

스포츠 하이라이트 생성

김미호*, 신성윤**, 전근환**, 이양원*
*군산대학교 컴퓨터정보과학과
**군장대학 컴퓨터응용학부
e-mail:syshin@cs.kunsan.ac.kr

Sports Highlight Abstraction

Mi-Ho Kim*, Seong-Yoon Shin**, Keun-Hwan Jeon**,
Yang-Weon Rhee*

*Dept of Computer Science, Kunsan National University

**Division of Computer Application, Kunjang College

요 약

다양한 장르의 비디오 데이터에서 하이라이트를 생성하는 것은 짧게 요약된 하이라이트 비디오 신을 생성하고자 하는 멀티미디어 콘텐츠의 제작자나 사용자에게 중요한 역할을 제공한다. 본 논문에서는 새로운 비디오 하이라이트의 생성 방법과 내용 기반, 즉 이벤트 기반의 비디오 인덱싱 방법을 제시한다. 경기 종목으로는 골을 넣어 득점하는 축구, 농구 그리고 핸드볼을 대상으로 하였으며 골을 넣어 득점하는 하이라이트 샷을 추출하기 위해 이벤트 규칙을 사용하였다. 비디오 인덱싱에서는 비디오 데이터 자체의 시각적 정보와 캡션 정보를 모두 이용하였다.

1. 서론

비디오 하이라이트(highlight)의 생성은 콘텐츠(contents) 제작자나 사용자 관점에서 볼 때 매우 중요한 역할을 수행하는 것으로서, 매우 긴 비디오 데이터에서 중요한 내용을 추출하여 보다 짧은 비디오 신(scene)을 생성하는 것을 말하여 내용에 기반한 다양한 방법들이 연구되어지고 있다.

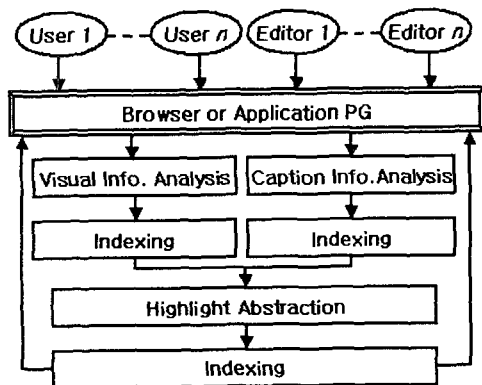
비디오 데이터는 시각적, 청각적, 그리고 문자 스트림(stream)들과 같은 다양한 정보들이 동기화 되어 사용자들에게 제공되는 것으로서 멀티미디어 시스템에서 다양한 형태와 시간적 차원을 갖는 핵심적인 요소이다[1]. 이러한 비디오 데이터로부터 사용자들은 필요한 내용을 보다 빠르고 편리하게 검색하여 재생 및 편집 할 수 있는 방법들을 요구하고 있으며 보다 효율적인 방법들이 연구 개발되어야 할 것이다.

비디오 하이라이트를 추출하기 위하여 비디오를 시간의 흐름에 따른 중요한 정보를 포함하는 샷(shot)으로 분류하는 작업이 우선적으로 수행되어야

한다[2,3,4]. 이렇게 생성된 샷들은 제작자나 사용자들이 쉽게 검색하여 편집하고 재생할 수 있도록 논리적으로 인덱싱(indexing)되어 브라우징(browsing)이 가능하도록 구축해야 한다.

본 논문에서는 스포츠(sports) 비디오에서 하이라이트를 자동으로 생성하는 새로운 방법을 제시하는데, 실험 대상으로 골을 넣어 승부를 가르는 구기 종목인 축구, 농구, 핸드볼을 선정하였다. 이러한 구기 종목들은 실제로 거의 비슷하게 반복되는 장면들이 대부분을 차지하기 때문에 하이라이트를 추출하는데 상당한 어려움을 준다. 그러므로, 시각적인 정보와 자막의 텍스트 정보를 함께 이용하여 골(goal)이라는 이벤트를 기반으로 하여 하이라이트를 추출하도록 한다.

전체적인 시스템 구조는 (그림 1)과 같으며 2장에서는 시각정보에 의한 비디오 인덱싱, 3장에서는 자막정보에 의한 비디오 인덱싱에 대해 설명하고 4장에서는 비디오 하이라이트 추출에 대해서, 5장에서는 실험 및 고찰을 수행하고, 6장에서 결론 및 향후 연구 과제를 제시하도록 한다.



(그림 1) 시스템 구성

2. 시각정보에 의한 비디오 인덱싱

시각정보에 의한 인덱싱을 위하여 먼저 [5,6]에서 제안한 컬러 히스토그램(color histogram)과 χ^2 히스토그램의 장점을 합성한 (식 1)의 방법을 이용하는데, 각 프레임(frame) 사이의 차이 값을 계산하고 차이 값의 크기에 따라 키 프레임을 추출하여 샷(shot)들을 구분하는 계산 방법이다.

$$d(I_i, I_j) = \sum_{k=1}^n \left(\frac{(H_i^r(k) - H_j^r(k))^2}{H_i^r(k)} \times 0.299 + \frac{(H_i^g(k) - H_j^g(k))^2}{H_i^g(k)} \times 0.587 + \frac{(H_i^b(k) - H_j^b(k))^2}{H_i^b(k)} \times 0.114 \right) / 3 \dots \dots \text{(식 1)}$$

이렇게 하여 얻어진 히스토그램의 차이 값에 따라 차이 값 변화가 큰 부분의 전환점이 되는 프레임을 찾아 이 프레임을 키 프레임(key frame)으로 설정하도록 한다. 스포츠 비디오에서 키 프레임은 첫 번째 프레임을 포함하여 추출된 각 샷들의 첫 번째 프레임이 된다.

키 프레임들은 추출된 순서에 따라 장면 변화가 큰 프레임을 중심으로 자동적으로 일련번호에 의해 물리적으로 인덱싱 되며, 이들에 대하여 분야별 또는 사건별로 의미를 부여하여 논리적인 구조로 인덱싱하여 접근하는 것도 가능하다.

논리적인 인덱싱은 사용자가 원하는 다양한 형태

로 구성이 가능한데, 물리적으로 인덱싱된 키 프레임 번호를 연결하여 스포츠 비디오의 어떤 장면이라도 직접 접근이 가능하도록 한다.

하나의 스포츠 비디오 스트림을 V 라 하면 (식 2)와 같이 V 는 각 샷들의 키 프레임 SK_i 들로 인덱싱되어 구성되며 이 SK_i 는 샷을 구성하는 각각의 프레임 SF_{ij} 들로 구성된다. 여기서 샷은 (식 1)에서 제안한 방법에 따라 장면변화가 큰 이벤트에 대한 경계점을 나타내며, 프레임은 이 샷들을 구성하는 최소 단위를 말한다.

$$V = \sum_{i=1}^m SK_i$$

$$SK_i = \sum_{j=1}^n SF_{ij} \dots \dots \dots \text{(식 2)}$$

3. 자막정보에 의한 비디오 인덱싱

비디오 데이터에서 자막은 시청자에게 중요한 내용을 전달하는 대표적인 수단이다. 스포츠비디오에서 자막은 거의 일정한 위치에 존재하게 되며 이벤트(event)가 발생할 경우 화면에 표시되어 내용을 전달한다. 보통 구기종목 스포츠비디오에서 자막은 다음 <표 1>과 같이 분류한다.

<표 1> 자막의 분류

점수	팀 + 점수 + 시간
골	점수 + 득점선수 + 시간
선수교체	투입선수 + 아웃선수
반칙	반칙(파울, 경고, 퇴장)표시 + 반칙선수
벤치	감독 + 코치 + 대기선수
시작/종료	팀 + 점수
선수명단	각 팀별 선수 명단
방송인	아나운서 + 해설자

이와 같은 자막 정보가 등장하는 첫 프레임을 자막정보에 의한 샷의 키 프레임으로 설정하여 물리적인 인덱싱을 수행하는데, 이는 스포츠비디오의 전체적인 하이라이트를 추출하기 위하여 골 이벤트에 따라 다시 인덱싱 된다.

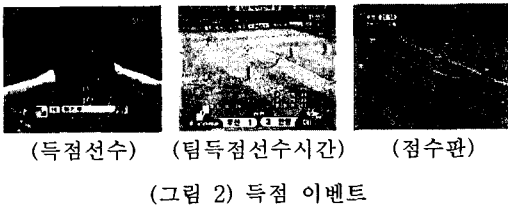
본 논문에서 자막정보에 의한 인덱싱은 골 이벤트를 기준으로 수행하여 하이라이트를 생성하도록

하므로 <표 2>와 같은 특징을 갖는 골 이벤트와 관련된 샷들만 다시 추출하도록 한다.

<표 2> 골 이벤트 자막의 특징

우선순위	특징
1	점수판 자막의 점수변화
2	선수이름
3	팀 + 점수 + 선수이름들 + 득점시간들

골과 관련된 자막 정보는 같은 구기종목이라도 점수판 자막의 점수변화를 제외하고는 서로 다르다. 축구의 골 득점의 경우 (그림 2)와 같이 자막의 크기나 위치도 약간씩 서로 다르기 때문에 키 프레임과 자막의 정확한 위치를 추출하도록 한다.



특히 점수판 자막의 경우 점수 변화의 인식은 자막의 문자인식 과정을 거치지 않고 점수 영역의 히스토그램을 이용하여 인식하도록 한다.

4. 하이라이트 추출 및 인덱싱

스포츠 비디오의 전체적인 하이라이트는 앞에서 추출한 시각적 정보에 의하여 인덱싱한 데이터와 자막정보에 의하여 인덱싱한 데이터를 조합하여 이들 중 필요한 샷들을 추출하여 구성한다.

스포츠 비디오 하이라이트는 <표 3>과 같은 생성 규칙에 의하여 구성될 수 있으며, (그림 3)과 같이 추출된 각 샷들의 키 프레임을 중심으로 전과 후에 일정한 수의 프레임을 추가하여 새로운 하이라이트 샷을 생성한다.

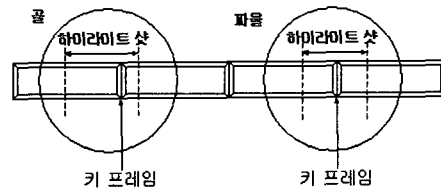
시각적인 샷(VSi)에 해당하는 샷들은 가끔 수동으로 다른 장면을 하이라이트에 삽입 할 경우 임의의 샷을 선택하여 자막 샷에 삽입하여 편집이 가능하다. 예로서 스포츠 하이라이트 중계 도중 관중의 환호와 같은 샷들이 이에 해당된다고 볼 수 있다.

이렇게 생성된 하이라이트를 구성하는 샷들은 시간

의 흐름에 따라 순서대로 인덱싱 되며, 실제로 약 100여분 정도의 축구 경기 하이라이트는 약 7분 내지 8분 정도의 요약된 하이라이트로 볼 수 있다.

<표 3> 하이라이트 생성 규칙

구분	구성
하이라이트	$\exists(VSi)$ $\exists(CSi)$
VS _i (시각적 샷)	연속 경기 샷 선수 초점 샷 반칙 샷 선수 교체 샷 골 샷 관중 샷 중계 샷 벤치 샷
CS _i (자막 샷)	점수 표시 샷 선수 표시 샷 선수 교체 샷 벤치 소개 샷 심판 소개 샷 아나운서 및 해설자 샷



(그림 3) 하이라이트 구성 샷 설정

5. 실험 및 고찰

실험 데이터는 축구, 농구 그리고 핸드볼 각각 5경기를 대상으로 수행하였다. 이 데이터는 실제 경기가 시작되는 시간부터 종료되는 시간까지의 데이터이며 이를 이용하여 측정한 하이라이트 생성 결과는 <표 4>와 같다.

<표 4>의 실험 결과에서 알 수 있는 것처럼 골과 반칙 및 작전 타임이 다른 경기보다 상대적으로 작으며 운동장에서의 패스 및 공 점유가 상당히 많은 부분

을 차지하는 축구 경기에서는 하이라이트 생성율이 상대적으로 높았으며, 골과 반칙 및 작전 시간이 많은 농구와 핸드볼에서는 상대적으로 하이라이트 생성율이 낮음을 알 수 있다.

<표 4> 평균 하이라이트 생성 결과

종목	평균 경기시간	평균 하이라이트 시간	요약율 (%)
축구	101분 03초	7분 36초	92.5
농구	46분 21초	6분 32초	85.9
핸드볼	52분 12초	7분 08초	84.4

또한 같은 종목의 경기라 할지라도 매 경기 경기에 따라서 어느 경기는 득점이 많지 않은 경기가 있는가 하면 또 어떤 경기에서는 상대적으로 득점이 다른 경기의 2배정도 및 그 이상이 있는 경우도 있기 때문에 실제 하이라이트 비디오의 생성율에 많은 영향을 미친다는 것을 알 수 있었다.

실험에서 발생하는 오류는 대부분 일정 위치에 존재하는 자막 위치의 아주 작은 변화나 갑자기 다른 부분에 초점을 맞추는 등의 방송 촬영의 방식에 의하여 많이 발생하지만 대부분의 데이터에서 원활하게 하이라이트를 추출하는 것을 알 수 있었다.

6. 결론 및 향후 연구 방향

본 논문에서는 스포츠 비디오 데이터에서 중요한 내용만을 포함하는 하이라이트를 생성하는 새로운 방법과 인덱싱에 대해 제안하였는데, 하이라이트 생성은 시각적인 정보와 자막에 의한 정보를 바탕으로 이벤트를 중심으로 하여 하이라이트 생성 규칙에 따라 생성하였다.

거의 비슷하게 반복되는 부분이 많은 긴 스포츠 비디오를 중요한 부분만을 추출하여 한눈에 시청 할 수 있도록 하는 새로운 방법으로서 기대효과가 크지만, 실제 비디오상의 시각적인 변화와 자막의 변화에 잘 적응하는 효율적인 요약 방법에 대한 연구가 계속 수행되어야 할 것 같다.

참고문헌

- [1] P.Aigrain, H. J. Zhang and D. Petkovic, "Content-based representation and Retrieval of Visual Media : a State-of-the-Art Review, *Multimedia Tools and Applications*, Vol.3, pp. 179-202, 1996.
- [2] M. A. Smith and T. Kanade, "Video Skimming and Characterization tyrough the Combination of Image and Language Understanding Techniques", *Proceedings of CVPR '97*, pp. 775-781, 1997.
- [3] R. Lienhart, S. Pfeiffer and W. Effelsberg, "Video Abstracting", *ACM Communication*, Vol. 40, No. 12, pp. 55-62, 1997.
- [4] L. He, E. Sanocki, A. Gupta and J. Grudin, "Auto-Summarization of Audio -Video Presentations", *Proceedings of ACM Multimedia'99*, pp. 489-498, 1999.
- [5] 이동섭, 이양원, "뉴스 비디오의 장면분할을 위한 키 프레임 추출 기법", *한국정보과학회 춘계 학술발표 논문집*, Vol. 25, No. 1, pp. 613-615, 1998.
- [6] 양원보, 김재원, 신성윤, 전근환, 배석찬, 이양원, "뉴스 비디오에서 앵커 인식", *제12회 산학연 멀티미디어 기술 학술대회 학술논문집*, pp. 296-301, 1998.11