자막 분석을 이용한 축구비디오 요약

신성윤, 강일고, 이양원

군산대학교 컴퓨터정보과학과
전라북도청 공보정보화과

Soccer Vodeo Summarization Using Caption Analysis

Seong-Yoon Shin, Il-Ko Kang, Yang-Won Rhee
Dept. of Computer Information Science, Kunsan National University
A Provincial Office of Jeollabuk-do
{syshin, ywrhee}@cs.kunsan.ac.kr

요 약

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

1. 서론

비디오 요약(summary) 생성은 대용량의 긴 비디오 데이터에서 중요한 내용들만을 추출하여 보다 핵심적인 내용을 짧은 시간에 보여줄 수 있는 비디오 신(scene)을 생성하는 것이다.

비디오 요약을 생성하기 위하여 먼저 비디오를 일정 기준에 따라 샷(shot)으로 분류하는 작업[1,2,3]이 수행되어야 하며, 샷들은 대표 프레임(representative frame)[4] 또는 키 프레임(key frame)[5,6]이라는 중요한 프레임을 갖게 되며 색인화에 사용된다.

비디오 요약 생성과 관련된 연구에는 이미지 템플리트(template), 통계학적 특징, 그리고 히스토그램(histogram) 기반 검색과 처리를 이용하는 방법 등이 많이 이용되고 있다[7].

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

[9]에서는 비디오의 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 이용하기도 하였다.

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

본 논문에서는 축구 경기 비디오 요약을 생성하는 새로운 방법을 제시하는데, 캡션을 이용하여 중요 이벤트(event)를 중심으로 비디오 요약 생성 규칙에 따라 요약된 비디오를 생성한다. 캡션에 대하여 템플리트 매칭과 지역적 차영상을 적용하여 캡션 키 프레임을 추출하고 색인화 하도록 한다. 색인화된 비디오 데이터는 비디오 요약 규칙에 따라 요약된 비디오로 생성된다.

전체적인 시스템 구조는 그림 1과 같으며 2장에서는 캡션이 갖는 세부적인 특징들을 파악하고 이를 분석하도록 하며, 3장에서는 캡션이 갖는 사전 지식을 바탕으로 템플리트 매칭과 지역적 차영상을 이용하여 캡션 키 프레임 추출하고 색인화 하도록 한다. 그리고 4장에서는 비디오 요약 생성 규칙에 따른 요약된 비디오 생성에 대해서 설명하고. 5장에서는 실험 데이터를 이용하여 결과를 분석하고 6장에서 결론을 맺도록 한다.

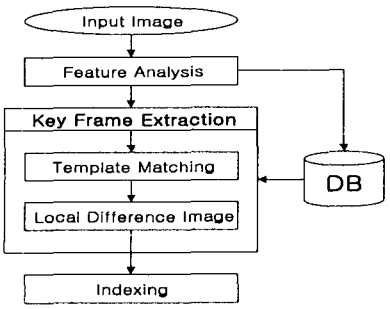

그림 1. 시스템 구조

2. 캡션 분석

캡션의 특징 추출을 위하여 축구 경기 비디오에 등장하는 대표적인 캡션을 표 1과 같이 10가지 종류로 제한하여 분류하였으며 분석을 통해 캡션들은 다음과 같은 명확한 특징을 갖는다.

(1) 각 캡션들의 위치는 종류별로 일정하다.

(2) 각 캡션들은 종류에 따라 크기가 일정하다.

(3) 캡션들은 일정 시간동안 나타났다가 사라진다.

(4) 캡션 자체에서 내용이 변하는 부분의 위치는 항상 고정되어 있다.

(5) 캡션 형성 영역을 구성하는 컬러는 거의 일정한 컬러값을 갖는다.

(6) 점수판 캡션의 간헐적 등장을 제외하고 모든 캡션은 어떠한 이벤트가 발생한 직후 등장했다가 사라진다.

(7) 캡션등장순서는 방송(경기)마다 다를 수 있다.

표 1. 캡션의 분류

구분	종류	특징	설명
CSi (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_{flt}	반칙 (fault)	팀과 선수 반칙 표시
	C_{bch}	벤치 (bench)	감독, 코치나 대기선수표시
	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_{bct}	중계석 (caster)	아나운서와 해설자 표시
	C_{plr}	선수이름 (player name)	선수 표시

3. 캡션 키 프레임 추출과 색인화

3.1 캡션 키 프레임 추출

캡션 키 프레임 추출은 캡션이 갖는 특성을 바탕으로 사전 지식과의 비교를 통한 탬플리트 매칭을 이용하여 캡션의 등장을 인식하고 캡션 내부의 일정 영역에 대한 지역적 차영상을 이용하여 캡션의 내용변화를 인식하는 방법으로 추출하는데, 다음과 같은 경우에 캡션 키 프레임으로 추출된다.

1) 캡션이 등장한 첫 번째 프레임

캡션의 등장에 대한 인식은 먼저 10가지 종류의 캡션 영역에 대하여 최소사각영역을 설정한 후, 캡션의 위치와 크기 및 컬러 정보를 바탕으로 다음 식의 탬플리트 매칭을 통한 유사성 측정 방법을 이용하여 추출한다.

Similarity=CIi(p, s, c)-MIj(p, s, c),
where i=1…m, j=1…n

여기서 MIj(p, s, c)는 사전 지식에 의한 탬플리트를 말하고 CIi(p, s, c)는 입력되는 캡션 영역을 나타낸다. 여기서 p는 캡션의 위치(position), s는 캡션의 크기(size), 그리고 는 c컬러(color)값을 말한다.

2) 캡션 내용이 변한 첫 번째 프레임

캡션이 등장하여 존재하고 있는 상태에서 캡션 내부에서 내용이 변화된 것을 추출하기 위한 방법으로서, 캡션의 유동영역을 설정한 다음 식과 같은 지역적 차영상을 이용하여 캡션 내용의 변화를 추출하고 내용이 변한 첫 번째 프레임을 키 프레임으로 추출한다.

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

여기서 Ia와 Ib는 입력되는 두 이미지의 두 캡션 영역을 나타낸다.

3.2 캡션 키 프레임 색인화

색인화는 다음 식과 같이 하나의 비디오 스트림을 V라 하면, V는 캡션 샷들의 키 프레임인 CSi들로 구성되며, 이 CSi는 샷을 구성하는 각각의 프레임 CFj들로 구성된다.

$$V=\sum_{i=1}^{n}CS_i$$

$$CS_i=\sum_{j=1}^{m}CF_j$$

논리적 색인화는 키 프레임들을 여러 형태로 연결하여 해당 캡션 샷에 접근 가능한데, 캡션 샷의 물리적/논리적 색인화의 구성형태는 그림 2와 같다.

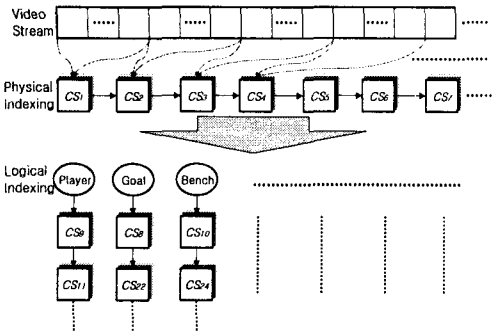

그림 2. 캡션 샷의 색인화 구조

4. 비디오 요약 생성

4.1 요약을 위한 샷의 재 설정

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

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

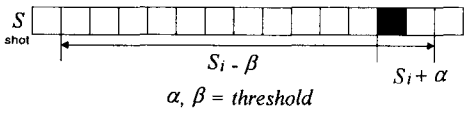

그림 3. 요약을 위한 샷의 재 설정 영역

4.2 요약 비디오 생성

비디오 요약은 CSi를 중심으로 기본 20개의 샷을 하나의 요약 비디오로 설정하여 구성한다. 요약 생성 규칙은 다음과 같으며, 그림 4는 비디오 요약의 한 예를 나타낸다.

1) 기본적 요약 구성 샷은 Csco,Cgol, Cbgn, Cend, Clst, Cbct 의 6개이다.

① 첫 번째 샷은 Cbgn 과 Cbct 둘 중 하나이며, 둘 중에서 하나가 선택되면 다른 하나는 바로 뒤에 따른다.

② 위의 ①이 결정된 다음에는 Clst샷이 따른다.

③ 골이 있어 득점을 한 경우에는 Cgol과 Csco샷이 골 득점 수에 맞게 추가된다.

④ 마지막 샷은 Cend 와 Cbct 둘 중 하나이며, 둘 중에 하나가 선택되면 다른 하나는 바로 앞에 추가된다.

⑤ 경우에 따라 Cbct는 나타나지 않을 수 있다.

2) 캡션 샷 Cchg, Cflt, Cplr 그리고 Cbch를 선택적으로 사용자가 추가하는 것을 원칙으로 한다.

3) 시간의 흐름과 무관하게 추가될 샷들을 선택할 수도 있다.

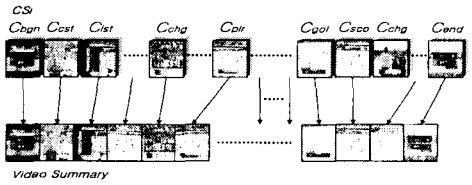

그림 4. 비디오 요약 생성 예

5. 실험 및 분석

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

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

표 2. 캡션 키 프레임 추출 결과

구분	캡션에 의한 키 프레임 추출 수
경기 A	53
경기 B	47
경기 C	57
경기 D	45

표 3. 평균 재현 시간(초)

구분	실제데이터	요약 비디오	생성율
평균재현시간	2852	327	11%

평균 재현 시간은 표 3과 같다. 여기서 생성율은 요약된 정도를 나타내는 것이며, 비디오에서 방송상의 문제에 따른 갑작스런 캡션 위치, 크기 그리고 컬러 정보가 변하지 않는다면 최적의 요약 비디오를 생성하는 것이 가능하다고 판단된다.

6. 결론

본 논문에서는 축구 비디오 데이터에서 경기의 전반적인 내용을 짧은 시간에 적은 양의 내용을 표현하여 이해할 수 있도록 이벤트-캡션을 기반으로 한 요약 비디오를 생성하는 방법을 제시하였다. 결론적으로, 중요한 이벤트 발생에 따라 등장하는 캡션에 의한 정보를 바탕으로 요약 비디오를 요약 생성 규칙에 따라 생성하였다.

또한 캡션의 등장과 내용 변화를 탬플리트 매칭과 지역적 차영상을 이용한 유사성 측정으로 인식하고 캡션 키 프레임을 추출하여 물리적/논리적 색인화를 유용하게 수행할 수 있는 기반을 제시하였다.

향후 방송에서 정확한 캡션의 분류를 통한 상영과 캡션 문자 인식에 대한 연구가 진척된다면 스포츠 비디오의 요약에 커다란 성과를 보게 될 것이다.

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., "Content-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 Characterization", 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, Swizerland, 2000