

캡션 분석에 의한 축구비디오 하이라이트

전근환*, 하태준**, 신성윤**, 이양원**, 류근호*

*충북대학교 전자계산학과

**군산대학교 컴퓨터정보과학과

e-mail:khjeon@kunjang.ac.kr

The Highlight of The Soccer Video Using Caption Analysis

Keun-Hwan Jeon, Tae-Joon Ha, Seong-Yoon Shin,
Yang-Won Rhee and Keun-Ho Ryu

*Dept. of Computer Science, Chungbuk Natl. Univ.

**Dept. of Computer Information Science, Kunsan Natl. Univ.

요약

비디오 데이터에서 캡션은 비디오의 중요한 부분과 내용을 나타내는 가장 보편적인 방법이다. 본 논문에서는 축구 비디오에서 캡션이 갖는 특징을 분석하고 캡션에 의한 키 프레임 추출하도록 하며, 하이라이트 생성 규칙에 따라 하이라이트를 생성하도록 한다. 키 프레임 추출은 이벤트 발생에 따른 캡션의 등장과 캡션 내용의 변화를 추출하는 것으로 템플릿 매칭과 지역적 차영상을 통하여 추출하며 샷을 재설정 하여 중요 이벤트를 포함한 하이라이트를 생성하도록 한다.

1. 서론

비디오 하이라이트(Highlight) 생성은 장시간의 대용량 비디오 데이터에서 중요한 내용들만을 추출하여 보다 함축적이면서 짧은 비디오 신(scene)을 생성하는 것을 말한다.

비디오 하이라이트를 생성하기 위하여 먼저 비디오를 샷(shot)으로 분류하는 작업[1,2,3]이 수행되어야 하는데, 이 샷들은 대표 프레임(frame)[4] 또는 키 프레임(key frame)[5,6]이라는 중요한 프레임을 갖게 되며 색인화를 위해 이용된다.

하이라이트 생성과 관련된 연구를 살펴보면, 먼저 비디오 하이라이트를 생성하는데 이미지 템플릿(template), 통계학적 특징, 그리고 히스토그램(histogram) 기반 검색과 처리를 이용하는 방법이 있다[7].

또한, 시각/청각적인 특징을 모두 고려하여 원래 비디오의 시놉시스(synopsis)를 표현하는 비디오 스킴(skim)을 구축하는 방법이 있는데, 이 또한 원래 비디오의 세그먼트(segment)들을 병합하는 방법을 이용한 것이다[8].

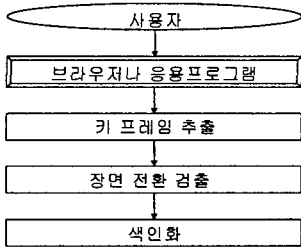
[9]에서는 스토리(story) 내용을 묘사하는데 비디오 포스터(video poster)를 제안하였으며, [10]에서는 비디오 내의 신을 구분하기 위해서 두드러진 모션(motion)이나 다양한 히스토그램 특징들을 분석하여 이용하였다.

또한 [11,12]에서는 고정된 환경의 고정된 도메인(domain) 레도 내에서 움직이는 객체를 추출하고 인식하여 그들의 모션을 분류하는 방법을 이용하였으며, [13]에서는 객체의 활동 레도를 시뮬레이션(simulation)하기 위하여 시공간적 통합 데이터세트(dataset)를 생성하여 이용하였다.

본 논문에서는 축구 경기 비디오 하이라이트를 생성하는 새로운 방법을 제시하는데, 캡션을 이용하여 중요 이벤트(event)를 중심으로 하이라이트 생성 규칙에 따라 하이라이트 비디오를 생성한다. 캡션정보에 대하여 템플릿 매칭과 지역적 차영상에 의하여 캡션 키 프레임을 추출하고 색인화 한다. 색인화된 비디오 데이터는 비디오 하이라이트 규칙에 따라 하이라이트 비디오로 생성된다.

전체적인 시스템 구조는 (그림 1)과 같으며 2장에

서는 캡션의 분석, 3장에서는 캡션 키 프레임 추출과 색인화, 그리고 4장에서는 비디오 하이라이트 생성에 대해서 설명한다. 5장에서는 실험과 결과를 분석하고 6장에서 결론을 맺도록 한다.



(그림 1) 시스템 구조

2. 캡션의 분석

본 논문에서는 실험 대상인 축구 경기 비디오에서 캡션을 위의 <표 1>와 같이 10가지 종류로 한정하여 분류하였는데, 캡션의 분석에 의해 캡션들은 다음과 같은 명확한 특성을 갖는다.

- (1) 캡션들의 위치는 일정함.
- (2) 캡션들 종류에 따라 크기는 일정함.
- (3) 일정 시간동안 나타났다가 사라짐.
- (4) 캡션 자체에서 내용이 변하는 위치는 고정.
- (5) 캡션 형성 영역은 거의 일정한 컬러값.
- (6) 점수판 캡션의 간헐적 등장을 제외하고 모든 캡션은 어떠한 이벤트가 발생한 직후 등장.
- (7) 캡션 등장 순서는 매 방송마다 다를 수 있음.

<표 1> 캡션의 분류

구분	종류	특징	설명
CS _i (Caption Shot)	C _{sco}	점수판 (score board)	팀과 점수 및 시간 표시
	C _{gol}	골 (goal)	점수와 선수 및 시간 표시
	C _{chg}	선수교체 (player change)	교체되는 양 선수 표시
	C _{ft}	반칙 (fault)	팀과 선수 반칙 표시
	C _{bch}	벤치 (bench)	감독, 코치나 대기선수 표시
	C _{bgn}	경기시작 (game begin)	양 팀 표시
	C _{end}	경기종료 (game end)	양 팀과 점수 표시
	C _{lst}	선수명단 (player list)	팀과 선수 리스트 표시
	C _{ct}	중계석 (caster)	아나운서와 해설자 표시
	C _{plr}	선수이름 (player name)	선수 표시

3. 캡션 키 프레임 추출과 색인화

(1) 키 프레임 추출

키 프레임 추출은 캡션이 갖는 특성을 바탕으로 템플릿 매칭을 이용하여 캡션의 등장을 인식하고 지역적 차영상을 이용하여 캡션의 내용변화를 인식하여 추출하는데, 다음과 같은 경우에 키 프레임으로 설정한다.

1) 캡션이 나타난 첫 번째 프레임

캡션의 등장에 대한 인식은 먼저 캡션 영역에 대하여 최소사각영역을 설정한 후, 캡션의 위치와 크기 및 컬러 정보를 바탕으로 식 (1)의 방법을 이용하여 추출한다.

$$Similarity = CI_i(p, s, c) - MI_j(p, s, c),$$

$$\text{where } i=1 \dots m, j=1 \dots n \quad (1)$$

여기서 $MI_j(p, s, c)$ 는 캡션의 위치(p :position)와 크기(s :size) 그리고 컬러(c :color)값을 갖는 템플릿을 나타내며, $CI_i(p, s, c)$ 는 입력되는 캡션 영역을 나타낸다.

2) 존재하는 캡션 내용이 변한 첫 번째 프레임

캡션이 나타난 다음에 캡션 내부에서의 내용 변화를 추출하기 위한 방법으로서, 캡션의 유동영역을 설정하여 식 (2)와 같은 지역적 차영상을 이용하여 캡션 내용의 변화를 추출하고 내용이 변한 첫 번째 프레임을 키 프레임으로 설정한다.

$$LD(x, y) = \sum_{y=1}^N \sum_{x=1}^N |I_a(x, y) - I_b(x, y)| \quad (2)$$

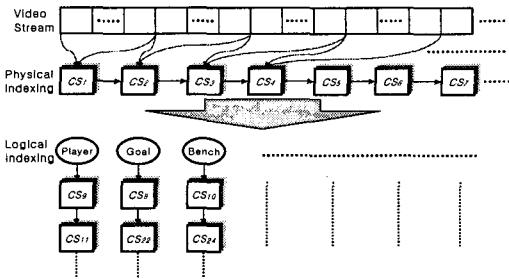
(2) 캡션에 의한 키 프레임 색인화

색인화는 식 (3)과 같이 하나의 비디오 스트림을 V 라 하면, V 는 캡션 샷들의 키 프레임인 CS_i 들로 구성되며, 이 CS_i 는 샷을 구성하는 각각의 프레임 CF_j 들로 구성된다.

$$V = \sum_{i=1}^n CS_i, \quad CS_i = \sum_{j=1}^m CF_j \quad (3)$$

논리적 색인화는 키 프레임들을 여러 형태로 연결

하여 해당 캡션 샷에 접근 가능한데, 캡션 샷의 물리적/논리적 색인화의 구성형태는 (그림 2)와 같다.



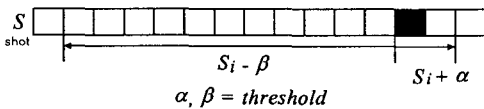
(그림 2) 캡션 샷의 물리적/논리적 색인화 구조

4. 비디오 하이라이트 생성

(1) 샷 설정

샷 설정은 키 프레임을 중심으로 캡션 샷의 크기를 일정하게 설정해 주는 것이다.

캡션은 이벤트 발생과 동반하여 등장하게 되므로 샷 설정은 (그림 3)과 같이 키 프레임을 기준으로 이벤트 영역(β)과 경기 재개 영역(α)을 합하여 하나의 샷으로 설정한다.



(그림 3) 키 프레임에 의한 샷 영역

(2) 비디오 하이라이트 생성

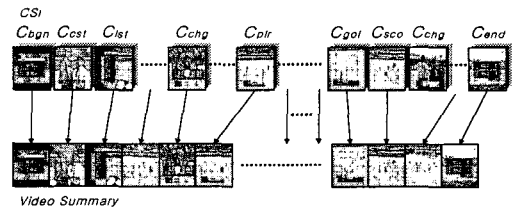
비디오 하이라이트는 CS_i 를 중심으로 기본 20개의 샷을 하나의 하이라이트로 설정하여 구성한다. 하이라이트 생성규칙은 다음과 같으며, (그림 4)는 하이라이트의 한 예를 나타낸다.

1) 기본적 하이라이트 구성 샷은 $C_{sco}, C_{gol}, C_{bgn}, C_{end}, C_{ist}, C_{bet}$ 의 6개이다.

- ① 첫 번째 샷은 C_{bgn} 과 C_{bet} 둘 중 하나이며, 둘 중에서 하나가 선택되면 다른 하나는 바로 뒤에 따른다.
- ② 위의 ①이 결정된 다음에는 C_{ist} 샷이 따른다.
- ③ 골이 있어 득점을 한 경우에는 C_{gol} 과 C_{sco} 샷이 골 득점 수에 맞게 추가된다.
- ④ 마지막 샷은 C_{end} 와 C_{bet} 둘 중 하나이며, 둘 중에 하나가 선택되면 다른 하나는 바로 앞에 추가된다.
- ⑤ 경우에 따라 C_{bet} 는 나타나지 않을 수 있다.

2) 캡션 샷 $C_{chg}, C_{flu}, C_{plr}$ 그리고 C_{bch} 를 선택적으로 사용자가 추가하는 것을 원칙으로 한다.

3) 시간의 흐름과 무관하게 추가될 샷들을 선택할 수도 있다.



(그림 4) 비디오 하이라이트 생성의 예

5. 실험 및 분석

펜티엄III 400MHz, Win98, Visual C++ 6.0 환경에서 구현하였다. 문화관광부 장관 배 고교 축구 4경기 전반전 AVI 비디오를 OSCAR II로 초당 5프레임으로 캡처 후 크기를 400X300으로 정규화 하여 사용하였다.

키 프레임 추출 결과는 <표 2>와 같은데, 캡션의 정확한 특성 때문에 실제 오류는 발생하지 않았으며, 단, 방송에 따라 모양, 형태, 위치가 다른 경우에는 이들을 다시 설정해 두어야 하는 문제점이 존재한다.

평균 재현 시간은 <표 3>과 같다.

<표 2> 캡션 키 프레임 추출 결과

<표 3> 평균 재현시간(초)

구분	캡션에 의한 추출 수	구분	실제 데이터	하이라이트 비디오	생성율
A경기	53	평균 재현 시간	2852	327	11%
B경기	47				
C경기	57				
D경기	45				

여기서 생성율은 함축된 정도를 나타내는 것이며, 비디오에서 갑작스런 캡션 위치, 크기 그리고 컬러 정보가 변하지 않는다면 100% 정확히 추출이 가능하다.

6. 결론

본 논문에서는 축구 비디오 데이터에서 중요한 이벤트 발생에 따라 등장하는 캡션에 의한 정보를 바탕으로 하이라이트 비디오를 하이라이트 생성 규칙에 따라 생성하였다.

장면전환검출 또한 캡션 특성을 바탕으로 문자 인식

과정을 거치지 않고 템플릿 매칭과 지역적 차영상을 이용한 유사성 측정으로 정확하게 캡션의 키 프레임을 추출하였다.

참고문헌

- [1] 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.
- [2] R. Lienhart, S. Pfeiffer and W. Effelsberg, "Video Abstracting", *ACM Comm*, Vol.40, No.12, pp.55-62, 1997.
- [3] L.He, E.Sanocki, A.Gupta and J.Grudin, "Auto-Summarization of Audio -Video Presentations", *Proc of ACM Multimedia'99*, pp. 489-498, 1999.
- [4] Sun, X., Kankanhalli, M., Zhu, Y. & Wu, J., "Cont-ent-Based Representative Frame Extraction for Digital Video", *Int. Conf. on Multimedia Computing and Systems*, pp. 190-193, 1998.
- [5] Smith, M.A. & Kanade, T., "Video Skimming for Quick Browsing based on Audio and Image Chara -cterization", *T.R. No. CMU-CS-95-186*, School of Computer Science, Carnegie Mellon Univ., 1995
- [6] Zhang, H.J., Low, C.Y. & Smoliar, S.W., "Video Parsing and Browsing using Compressed Data", *Multimedia tools and App. 1*, pp. 89-111, 1995
- [7] W. Chang, G. Sheikholeslami, J. Wang and A. Zhang, "Data Resource Selection in Distributed Visual Information Systems", *IEEE Trans. on Knowledge and Data Engineering*, Vol. 10, No. 6, pp. 926-946, 1998
- [8] M. Smith and T. Janade, "Video Skimming for Quick Browsing based on Audio and Image Characterization", *Tech. Report CMU-CS-95-186*, Computer Science Department, Carneigie Mellon University, July 1995
- [9] M. Yeung and Boon-Lock Yeo, "Video Visualization for Computer Presentation and Fast Browsing of Pictorial Content", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 7, No. 5, pp. 771-785, 1997
- [10] N. Vasconcelos and A. Lippman, "A Spatiotemporal Motion Model for Video Summarization", *CVPR*, Santa Barbara, 1998
- [11] G. Medioni, R. Nevatia and I. Cohen, "Event detection and Analysis from Video Streams", *DARPA98*, pp. 63-72, 1998
- [12] R. Rosales and S. Sclaroff, "3D Trajectory for Tracking Multiple Objects and Trajectory Guided Recognition of Actions", *CVPR*, June 1999
- [13] D.Pfoser and Y.Theodoridis, "Generating Semantics-Based Trajectories of Moving Objects", *Int. Workshop on Emerging Technologies for Geo-Based Applications*, Ascona, Switzerland, 2000