

장면자막의 구조적 특징을 이용한 축구 비디오 하이라이트 생성

신성윤*, 이양원*

*군산대학교 컴퓨터정보과학과
e-mail:syshin@cs.kunsan.ac.kr

Creation of Soccer Video Highlight Using The Structural Features of Caption

Seong-Yoon Shin*, Yang-Won-Rhee*

*Dept of Computer Info. Science, Kunsan Natl. University

요 약

논문에서는 자막의 구조적 특징을 이용하여 축구 비디오 하이라이트를 생성하는 방법을 제시한다. 자막의 구조적 특징은 자막이 갖는 시간적 특징과 공간적 특징으로서 이러한 구조적 특징을 이용하여 자막 프레임 구간과 자막 키 프레임을 추출한다. 그리고 하이라이트 비디오는 자막 키 프레임들에 대한 장면 재설정과 논리적 색인화 및 하이라이트 생성 규칙을 이용하여 생성한다. 마지막으로, 브라우저를 통한 사용자의 항목 선택에 의하여 하이라이트 비디오와 비디오 세그먼트에 대한 검색과 브라우징을 수행할 수 있다.

1. 서론

비디오 데이터베이스에서 비디오 내용은 샷이나 신의 구조로 묘사된다[1]. 비디오 색인화는 자막 텍스트의 추출과 인식을 사용하여 효율적으로 수행될 수 있다. 예를 들면, 광학 문자 인식(OCR) 기술[2,3,4,5]을 이용한 하드카피 다큐먼트의 자동 변환과 음성 인식(VR) 기술[6,7]을 이용한 언어의 모사에서 놀라울만한 성공을 거두었다. 두 가지 경우에서 완벽하진 않지만 전통적인 정보 검색 기술로 색인화될 수 있는 ASCII 텍스트 형태의 출력을 생성한다.

비디오 요약은 비디오를 시청하고자 하는 사람에게 비디오를 시청할 가치가 있는지를 결정하는데 있어서 유용하다. 비디오 요약 시퀀스는 전체 비디오의 내용에 대하여 중요한 개요를 제공하기 때문에 다큐멘터리에 적합한 반면, 비디오 하이라이트는 관심 분야의 비디오 세그먼트들만을 포함하기 때문에 비디오 예고편이나 스포츠 하이라이트에 알맞다[8].

비디오 요약 시퀀스 생성 방법에는 비디오 스키밍[9], 장면 변화 그래프[10], 클러스터 유효성 분석[11], 그리고 비디오 망가[12]가 있다. 그리고 하이라이트 생성 방법에는 무비 트레일러[13,14]와 이벤트-기반 스포츠 요약[15]이 있다.

Christal 등[9]은 다큐멘터리나 뉴스 방송을 요약

하기 위하여 비디오 스키밍을 제시하였다. Yeung 등[10]은 장면 변화 그래프라는 스토리 흐름을 이용한 샷 기반 구조를 제시하였다. Hanjalic 등[11]은 키 프레임을 추출하고 비디오 샷을 설정하였다. 그리고 나서 그 키 프레임들을 포함하는 비디오 요약 시퀀스를 생성하기 위하여 클러스터 유효성 분석을 사용하였다. Uchihashi 등[12]은 비디오 망가라는 코믹 만화 형태를 닮은 비디오 요약 생성 방법을 제시하였다. 이들은 비디오 세그먼트의 회소성과 지속성을 기초로 한 중요도 측정을 사용하였다.

또 다른 형태의 비디오 요약은 하이라이트 추출이다. Lienhart 등[13]과 Pfeiffer 등[14]은 저수준 시각적 및 오디오 특징, 모션 정보, 그리고 컬러 정보를 탐색하여 무비 트레일러를 자동으로 생성하는 방법을 제시하였다. Babaguchi[15]는 이벤트 기반 비디오 색인화 방법을 이용한 스포츠 비디오 요약 방법을 제시하였다. 이 방법은 비디오 요약에 유용하긴 하지만 의미를 표현하는 중요한 특징들을 많이 유실한다.

본 논문에서는 자막의 구조적 특징을 이용하여 축구 비디오 하이라이트를 생성하는 방법을 제안한다. 자막의 구조적 특징들은 시간적인 등장 특성과 공간적 위치, 크기 그리고 컬러 특징들을 이용한다.

그러한 자막의 구조적 특징을 이용하여 입력 프레임들로부터 자막 프레임 구간과 자막 키 프레임을 추출한다. 그리고 자막 키 프레임들은 물리적 및 논리적으로 색인화 한다. 축구 비디오 하이라이트는 사전 정의된 하이라이트 생성규칙에 의하여 생성된다.

2. 자막의 구조적 특성

2.1 자막 영역의 분석

축구 비디오에 등장하는 자막은 축구 경기의 핵심 내용을 알리는 중요한 요소이다. 축구 비디오의 자막 영역은 드라마나 다큐멘터리의 자막 영역과는 달리 포지션, 크기, 존재, 변화, 컬러, 등장시점, 그리고 등장순서에서 분명한 특징을 갖는다.

2.2 자막 장면의 분석

축구 비디오의 자막은 제한된 수의 형태로서 등장한다. 본 논문에서는 비디오 내용의 의미에 따라서 다음과 같이 13 형태로 분류한다.

- a) 팀(Ctem): 양 팀의 이름
- b) 경기장(Cpic): 경기장 이름
- c) 중계석(Cbct): 중계자 이름
- d) 심판(Cref): 심판 이름
- e) 선수명단(Clst): 양 팀의 선수 명단 리스트
- f) 경기시작(Cbgn): 경기의 시작
- g) 점수판(Csco): 경기의 점수
- h) 벤치(Cbch): 양 팀의 벤치
- i) 선수(Cplr): 슛, 어시스트, 프리킥, 일반 반칙 등을 범한 선수 이름
- j) 선수교체(Cchg): 교체되어 들어가고 나오는 선수 이름
- k) 골(Cgol): 골을 넣은 선수 이름
- l) 반칙(Cfol): 경고나 퇴장 등 중요 반칙을 범한 선수 이름
- m) 경기종료(Cend): 경기의 종료

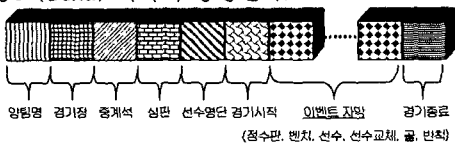
2.3 자막의 시간적 특성

자막들은 등장 순서에 따라 <표 1>과 같이 시작, 이벤트, 종료 자막으로 분류된다. 자막들은 일반적으로 (그림 1)과 같은 시간적 구조를 갖는다.

<표 1> 자막의 시간적 분류

시간적 분류	구성 자막
시작 자막	Ctem, Cpic, Cbct, Cref, Clst, Cbgn
이벤트 자막	Csco, Cbch, Cplr, Cchg, Cgol, Cfol
종료 자막	Cend

본격적인 경기가 시작되기 전에 대부분의 자막들은 Ctem, Cpic, Cbct, Cref, Clst, 그리고 Cbgn 으로 등장하게 되는데, 이 자막들의 등장 순서는 경기마다 약간씩 다르게 등장하는 경우도 있고 다음으로 이벤트 자막이 등장한다. 그리고 경기가 종료될 때 경기종료(Cend) 자막이 등장한다.



(그림 1) 자막의 시간적 흐름

(그림 1)에서와 같이 이벤트 자막은 Csco, Cbch, Cplr, Cchg, Cgol, 그리고 Cfol과 같은 이벤트를 나타낸다. 이 이벤트 자막들은 등장 순서가 일정하지 않으며 축구 경기 수행 중 특정 이벤트가 발생한 직후에 등장한다. 그리고 마지막으로 경기 종료료를 알리는 Cend 자막이 등장하여 경기 결과를 보여준다.

자막의 등장 특성은 <표 2>와 같다.

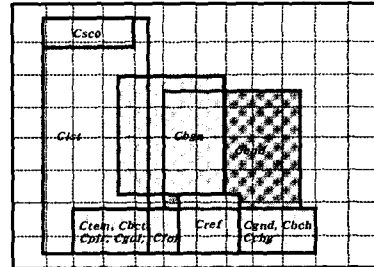
<표 2> 자막의 등장 특성

자막	세부 구성 자막	등장 특성
시작 자막	Ctem, Cpic, Cbct, Clst, Cref, Cbgn	- 경기 시작의 개요를 전달 - 독립적으로 등장
이벤트 자막	Csco	- 축구 경기의 전반에 걸쳐 장·단기적으로 등장 - 등장 시기가 임의적임 - 독립적으로 또는 다른 자막과 중복되어 등장
	Cbch, Cplr, Cgol, Cchg, Cfl	- 이벤트 발생 직후에 등장 - Csco 와 무관하게 등장
종료 자막	Cend	- 경기의 결과를 전달 - 방송 종료

2.4 자막의 공간적 특성

자막의 공간적 구조는 축구 비디오에서 자막의 등장하는 위치 영역의 정보이다. 자막은 그림 2와 같이 6개의 제한된 영역에 등장한다.

(그림 2)에서 Csco, Clst, Cbgn, Cend, 그리고 Cref 자막의 영역이 중복되기는 하지만 서로 다른 자신들만의 크기를 갖고 있다. Ctem, Cpic, Cbct, Cbch, Cplr, Cgol, Cfol, 그리고 Cchg 자막은 각각 크기에서 거의 차이가 없으며 프레임의 하단 부 중앙에 등장한다.



(그림 2) 자막의 공간적 구조

본 논문에서는 자막 영역의 특징을 바탕으로 13 개의 자막 장면들에서 자막 키 프레임을 추출한다.

3. 자막 프레임 구간과 키 프레임 추출

3.1 자막 프레임 구간 추출

(1) 자막 위치의 유사성 측정

$$|x_{fi} - x_i| < T_x \text{ AND } |y_{fi} - y_i| < T_y \quad (1)$$

(식 1)에서 x_{fi} 와 y_{fi} 는 입력되는 프레임의 좌표값이고, x_i 와 y_i 는 구조적 특징에 의해 얻어진 참조값이며, T_x 와 T_y 는 위치 임계값이다.

(2) 자막 크기의 유사성 측정

$$\frac{A_m}{A_f} > T_A \text{ AND } \frac{A_m}{A_i} > T_A$$

where

$$A_m = (\max(x_f, x_i), \max(y_f, y_i), \min(x_f, x_i), \min(y_f, y_i))$$

(식 2)에서 A_m 는 입력 프레임과 참조 프레임 사이의 중복되는 자막 영역의 면적이다. A_f 와 A_i 는 각각 입력 프레임의 자막 영역 면적과 기준 자막 영역의 면적을 나타내고 T_A 는 크기 임계값을 나타낸다.

(3) 자막 컬러의 유사성 측정

$$C_{f_i}^R - C_i^R < T_c \text{ AND}$$

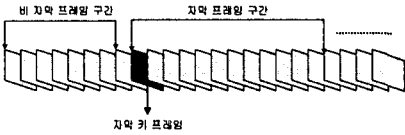
$$C_{f_i}^G - C_i^G < T_c \text{ AND}$$

$$C_{f_i}^B - C_i^B < T_c$$

식 3에서 $C_{f_i}^{RGB}$ 와 C_i^{RGB} 는 각각 입력되는 프레임 자막 영역의 RGB 최대 픽셀수와 기준 자막 영역의 RGB 최대 픽셀수를 나타내고 T_c 는 컬러 임계값을 나타낸다.

3.2 자막 키 프레임 추출

본 논문에서는 (그림 3)과 같이 자막 키 프레임은 자막 프레임 구간의 첫 번째 프레임으로 한다. 자막 키 프레임은 비디오 검색, 브라우징, 그리고 하이라이트 생성을 위하여 중요한 역할을 수행한다.

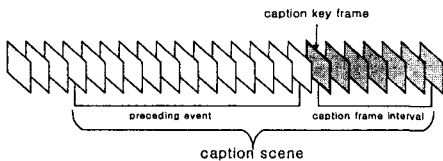


(그림 3) 자막 키 프레임 추출

4. 하이라이트 생성

4.1 장면 재설정

재설정이 필요한 장면들은 반드시 자막 프레임 구간과 선행하는 이벤트를 포함해야 한다. (그림 4)는 재설정이 필요한 장면의 장면 재설정 방법을 보여준다. 자막 프레임 구간은 자신의 자막을 포함하는 모든 프레임들을 포함한다. 선행하는 이벤트는 중요한 이벤트를 나타내기 위해 충분해야 한다.



(그림 4) 장면 재설정

일반적으로 선행하는 이벤트의 충분한 표현 시간은 10초에서 20초 사이이다. 따라서 본 논문에서는 선행 이벤트 시간을 20초로 하였다. 이는 초당 30 프레임의 표준에 따라서 선행 이벤트는 총 600 프레임을 갖는다. 따라서 자막 장면은 선행 이벤트(20초, 600 프레임)와 자막 프레임 구간으로 구성된다.

4.2 하이라이트 생성

하이라이트를 생성하기 위해서는 우선적으로 하이

라이트 기준 장면 수가 결정되어야 한다. 하이라이트 기준 장면 수는 골의 수를 고려하여 유연하게 결정되어야 한다. 하이라이트 생성 규칙은 다음과 같다.

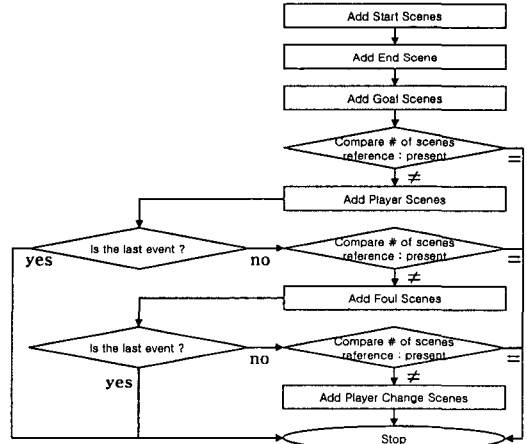
(1) 하이라이트를 구성하는 기본 요소는 C_{tem} , C_{gnd} , C_{bct} , C_{ref} , C_{lst} , C_{bgn} , C_{plr} , C_{chg} , C_{gol} , C_{fol} , 그리고 C_{end} 장면이다. 세부적인 규칙은 다음과 같다.

- 1) 시작 장면
 - ① 가장 처음 장면은 C_{tem} 으로 구성한다.
 - ② 다음은 $C_{gnd}-C_{bct}$ 또는 $C_{bct}-C_{gnd}$ 로 구성한다.
 - ③ 다음은 $C_{ref}-C_{lst}$ 또는 $C_{lst}-C_{ref}$ 로 구성한다.
 - ④ 마지막 장면은 C_{bgn} 으로 구성한다.
- 2) 종료 장면
 - ① 종료 장면은 C_{end} 로 구성한다.
- 3) 이벤트 장면

이벤트 장면은 다음의 이벤트 우선순위에 따라 하이라이트에 포함된다. 각 단계에서 생성된 하이라이트 장면들의 총 수가 하이라이트 기준 장면 수와 비교되는데, 두 수가 같거나 마지막 이벤트가 하이라이트에 포함될 때까지 수행된다.

- ① 가장 최우선 이벤트는 C_{gols} 로 구성한다..
 - ② 두 번째 이벤트는 C_{plrs} 로 구성한다.
 - ③ 세 번째 이벤트는 C_{fols} 로 구성한다.
 - ④ 마지막 이벤트는 C_{chgs} 로 구성한다.
- (2) 만일 지금까지 생성된 하이라이트 장면의 총수가 기준 장면의 수보다 적으면 최종 하이라이트 비디오는 현재의 하이라이트가 된다.

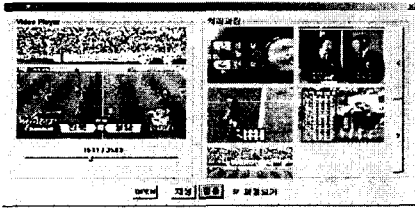
(그림 5)는 하이라이트 생성 흐름을 나타낸다.



(그림 5) 하이라이트 생성 흐름

5. 실험결과

(그림 6)은 자막 영역과 자막 프레임 구간의 추출을 통하여 자막 키 프레임을 추출을 나타낸다.



(그림 6) 자막 키 프레임 추출 화면
 <표 3>은 자막 키 프레임 추출 결과를 보여준다.
 <표 3> 자막 키 프레임 추출 결과

경기	자막 키 프레임 수
Video A	62
Video B	54
Video C	67
Video D	58

(그림 7)은 축구 비디오에서 하이라이트 생성 규칙에 따라 하이라이트를 생성하는 것을 보여준다.



(그림 8) 하이라이트 생성 화면

하이라이트 생성에서 하이라이트 기준 장면 수는 20으로 설정하였다. 입력 비디오의 평균 재생 시간은 58.3분이고 하이라이트 비디오의 평균 재생 시간은 6.7분이다. 비디오 시청자들은 전체 비디오를 보지 않고 6.7분 내에 전체 비디오 내용을 시청하고 이해할 수 있다.

6. 결론

본 논문에서는 자막의 시간적 특징과 공간적 특징과 같은 구조적 특징을 이용하여 축구 비디오 하이라이트를 생성하는 효율적인 방법을 제시하였다. 자막의 구조적 특징은 자막 정보의 분석에 의해 추출되고 자막 프레임 구간과 자막 키 프레임은 이들 특징들을 이용한 추출하였다. 또한 하이라이트를 생성하기 위하여 효율적인 하이라이트 생성 규칙을 이용하였다. 사용자들은 원하는 비디오를 빠르게 편리하게 시청하고 이해할 수 있다.

참고문헌

[1] G. Davenport, T. Smith, and N. Pincever, "Cinematic Primitives for Multimedia," *Computers and Graphics*, Vol. 15, pp. 67-74, 1991.
 [2] I. Guyon, "Applications of Neural Networks to Character Recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.

5, pp. 353-382, 1991.

[3] S. Harmalkar and R. M. K. Sinha, "Integrating Word Level Knowledge in Test Recognition," In *Proc. of ICPR*, pp. 758-760, 1990.
 [4] F. Li and S. S. Yu, "Handprinted Chinese Character recognition Using Probability Distribution Feature," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 8, pp. 1241-1258, 1994.
 [5] M. A. O'Hair and M. Kabrisky, "Recognizing Whole Words as Single Symbols," In *Proc. of ICDAR*, pp. 350-358, 1991.
 [6] J. Hernando, "Voice Signal Processing and Representation Techniques for Speech Recognition in Noisy Environments," *Signal Processing*, Vol. 36, No. 3, pp. 393, 1994.
 [7] Y. Pan, J. Wu, S. Tamura, and K. Okazaki, "Neural Network Vowel-Recognition jointly Using Voice Features and Mouth Shape Image," *Pattern Recognition*, Vol. 24, pp. 921-927, 1991.
 [8] K. Hang-Bong, "Generation of Video Highlights Using Video Context and Perception," *Proc. of SPIE, Storage and Retrieval for Media Databases 2001*, Vol. 4315, pp. 320-399, 2001.
 [9] M. Christal, M. Smith, C. Taylor and D. Winkler, "Evolving Video Skims into Useful Multimedia Abstractions," *Proc. CHI'98*, pp. 171-178, 1998.
 [10] M. Yeung, B. Yeo and B. Liu, "Segmentation of Video by Clustering and Graph Analysis," *Computer Vision and Image Understanding*, Vol. 71, No. 1, pp. 94-109, 1998.
 [11] A. Hanjalic and H. Zhang, "An Integrated Scheme for Automated Video Abstraction Based on Unsupervised Cluster-Validity Analysis," *IEEE Trans. Cir. & Sys. for Video Tech.*, Vol. 9, No. 8, pp. 1280-1289, Dec. 1999.
 [12] S. Uchihashi, J. Foote, A. Girgenson and J. Boreczky, "Video Manga: Generating Semantically Meaningful Video Summaries," *Proc. ACM MM'99*, 1999.
 [13] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video Abstracting," *Communications of the ACM*, Vol. 40, No. 12, pp. 54-62, 1997.
 [14] S. Pfeiffer, R. Lienhart, S. Fisher and Effelsberg, "Abstracting Digital Movies Automatically," *Int. Jour. Visual Communication and Image Representation*, Vol. 7, No. 4, pp. 345-353, 1996.
 [15] N. Babaguchi, "Towards Abstracting Sports Video by Highlights," *Proc. IEEE Int. Conf. Multimedia and Expo(III)*, pp. 1519-1522, Aug. 2000.