum_{x,y}^{32} I(x,y)\times Mask(x,y) \quad (3)$$

외곽박스 제거 후 프레임을 32x32 화소 크기로 축소한다. 그림 4.는 축소한 프레임 영상과 같은 크기의 -1과 1로 구성된 마스크이다. 흰색은 1을 검정영역은 -1을 나타낸다. 제시한 마스크는 프레임과 식 (3)과 같이 축소된 영상과 중첩 연산 $P(i)$를 실행한다. 패턴 연산값 $P(i)$를 양수와 음수로 구분하여 양수의 경우 1, 음수의 경우 0으로 지정한다. 16개의 패턴으로부터 총 2바이트의 해쉬(hash)화한 대략적 프레임 기술자를 얻을 수 있다. 프레임 서술자간에 정합시 이를 우선 비교하여 나머지 정합을 수행할지의 여부를 판단하는 필터링 역할을 수행한다. 추가적으로 프레임 서술자는 16개의 서로 패턴 연산값 $P(i)$를 그림 5.의 의사코드(pseudo code)처럼 패턴 연산값과 연산값의 절대 값을 비교를 통하여 상대값이 크고 작음을 이용하여 해쉬화한다. 총 240비트의 프레임 서술자를 뽑아낼 수 있다. 모든 프레임 서술자는 그레이(gray) 채널에서만 이루어진다. ABS(x)는 절대 값을 의미한다.

```
loop j=1+1 to 16
  if P(i) > P(j)
    bits = 1
    bits << 1
  if ABS( P(i)>P(j))
    bits = 1
    bits << 1
end loop
```

그림 5. 프레임 서술자를 위한 의사 코드
Fig. 5. pseudo code for frame descriptor.

4. 대표 프레임 선택

프레임 서술자의 경우 동영상의 모든 프레임에서 추출하기에는 동영상의 특성인 대용량으로 인해 많은 저장 공간이 필요하다. 또한 서술자 정합시에도 모든 프레임을 정합하기는 많은 연산시간을 소요된다. 동영상 특성상 유사한 프레임의 반복의 경우가 높고 같은 장면 내에서는 유사성이 높다. 이를 이용하여 반복도가 높은 특정 프레임을 선택하여 서술자를 추출함이 효율적이다. 이전 연구들에서 동영상의 특성인 시간적 중복 성을 이용하여 변화가 정해진 값 이상 발생했을 때마다 서브 샘플링을 하는 방법[14]이나 일정한 시간마다 샘플링하는 방법[13]이 제안되었다.

본 논문에서는 장면전환 검출에서 사용되는 이웃하는 프레임의 히스토그램 차이값이 국소 최대값(local maxima)을 가지는 프레임을 선택하여 프레임 서술자를 추출하였다. 하나의 장면분할 내에서 길이에 따라 최대 3 프레임까지 선택한다. 주로 장면전환이 일어난 다음 프레임을 선택하며 상당히 긴 장면길이를 가지는 경우는 최대 3장까지 장면전환 수치가 높은 프레임을 추가로 선택한다. 이는 장면전환 오류로 인한 대표 프레임 불일치를 해결하는데 도움을 준다.

5. 서술자 정합 (matching descriptor)

대부분의 동영상 서술자 연구에서는 질의 서술자를 원본 서술자의 정확한 일부라는 가정 하에 정합을 수행한다. 그림 6.(a)는 대부분의 연구에서 제안하는 질의 동영상의 구조를 나타낸다. 그러나 실제 서비스되고 잇는 동영상을 분석해보면 영상에 가해지는 변형뿐만 아니라 영상 앞, 중간, 뒤에 다른 콘텐츠가 삽입되어 있는 그림 6.(b)과 같은 경우가 흔히 존재한다. 또한 실제 동영상 검색 환경에서 대부분의 서술자 정합은 동영상의 시간정보 즉, 초당 프레임율을 이용하여 동영상 서술자 정합을 수행한다. 그러나 이는 실제 동영상 서비스 환경에서 차이가 있다. 중복된 동영상일지라도 변형되는 과정에서 실제 프레임율이 다른 경우가 많다. 10초마다

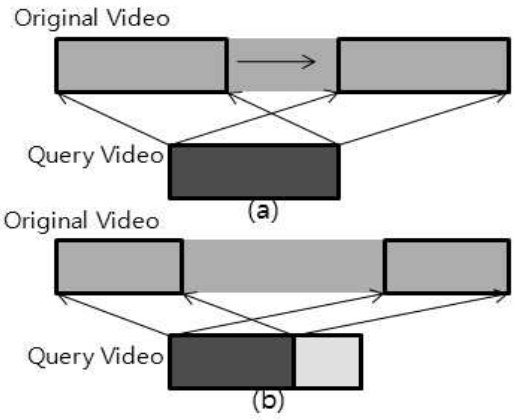


그림 6. 동영상 서술자 질의 방법
(a) 직접 질의 (b) 부분 질의
Fig. 6. matching scenario for video descriptor.
(a) direct matching (b) partial matching

3000장 동영상이 재생되는 것이 10초마다 3030또는 2970장 재생으로 왜곡되는 경우가 흔하게 발생한다. 이러한 프레임율 왜곡은 작은 길이의 질의 동영상 정합시에는 큰 문제가 되지 않지만 긴 길이의 동영상에서는 정합시 오류의 원인을 제공할 수 있으므로 동영상에 시간정보를 이용함은 신중해야 한다. 본 논문에서는 프레임율의 허용 오차를 두어 최종 정합 과정에 이를 반영한다.

본 논문에서 동영상 서술자 정합시 우선 각각 서술자에서 대표되는 20개에 움직임 분포 서술자를 선출한다. 선출된 서술자는 움직임 활동 정도가 강한 분할구간을 기준으로 전체 동영상의 전역에서 균일하게 선택한다. 선택된 20개에 움직임 분포 서술자끼리 정합을 실행한다. 움직임 분포 서술자가 유사한 경우 장면 내 선택된 프레임을 기준으로 프레임 서술자 정합을 실행한다. 프레임 서술자 정합 또한 설정된 문턱값 이하에 프레임이 나올 경우, 정합 된 프레임을 기준으로 전, 후 설정된 최소 정합 길이에 비례하여 구간을 확장하여 구간내 모든 프레임끼리 정합을 시도한다.

그림. 7은 정합 구간 내에서 정합된 프레임 쌍의 위치를 x로 표시하였다. 이를 이용하여 정합 프레임을 점으로 간주하여 2차원 공간상에 일차원 선형 정합(line fitting) 실행하여 직선의 기울기를 얻고 기울기가 허용오차를 만족하면 정합 되었다고 본다. 또한 정합 점과 선형 정합에 의해 구해진 선과의 거리 평균을 이용하여 최종 정합도를 나타낸다. 선형 정합의 기울기 값은 0.9에서 1.1로 제한하여 이를 벗어나면 올바르게 정합 되지 않았다고 본다. 정상적인 중복 동영상끼리는 기울기가 1로 정합 되어야 하나 프레임율 왜곡 문제에 의한 정합오류를 감안하여 오차의 범위를 고려한 것이다. 그림 8.은 동영상 서술자 정합 순서도이다.

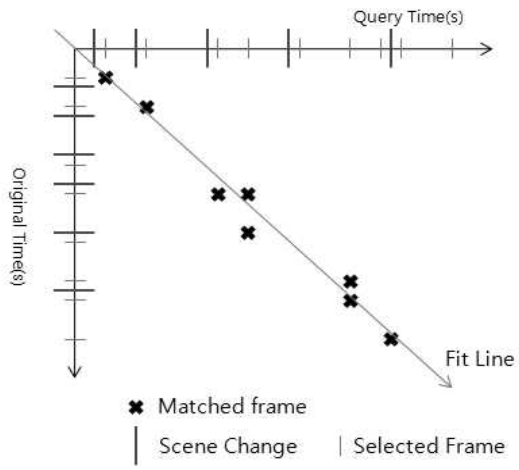


그림 7. 정합된 프레임들을 이용한 선형정합(line fitting)
Fig. 7. Line Fitting using matched frames.

Ⅲ. 실 험

제안된 중복 동영상 검출 방법 평가를 위하여 실제 국내 동영상 웹 서비스 사이트를 통해 서비스 되고 있는 다양한 장르의 실험 동영상 536개를 구성하였다. 시간은 최소 1분에서 최대 2시간에 내외로 구성되었고 해당 장르별 길이의 분포와 평균은 표 2.와 같다. 표 2.는 동영상 중복 검출에서 원본데이터의 역할을 한다. 표 3.은 이 원본 동영상을 이용하여 실제 자주 발생하는 동영상 변형을 기반으로 본 실험에서 사용된 변형 종류와 강도에 대하여 나타낸다. 실제 다운로드 및 서비스되는 대부분의 변형은 이러한 형태를 벗어나지 않는다. 또한 변형 질의 동영상은 제작시 원본에서 임의 지점에서 30% 잘라내어 그 자체를 질의로 사용하는 직접 방법 그리고 앞, 뒤에 다른 동영상을 붙여서 질의로 사용하는 간접 방법 두 가지로 나누었다. 그림 6.과 같이 직접, 부분 질의 정합으로 나누었다.

본 실험에서 그림 8.에 정합 순서도의 각 단계에서 사용된 문턱값들은 다음과 같다. $th_{mot}=20$, $th_{frm}=18$, $low_{slope}=0.9$, $Hi_{slope}=1.1$, $th_{dist}=7$.

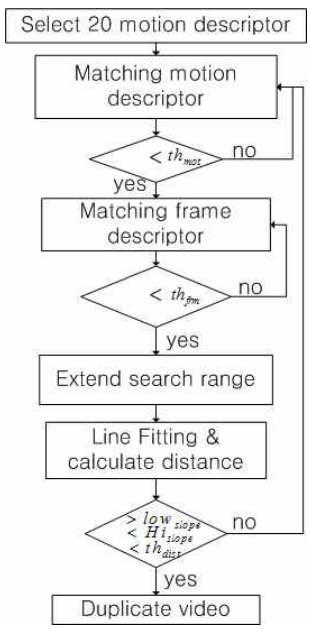


그림 8. 동영상 서술자 매칭 순서도
Fig. 8. Flow chart for matching video descriptor.

표 2. 실제 서비스되는 동영상 원본의 종류
Table 2. Video test set in real video service.

장르	갯수	시간(분)	평균
Animation	69	5~25	15
Movie	78	30~130	60
Documentary	64	15~60	40
drama	68	20~50	40
music	78	2~5	4
news	65	1~15	3
sports	50	3~50	20
UCC	64	1~15	5
합계	536	1~130	25

표 3. 동영상 변형 종류 및 정도
Table 3. Modification kind and degree.

Modification	Level	Contents
Brightness change(BC)	Light	+10 (Only Y channel)
	Heavy	-15 (Only Y channel)
Frame Reduction(FR)	Light	25,30 fps-> 15 fps
Pillar Letter Box(PLB)	Light	3:4->16:9, 16:9->4:3
Resize(RR)	Light	CIF(352x288)
	Heavy	QCIF(176x144)
Compression & Resize (CR)	Light	512kbps (CIF)
	Heavy	256kbps (CIF)
Text Logo Overlay (TLO)	Light	overlay 10%
	Heavy	overlay 20%

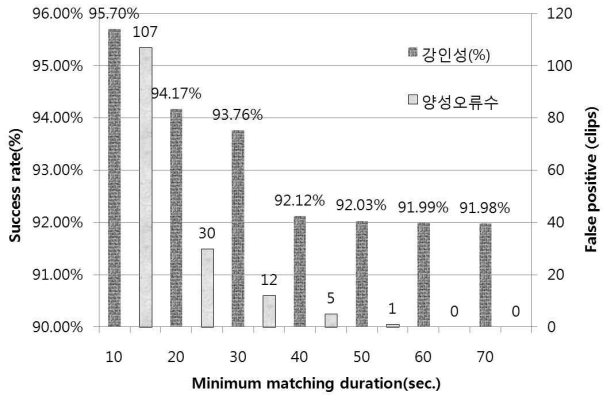
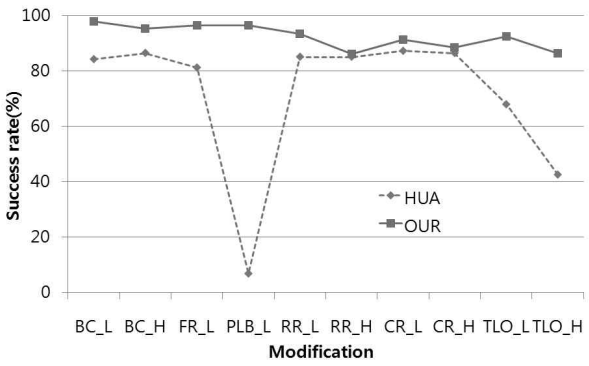
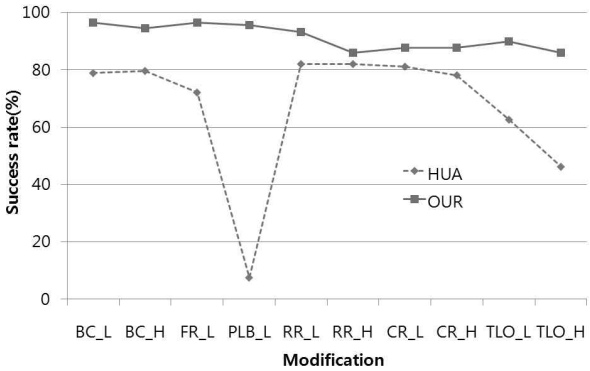


그림 9. 최소 정합 구간에 따른 인식률과 오인수 개수
Fig. 9. Success rate and false alarm according to minimum matching duration.

준비된 원본과 변형 동영상을 기준으로 강인성과 양성오류수(false positive)로 평가하였다. 그림. 9는 최소 정합 길이를 변수로 하여 10초에서 70초까지 실험한 결과이다. 강인성은 원본과 원본의 일부을 변형한 질의와의 정합시 성공률이며, 양성 오류수은 서로 다른 동영상간의 정합에서 성공되었다고 실수한 경우의 수이다.



(a)



(b)

그림 10. 질의 동영상 형태 및 변형에 따른 결과
(a) 직적질의 (b) 간접질의
Fig. 10. Result of success rate query video type and modification kind each. (a) direct query (b) partial query.

그림 9.에서 나타내듯 최소 정합길이가 60초 이상인 경우 오인식의 경우는 없어짐을 알 수 있다. 실험 결과뿐만 아니라 실제로 서비스되는 동영상에서 30초 내외의 경우에는 매우 유사한 영상이 흔히 존재한다.

비교 대상 알고리즘에는 프레임 서술자로는 영상을 3x3으로 나누어 블록의 랭킹을 이용하고 정합은 본 논문과 부분 질의 정합이 가능한 알고리즘인 Hua[13]방법과 비교하였다. 제시한 알고리즘은 양성 오류의 수가 없어지는 최소 정합 구간인 60초로 설정하였다. 그림 10.은 직접 질의와 부분 질의에 따른 강인성 결과이다.

표 4.는 비교 알고리즘과의 성능 평가 결과이다. 직접, 부분매칭을 나누어 강인성에 해당되는 인식률로 나타내었다. 평균 정합 속도는 일대일 동영상 서술자 끼리 정합시 평균 소요 시간을 나타낸다. 또한 서로 다른 동영상 3,055,500번 정합 중에서 잘못된 정합의 개수로 양성 오류 개수를 나타내었다. 전체적으로 제안한 알고리즘이 속도에서는 비교대상 알고리즘보다 455배 이상

표 4. 전체적 정합 결과
Table 4. Result of matching.

	Hua	Our
직접 정합 인식률(%)	71.29	92.43
부분 정합 인식률(%)	67.094	91.37
양성 오류 수(개수)	479	0
평균 매칭속도(ms)	32.62	0.087

빨랐고 인식률도 20%이상 높았다. 또 하나 제시하는 정합 방법의 장점으로 동영상의 길이의 거의 상관없이 정합시간이 일정하다. 그러나 비교대상 알고리즘의 경우 길이에 따른 시간 차이가 크다.

Ⅳ. 결 론

본 논문은 장면전환을 기반으로 분할된 동영상에서 매크로블록의 움직임 정보와 선택된 대표 프레임을 통하여 프레임 서술자를 이용하여 중복 동영상 검출 및 검색 방법을 제안하였다. 제안된 방법은 동영상의 프레임 크기 및 길이에 비례정도가 낮으며 빠르면서 우수한 정합 결과를 얻었다. 한 동영상에서 움직임의 활동성이 큰 20개만의 움직임 분포 서술자를 선택하여 사용한다. 선택되는 구간의 수를 늘이면 더 높은 인식률을 얻을 수 있으나 이 이상의 선택의 경우는 성능의 증가에 비해 정합 속도의 저하의 폭이 커진다. 그러나 더 높은 인식률이 필요한 시나리오라면 선택되는 움직임 분포 서술자를 늘려서 사용할 수 있다. 제안한 선택된 최대 20개의 움직임 서술자와 장면내 최대 3장의 프레임 서술자를 이용한 동영상 내용기반 중복 검출 방법은 실제 서비스에서 충분히 사용할 수 있는 정도의 인식률 및 정합 속도를 나타낸다.

본 논문에서 제시한 중복 검출 및 검색 시나리오는 실제 콘텐츠 환경에서 적용하기에는 설정의 문제가 발생한다. 최소 정합길이를 어떻게 설정하느냐의 문제이다. 중복 검출 시나리오의 특성상 충분히 긴 길이의 동일 부분을 원본과 질의 동영상이 가지고 있다. 실제 조사에 의하면 동영상간의 유사 구간이 상당히 많다는 것이다. 뉴스의 경우 20~30초 정도의 동일한 영상을 사용하여 방송하는 경우가 흔하다. 또한 같은 스튜디오에서 진행자의 의상만 바뀐 영상도 존재한다. 이러한 예는 중복 동영상이라고 간주하기 위해서는 충분한 길이의 구간에서 일치하여야 함을 알 수 있다. 그러나 요구하는 최소 정합길이가 길어질수록 오인식의 경우는 낮아지나 정합의 효율이 조금씩 떨어진다. 본 실험 결과에 의하면 60초 정도가 적당하다고 보았다.

비교 대상 알고리즘과의 실험에서는 특히 RR_H에서 인식률 낮게 나왔다. 이 변형은 프레임 사이즈 QCIF로 원본에 비해 최소 16배 이상 줄이는 변형이다. 그로인해 원본 영상에서 크지 않은 움직임 벡터는 작은 프레임사이즈로 인해 0의 값으로 수렴하기 때문에 움직임 서술자에 악영향을 주었다. 그러나 실제로 동영상 변형에서 이렇게 과도한 프레임 사이즈 변형은 자주 존재하지 않는다. 전체적 결과를 보면 속도와 인식률에서 충분히 높은 성능을 보이며, 특히 오인식률이 낮은 것은 실제 어플리케이션 상황에서도 안정적으로 사용할 수 있는 기술이라 하기에 충분하다.

참 고 문 헌

[1] V. E. Ogle, “Chabot :Retireval from a Relational Database of Image”, *IEEE Computer,* vol. 28, no. 9, pp. 40-48, Sep. 1995.

[2] A. Mojsilovic, J. Hu, “A Method for Color Content Matching of Images,” *Proc. of the 2000 Int. Conf. on Multimedia and Expo,* vol. 2, pp. 649-652, Jul. 2000.

[3] B. S. Manjunath and W. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Trans. Pattern Anal. Machine Intell.,* vol 18, pp.837-842, Aug. 1996.

[4] N. Dimitrova and F. Golshani, “Motion Recovery for Video Content Classification,” ACM Trans. on Information Sys., vol. 13, no. 4, pp. 408-439, Oct. 1995.

[5] S. Dagtas, W. Al-Khatib, A. Ghafoor and R. L. Kashyap, “Models for Motion-Based Video Indexing and Retrieval,” *IEEE Trans. on Image Processing,* vol. 9, no. 1, pp. 88-101, Jan. 2000.

[6] A. Yoshitaka, Y. Hosoda, M. Yoshimitsu, “VIOLONE : Video Retrieval by Motion Example,” *J. of Visual Languages and Computing,* vol. 7, no. 4, pp. 423-443, 1996.

[7] K. W. Lee, W. S. You and J. Kim, “Video Retrieval based on the Object′s Motion Trajectory,” *Proc. of SPIE in Visual Comm. and Image Processing,* vol. 4067, pp. 114-124, 2000.

[8] Kim, C., “Content-based image copy detection,” Signal Processing: Image Communication., vol. 18, no. 3, pp. 169 184, Mar. 2003.

[9] Chong-Wah Ngo, Xiao Wu, Alexander G. Hauptmann “Practical elimination of near-

duplication from web video search", *ACM Multimedia,* pp. 218. 2007.
[10] Dugad R, Ratakonda K, Ahuja N. "Robust video shot change detection", *IEEE workshop on Multimedia Signal Processing, Redondo Beach, CA,* December 1998. p.376-81.
[11] Jing Huang, S. R. Kumar, M. Mitra, Wei-Jing Zhu, R. Zabih, "Image indexing using color correlograms," *IEEE Proc. Computer Vision and Pattern Recognition*, pp. 762-768, 1997.
[12] C. Kim, "Content-based image copy detection", signal processing: *Image Communication,* Vol 18. no.3 pp.169-184, 2003.
[13] X. S. Hua, X. Chen, and H. J. Zhang, "Robust video signature based on ordinal measure", *International conference on Image Processing,* 2004.
[14] C. Kim and B. Vasudev, "Spatiotemporal sequence matching for efficient video copy detection", *IEEE Trans. Circuit Systems Video Technology.* 15 (1) 2005, pp. 127-132

저 자 소 개

진 주 경(학생회원)
2003년 인하대학교 전자공학과 학사 졸업.
2005년 인하대학교 전자공학과 석사 졸업
2011년 인하대학교 전자공학과 박사 졸업
2011년 현재 인하대학교 포스트닥
<주관심분야 : 영상처리, 멀티미디어 신호처리, 패턴인식, 내용 기반 검색>

정 동 석(정회원)
1977년 서울대학교 전기공학과 학사 졸업
1985년 Virginia Tech 전자공학과 공학 석사
1988년 Virginia Tech 전자공학과 공학 박사
1988년～현재 인하대학교 전자공학과 교수
1990년～1994년 전자공학회 논문지 편집위원
1990년～1994년 통신학회 논문지 편집위원
2000년～2004년 정보전자공동연구소 소장
2010년～현재 인하대학교 IT공대학장
<주관심분야 : 영상처리, 컴퓨터 비젼, 패턴인식, 내용기반 멀티미디어검색>

나 상 일(학생회원)
2002년 인하대학교 전자공학과 학사 졸업
2004년 인하대학교 전자공학과 석사 졸업
2010년 인하대학교 전자공학과 박사 졸업
2008년～현재 ETRI 연구원
<주관심분야 : 영상처리, 컴퓨터 비전, 패턴인식, 내용 기반 검색>