스포츠 비디오를 위한 자막 위치검색 시스템

임정훈[0] 곽순영 국나영 이지현 이양원
군산대학교 컴퓨터과학과 인공지능연구실
dudrjqnr@kunsan.ac.kr kwagsoonyoung@hanmail.net lolajudy@hanmail.net jhlee@kunsan.ac.kr
ywrhee@cs.kunsan.ac.kr

Korea Information Science Society Caption position retrieval system for sports video

Joung-Hun Lim[0] Soon-Young Kwag Na-Young Ji-Hyun Lee Guk Yang-Won Lee
Dept. of Computer Infomation Sience, Kunsan National University

요 약

하이라이트를 구성하는데 종전에는 사람의 수작업에 의해서 이루어졌다. 요즘은 이런점을 연구를 통해 계속 자동화시키고 있는 추세이고 많은 논문들이 나오고 있다.

이 논문은 낮은 해상도의 동영상을 향상시키기 위해 Shannon Upsampling을 수행하고 적당한 임계치를 찾아내 이진영상을 만들어 전처리를 수행하고 수평 수직 히스토그램 기법과 다중프레임조합을 혼합해 자막위치를 찾는 방법을 제안한다. 이는 기존의 에지를 사용하는 방법들에 비해 간단하고 비교적 빠른 성능을 보인다.

1. 서 론

스포츠에 대한 사용자층이 다양해지고 있다. 더불어 스포츠 동영상의 대한 수요도 늘어나고 있는 실정이다. 이에 따라 동영상을 사용자의 구호에 맞춰 제작함으로써 보다 적은 시간에 많은 자료를 검색해서 사용자가 원하는 하이라이트만을 볼수 있는 있는 환경을 구성하고자 한다.

비디오 하일라이트 생성은 원래의 비디오 보다 짧고 의미있는 하일라이트 신을 생성하기를 원하는 멀티미디어 컨텐츠 제작자나 멀티미디어 사용자들에게 중요한 역할을 제공한다. 이에 대용량의 비디오를 하일라이트로 구성하는것은 중요한 역할을 수행한다.

이전까지는 수동으로 동영상을 편집하고 키워드를 삽입하고 제작자 관점으로 사람의 시각으로 구분해서 쿼리를 넣음으로써 객관적이지 못한 결과를 가져왔다. 또한 쏟아져 나오는 스포츠 동영상에 대해 사람의 손을 쓴다는 것은 거의 불가능한 실정이다. 이에 따라 특정한 조건에 따른 자동 편집에 대한 연구가 이루어져야 한다. 현제 광범위한 분야에서 내용과 이벤트(event)를 기반으로 하는 다양한 방법들이 연구되어지고 있다.

이 논문에서 제안하고자 하는 것은 자막내용 파악에 앞서 자막의 위치를 탐색해서 자막인지 여부와 어떤 종류의 자막인지를 파악하는데 있다. 방법은 우선 전처리를 통해 영상의 해상도를 높이게 된다. 해상도를 높이는데는 Shannon upsampling[3]기법을 사용했다. 그 다음은 영상처리 시간을 단축시키기위해 Othu[8] 메소드를 사용해 적절한 문턱치를 계산해서 이진영상을 만들어준다.
전처리를 한 다음 과정은 인접한 여러개의 프레임을 AND 연산을 통해 배경을 지우게 된다. 그럼 글자부분만 남게 되는데 이를 수평수직 히스토그램을 통해 위치를 파악하게 된다.

본 논문에서는 자막위치검색시스템을 설계하는데 있어서 3장에서 전체 시스템의 구성을 보이구 4장에서 전처리과정으로 Shannon Upsampling과 Othu Method를 보이고 5장에서 제안한 자막특성과 축구자막의 고유위치분석과 자막을 처리하는 방법을 기술하고 6, 7장에서 실험과정과 결론을 맺는다.

2. 관 련 연 구

전처리과정과 자막위치검색에 대한 다양한 연구가 이루어지고 있다. 이 중에서 대표적인 방법 몇가지를 소개한다.

[2]에서는 수평과 수직 각각의 색상변화 빈도수를 분석하여 자막 후보 영역을 선택하고 자막 색상의 단일성을 평가해서 자막영역을 찾는다.

Bernsen[4], Mardia[5], Niblack[6] 메소드에서는 국부적으로 적당한 문턱치 메소드를 사용했다. 지역적으로 계산된 문턱치에 의해서 영상이 변하게 된다. Niblack 메소드[6]에서는 영상을 여러 부분으로 나눠서 나눈 지역들에 m+ k*s를 계산해서 문턱치를 뽑아내게 된다. 여기서 m은 평균, k는 사용자 정의값, s는 표준편차를 의미한다.

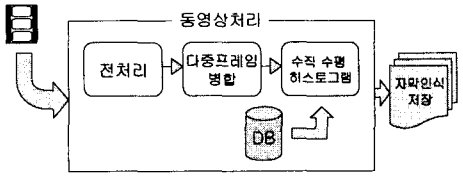
[7]에서는 칼라표지에 대해 자동 택스트 위치 식별한다. 칼라에 대한 민감한 변화를 감소하기 위해 클러스터링 기법을 사용해서 전처리를 수행한다. 하향식방법으로 이미지를 조각내고 상향식방법으로 region growing을 조합해서 텍스트 영역을 찾아낸다.

과거 연구의 문제점은 복잡한 연산을 수행해서 연산시간과 속도에 많은 단점을 보이고 있다. 본 논문은 이런 점을 해결하고 있다.

3. 전체 시스템 구성

[그림1]의 여러단계를 살펴보면 우선 동영상에서 프레임을 얻어내 처리를 한다. 전처리과정에서는 프레임의 영상을 향상시키기 위해 Upsampling을 수행한다. 얻어

진 영상은 다중프레임병합을 통해 AND연산을 하게되고 이를 통해서 배경을 지우게 된다. 결과적으로 자막영역만 남게 된다. 그 다음은 수직 수평 히스토그램의 xy사상을 통해 자막의 위치를 찾고 데이터베이스에 저장되어 있는 사전지식을 이용해서 매칭을 하고 확인된 자막의 종류와 위치를 저장하게 된다.



[그림1] 전체 시스템 구조도

4. 전처리

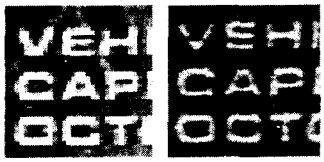
동영상 프레임을 우선 Shannon Upsampling으로 영상의 해상도를 높이게 된다. 이는 해상도가 낮은 영상은 자막을 분간하기가 쉽지 않기 때문이다. 그 다음은 Othu method를 통해서 자동으로 이진화된 영상을 만들게 된다. 이 과정은 처리시간단축에 중요한 역할을 수행한다.

4.1 Shannon Upsampling

영상을 캡쳐하는데 있어서 특별히 영상처리를 하지 않으면 대부분 낮은 해상도를 갖게 된다. 이를 해결하기 위해 Upsampling을 수행하게 된다. 이는 영상에 해상도를 높이는 방법으로 여러 방법중 Shannon Upsampling 방법을 통해서 이를 해결한다. 이 방법은 저역통과필터에 역변환과 비슷하다.

$$i(X,Y)=\sum_{x=1}^{N}\sum_{y=1}^{M} i(x,y)sinc(X-x,Y-y)$$

여기서 sinc함수는 FFT 와 행렬 마스크를 사용한다.



[그림2] (a) 블러링된 안티엘리아싱 이미지
(b) Shannon Upsampling 이미지

영상이 확대되었을 시에 [그림2](a)는 이웃화소 보간법에서 나타나는 계단현상을 보이게 된다. [그림2](b)는 주변화소사이의 그라디언트 효과가 발생하기 때문에 해상도가 높은 영상이 생성되게 된다.

4.2 Othu method

이 메소드는 히스토그램을 이용한 최적의 문턱치를 구하기 위한 알고리즘이다. 두 클래스(Peak) 사이에 가장 작은 포인트를 선택한다. T*은 문턱치 결과값을 나타내고 η의 최소값을 계산하게 된다. 확률적인 방법으로써 프레임의 수가 늘어남에 따라 보다 정확한 문턱치를 계산하게 된다.

$$T^{*}=\arg\min_{T\in D}\eta \quad , \quad D:\{f_t^{1},t=1,2,....\#frames\}$$

여기에서

$$\eta(t)=\frac{\sigma_b^{2}}{\sigma_t^{2}} \qquad \sigma_b^{2}=w_0w_1(\mu_0\mu_1)^{2}$$

$$w_0=\sum_{i=0}^{t}p_i \qquad w_1=1-w_0$$

$$\mu_0\approx\frac{\mu_t}{w_0} \qquad \mu_1=\frac{\mu_T-\mu_t}{1-w_0} \qquad \mu_t=\sum_{i=0}^{t}i\cdot p_i$$

5. 자막 위치 검색 시스템

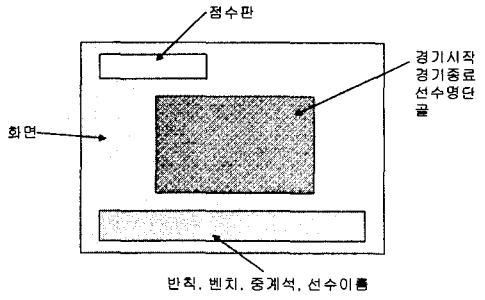
5.1 자막의 특성

현 국내 3개 방송의 스포츠 비디오 분석에 따르면 자막을 분석하는데 있어서 아래의 명확한 특성을 갖는다.

1. 각각의 자막들이 나타나는 위치는 고정되어 있다.
2. 각각의 자막들은 종류별로 일정한 크기를 갖는다.
3. 모든 자막들은 일정 시간동안 나타났다가 사라진다.
4. 자막 자체에서 내용이 변하는 위치는 고정되어 있다.
5. 자막을 형성하는 영역은 일정한 컬러값을 갖는다.
6. 자막은 어떠한 이벤트가 발생한 직후 나타난다.
7. 자막의 종류별 등장 순서는 매 방송마다 다를 수 있다.

5.2 자막의 위치 분석

[그림3]에 보는바와 같이 축구 비디오에서 자막은 10개로 분류할 수 있다. 각각은 고유한 위치를 가지고 있다.



[그림3] 축구 자막의 위치 분류

5.3 다중프레임 병합

영상은 계속 흘러가는데 자막은 일정시간 동안은 변하지 않는다. 이유는 동영상을 보는 사람으로 하여금 자막을 다 읽을 수 있는 시간적인 여유는 있어야 하기 때문이다. 물론 시간대는 불규칙적이다.

초당 30프레임의 재생을 하는 동영상이 있다. 이를 5프레임간격으로 여섯 개의 프레임을 덧셈연산하면 자막만 남고 배경영상은 지워지게 된다. 이유는 자막은 같은 위치에 일정한 시간동안 변하지 않지만 배경은 계속 변하기 때문에 AND연산을 하면 0의 값을 가지기 때문이다.

[그림4]는 동영상 스트림에서 일정한 간격으로 뽑아내 병합하는 과정을 보인다.

[그림4] 동영상 스트리밍과 병합

5.4 자막 위치 검색

수직 수평 히스토그램 기법을 이용한다.
자막의 x값을 알기 위해서는 y축 사상을 통해 수평 히스토그램이 적용된고 y위치값을 알기 위해서는 x축 사상을 통해 수직 히스토그램이 적용되게 된다.
[그림5]는 한 프레임에 대한 수직 수평 각각의 히스토그램을 보인다.

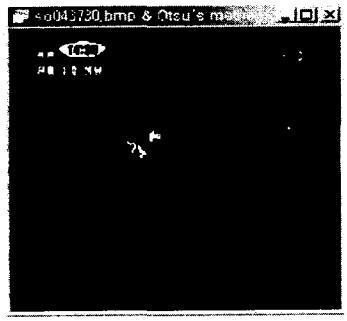
[그림5] 다중프레임 병합전 수직 수평 히스토그램 결과

6. 구현

본 눈문의 실험은 펜티엄 Ⅳ 1.9Hz PC와 윈도우 2000환경에 Visual C++ 6.0 언어로 구현하였다.

비디오 자료는 2001.4월 펼쳐진 제3회 문화관광부 장관 배 고교축구 4 경기의 전반전을 대상으로 AVI 압축형태의 비디오를 OSCAR Ⅱ 캡쳐 보드로 초당 5프레임을 실험 데이터로 캡쳐하여, 프레임 크기를 400X300으로 정규화 하여 사용하였다.

다중프레임 병합에서 프레임은 많은 프레임이 병합될수록 오류범위는 줄어들게 된다. 조건은 자막이 달라지거나 없어지는 시점을 넘어서면 안된다는 것이다. 이 논문에서는 경험적인 방법을 적용해서 적절한 6프레임 병합을 하게 되었다. 오차범위는 10%로 자막이 아닌 화면 영상의 변화가 심한 영역일수록 오차는 줄어든다.



[그림6] 다중프레임 병합후 결과 영상

[그림6]에서는 중앙에 잘못인식한 자막영역으로 보이고 있다.

[표1] 프레임당 평균계산시간 비교

	평균 계산 시간
top-down & bottom-up region growing	3.4초
제안된 다중프레임 병합	2.2초

[표1]은 상향식과 하향식을 혼합한 region growing 방법과 제안된 다중프레임방법에 따른 평균계산시간을 비교하였다. 제안된 방법이 두배 정도의 빠른 계산를 보이고 있다.

7. 결론

본 논문에서는 축구 비디오 데이터에서 자막의 위치를 검색해서 하이라이트를 생성하는 방법에 대해 제안하였는데, 영상향상과 적절한 문턱치 계산기법을 통한 전처리를 수행했고 자막특성을 이용한 자막의 위치를 찾아냈으며 다중프레임기법을 통한 간단한 알고리즘을 보였다.

자막 위치 검색을 하는데 있어서 다중프레임기법을 사용해서 여러 프레임에 AND연산을 하는데는 많은 시간을 요구하게 되었다. 이러한 문제를 해결하기 위해 배경을 지우는데 적은 시간을 요구하는 알고리즘에 대한 연구가 계속 수행되어야 할 것 같다.

8. 참 고 문 헌

[1] 류근호, 전근환, 신성윤, 이양원, "멀티미디어 : 축구 비디오 하이라이트 생성", 정보처리학회논문지, Vol.8, No.4, 2001
[2] 임문철, 김우생, "비디오에서 색상변화 빈도수를 이용한 자막영역 추출기법", 춘계학술발표 논문집, Vol.2, No.1, 1999
[3] Huiping Li, Omid Kia and David Doermann, "Text Enhancement in Digital Video", Proceedings of CIKM-99, 8th ACM International Conference on Information and Knowledge Management, 1999
[4] Bernsen J. "Dynamic thresholding of grey-level images.", In Proceedings of ICPR, pages 1251-1255, 1986.
[5] K.V.Mardia and T.J. Hainsworth, "A spatial thresholding method for image segmentation", In PAMI(10), No 6,, page 919-927, November 1988.
[6] Niblack W, "In An introduction to image processing", Englewood Cliffs, N.J.:Prentice Hall, pages 115-116, 1986.
[7] K. Sobottka, H. Bunke, and H. Kronenberg, "Identification of text on colored book and journal covers.", In Proceedings of the 5. Int. Conference on Document Analysis and Recognition, pages 57-62, September 1999.
[8] Bilge Gunsel and A. Murat Tekalp, "Content-based video abstraction," in Proceedings of IEEE Int'l Conference on Image Processing, Chicago, IL, October 4-7 1998.