

축구 비디오 하이라이트 생성

전 근 환[†]·신 성 윤[†]·이 양 원^{**}·류 근 호^{***}

요 약

비디오 하이라이트(highlights)는 원래의 비디오 보다 짧고 많은 양의 의미를 갖는다. 기존의 파노라마 형태의 추상화 기법은 여러 프레임을 하나의 프레임으로 모자이크하는 형태이었고, TV 드라마 하이라이트 방법은 카메라의 이동이나 특수효과에 의존하기 때문에 스포츠 비디오에 적용은 부적합하다. 이 논문에서는 축구 비디오를 대상으로 시각정보와 자막을 이용하는 새로운 비디오 하이라이트 생성 방법과 이벤트 기반 비디오 인덱싱 방법을 제안한다. 하이라이트 생성은 하이라이트 생성 규칙에 따라 자막에 의해 추출된 샷을 중심으로 시각정보에 의해 추출된 샷을 합성하여 생성하였고, 인덱싱은 자막으로 추출된 샷은 주요소로, 시각정보에 의해 추출된 샷은 부가적 요소로 구성하였다. 실험에서는 샷 추출기법 중 대표적인 컬러히스토그램과 χ^2 히스토그램과의 성능을 비교하여 제안한 하이라이트 기법이 다른 방식보다 우수함을 증명하였다.

Creating highlights of Soccer video

Keun Hwan Jeon[†] · Seong Yoon Shin[†] · Yang Weon Rhee^{**} · Keun Ho Ryu^{***}

ABSTRACT

Creating Video Highlights is not only shorter than original video, but also has more meaningful highlight scene. In current, the panoramic-abstracting scheme and the TV-drama highlights scheme are inadequate for sports video because panoramic-abstracting scheme create one panoramic frame which is integrated by many frames, and TV-drama highlights scheme depend on camera position and special editing. In this paper, we propose the new scheme of creating highlights by using a visual and caption information and an event-based video indexing, especially for soccer video. Highlight creation is performed on some pieces of shots extracted by caption and visual information by highlight creation rule. In indexing, the shots extracted by the captions are the main elements, and the shots extracted by the visual informations are optionally added to the main elements. Using simulation model, the performance of the proposed scheme is compared with those of two representative shot extraction schemes, the color histogram and the χ^2 histogram. The performance results indicate that proposed highlight scheme is better than other schemes.

키워드 : 비디오 하이라이트(video highlight), 키 프레임(key frame), 컬러, 크샤이-제곱 히스토그램(color and χ^2 histogram), 비디오 인덱싱(video indexing)

1. 서 론

비디오에서 하이라이트(highlight) 생성은 대용량의 긴 비디오 데이터에서 중요한 내용을 추출하여 보다 함축적이면서 짧은 비디오 씬(scene)을 생성하는 것으로서, 멀티미디어 콘텐츠(multimedia contents) 제작자나 사용자 관점에서 볼 때 매우 중요한 역할을 수행한다.

다양한 멀티미디어 데이터 중에서 특히, 비디오 데이터는 시각적 요소, 청각적 요소, 그리고 문자 스트림(stream)들과 같은 다양한 정보들이 동기화 되어 사용자들에게 제공되는 것으로서 멀티미디어 시스템에서 다양한 형태와 시간적 차원을 갖는 핵심적인 요소이다[1]. 이처럼 중요한 비디오 데이터에서, 사용자들에게 필요한 중요 내용을 보다 빠르고

편리하게 검색, 재생 그리고 편집 할 수 있는 기능을 제공하기 위해서는 저장자료의 빠른 검색이 필요하다.

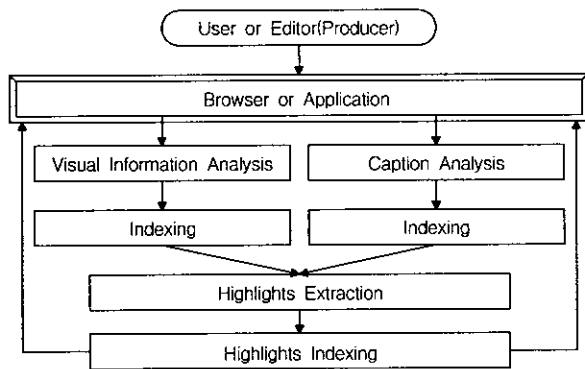
또한, 비디오 하이라이트를 생성하기 위하여 먼저, 비디오를 시간의 흐름에 따라 중요한 정보를 포함하는 샷(shot)으로 분류하는 작업이 수행되어야 한다[2-4]. 이렇게 생성된 샷들은 제작자나 사용자들이 쉽게 검색하여 편집하고 재생할 수 있도록 논리적으로 인덱싱(indexing)되어 브라우징(browsing)이 가능하도록 구축되어야 한다. 인덱싱을 위하여 샷들은 대표 프레임(frame)[5] 또는 키 프레임(key frame)[6, 7]이라는 샷을 대표하는 중요한 프레임을 갖는다.

비디오에서 하이라이트 추출은 곧 비디오를 요약하는 것과 유사하다. 이 하이라이트 추출은 표준 시간적 간격에서 이산적인 프레임들을 선택하는 것으로서 실제 비디오를 대상으로 수행되어 왔다. 실제로 스포츠 비디오는 거의 비슷하게 반복되는 장면들이 비디오의 대부분을 차지하기 때문에 하이라이트 추출은 필수적이다.

† 정 회 원 : 군장대학 컴퓨터응용학부 교수
 ** 중신회원 : 군산대학교 컴퓨터정보학과 교수
 *** 중신회원 : 충북대학교 전기전자컴퓨터공학부 교수
 논문접수 : 2001년 7월 12일, 심사완료 : 2001년 8월 21일

이 논문에서는 시각정보와 자막에 의한 키 프레임 인덱싱 방식을 사용하여 하이라이트 추출 방안을 제안한다. 하이라이트 추출은 시각적인 정보와 자막 정보를 함께 이용하여 중요 이벤트(event)를 중심으로 하이라이트 생성 규칙에 따라 하이라이트를 생성하도록 한다. 먼저, 시각정보에 대하여 컬러 히스토그램(color histogram)과 χ^2 히스토그램을 합성한 방법으로 시각적 키 프레임을 추출하고, 자막정보에 대하여 템플릿 매칭(template matching)과 지역적 차영상에 의하여 자막 키 프레임을 추출하여 이들을 기반으로 인덱싱을 구축한다. 이렇게 인덱싱된 비디오 데이터는 하이라이트 생성 규칙에 따라 시각적 샷과 자막 샷이 합성되어 하이라이트로 생성된다.

하이라이트 생성을 위한 전체적인 시스템 구조는 (그림 1)과 같으며 2장에서는 관련연구를, 3장에서는 시각정보에 의한 키 프레임 추출, 4장에서는 자막에 의한 키 프레임 추출 그리고 5장에서는 하이라이트 생성에 대해서 설명한다. 6장에서는 실험과 결과를 분석하고 7장에서 결론을 맺도록 한다.



(그림 1) 축구 하이라이트 생성 시스템 구조

2. 관련 연구

스포츠 경기에서 해당 경기의 전체가 흥미진진한 경우는 매우 드물다. 따라서 특정 경기의 중요 장면들만을 요약하기 위한 시도가 많은 연구자들에 의해 수행되었다.

[8]에서는 비디오를 요약하기 위해 이미지 템플릿(template), 통계학적 특징, 그리고 히스토그램(histogram) 기반 검색과 처리를 이용하였고, [9]에서는 시각적인 특징과 청각적인 특징을 모두 고려하여 원래 비디오의 시놉시스(synopsis)를 표현하는 비디오 스킴(skim)을 구축하였는데, 이 방법은 원래 비디오의 세그먼트(segment)들을 병합하는 방법을 이용한 것이다.

[10]에서는 스토리(story) 내용을 묘사하는데 비디오 포스터(video poster)를 제안하였으며, [11]에서는 비디오 내의 씬을 구분하기 위해서 두드러진 모션(motion)이나 다양한 히스토그램 특징들을 분석하여 이용하였다. 또한 [12, 13]에서는 고정된 환경의 고정된 도메인(domain) 궤도 내에서

움직이는 객체를 추출하고 인식하여 그들의 모션을 분류하는 방법을 이용하였고, [14]에서는 객체의 활동 궤도를 시뮬레이션(simulation)하기 위하여 시공간적 통합 데이터세트(dataset)를 생성하여 이용하였다.

위와 같은 연구들은 서로 다른 표현모델을 이용하여 여러 프레임들 속에 나타나는 객체 정보를 추출하여 전체 프레임 정보를 파노라마 형태의 한 프레임 속으로 함축(abstracting)시키는 연구 결과를 제시하였다. 하지만 이러한 파노라마 형태 요약은 중요 객체(축구선수)가 상당수(22명) 존재하는 축구와 같은 스포츠 비디오에 적용시키기는 부적합하다.

MoCA 프로젝트[3]에서는 TV 드라마를 대상으로 샷(shot)과 씬을 탐색하여 비디오 하이라이트 구축을 시도하였다. 샷은 카메라의 움직임과 컷(cut), 페이드(fade), 디졸브(dissolve) 등과 같은 편집 효과 탐색 방법을 이용하여 탐색하였고, 씬은 카메라 각도 변화에 따른 배경 컬러의 양을 이용한 탐색 방법과 오디오의 변화를 이용한 탐색 방법을 적용하였다. 탐색된 샷과 씬을 이벤트클립, 대화클립, 기타 클립으로 나누어 하이라이트를 생성하였다.

MoCA 프로젝트에서 제안한 하이라이트 방법을 축구와 같은 스포츠 비디오에 적용시키기 위해서는 먼저 샷과 씬 탐색기법부터 적합하지 않다. 드라마의 경우에는 이벤트의 변화를 카메라 각도와 편집효과를 이용하여 탐지할 수 있으나, 축구 비디오에서는 수 많은 카메라가 투입되며, 카메라의 변화나 사운드의 변화를 이용하여 파울, 선수교체, 골인 등과 같은 이벤트의 탐색은 불가능하다.

이 논문에서는 비디오에서 기본적으로 존재하는 장면의 변화와 스포츠 비디오에서 중요한 의미를 전달하는 자막을 이용하여 하이라이트를 구성하는 기법을 제시한다.

3. 시각정보에 의한 키 프레임 추출

3.1 장면 변화에 따른 키 프레임 추출

키 프레임 추출은 각각의 샷들을 대표하는 프레임을 추출하는 것으로서, 장면 변화가 크게 발생하는 지점의 첫 번째 프레임을 키 프레임으로 설정하여 비디오 인덱싱을 통한 비디오 편집, 검색 또는 브라우징 등에 중요하게 사용된다.

이와 같이 장면을 분할하여 키 프레임을 추출하기 위한 방법에는 프레임간의 차이를 이용하는 방법[15, 16], 컬러 히스토그램을 이용하는 방법[17-20], 객체의 이동을 계산하여 사용하는 방법[21, 22], 에지(edge) 변화를 추출하여 이용하는 방법[23] 그리고 전체적인 웨이블릿(wavelet)을 측정하여 이용하는 방법[24] 등이 사용된다.

일반적으로 갑작스런 장면 변화의 검출에는 식 (1)과 같은 컬러히스토그램의 차이값을 계산하여 사용하는 방법이 보편적으로 사용된다[17-20]. 여기서 $d(I_i, I_{i+1})$ 은 이웃하는 두 프레임 (I_i, I_{i+1}) 간 유사성을 측정하는 함수이며 $H_i(j)$ 는

i 번째 프레임의 j 번째 히스토그램 값을 나타낸다.

$$d(I_i, I_{i+1}) = \sum_{k=0}^{N-1} |H_i(k) - H_{i+1}(k)| \quad (1)$$

컬러 히스토그램은 식 (1)의 방법이 많이 사용되지만 식 (2)와 같이 RGB 컬러 공간대를 따로 계산하여 사용할 수도 있다[10].

$$d_{RGB}(I_i, I_j) = \sum_{k=1}^n (|H_i^r(k) - H_j^r(k)| + |H_i^g(k) - H_j^g(k)| + |H_i^b(k) - H_j^b(k)|) \quad (2)$$

또한 컬러 히스토그램이나 템플릿 매칭(template matching) 방법보다 좋은 결과를 나타내기 때문에 많은 연구에서 사용하는 χ^2 히스토그램 방법이 있다[20]. 여기서 $H_i(k)$ 는 i 번째 프레임의 히스토그램에서 k 번째 bin의 값을 의미한다.

$$d(I_i, I_j) = \sum_{k=1}^n \frac{(H_{i(k)} - H_{j(k)})^2}{H_{j(k)}} \quad (3)$$

위의 식 (3)의 변형된 형태가 다음 식 (4)와 같다.

$$d(I_i, I_j) = \sum_{k=0}^n \sqrt{H_{i(k)}^2 - H_{j(k)}^2} \quad (4)$$

이 연구에서 제안하는 키 프레임 추출 방법은, 시각정보에 의한 인덱싱을 위하여 컬러 히스토그램과 χ^2 히스토그램의 장점을 합성한 식 (5)의 방법을 이용하는데, 각 프레임 사이의 차이 값을 계산하고 차이 값의 크기에 따라 키 프레임을 추출하여 샷들을 구분하는 계산 방법이다.

$$d(I_i, I_j) = \sum_{k=1}^n \left(\frac{(H_i^r(k) - H_j^r(k))^2}{H_i^r(k)} \times 0.299 + \frac{(H_i^g(k) - H_j^g(k))^2}{H_i^g(k)} \times 0.587 + \frac{(H_i^b(k) - H_j^b(k))^2}{H_i^b(k)} \times 0.114 \right) / 3 \quad (5)$$

식 (5)를 이용하여 얻어진 히스토그램의 차이 값에 따라 차이 값 변화가 큰 부분의 전환점이 되는 프레임을 찾아 이 프레임을 키 프레임으로 설정하도록 한다. 스포츠 비디오에서 키 프레임은 첫 번째 프레임을 포함하여 추출된 각 샷들의 첫 번째 프레임이 된다.

3.2 시각정보에 의한 샷의 분석

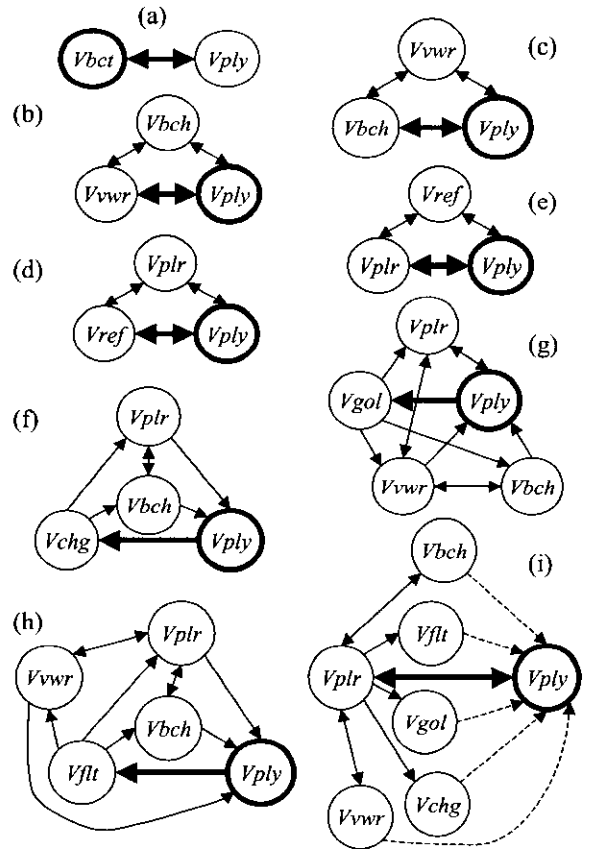
시각정보에 의한 샷을 분석하면 실제 장면 변화가 큰 부분에서 발생하므로 상당히 많은 수의 샷들이 존재하게 된다. 이렇게 많은 수의 샷들을 각각의 개별적인 개념으로 서로 다르게 분류하는 것은 상당히 복잡하므로, 이 연구에서는 샷들을 크게 <표 1>과 같이 9가지 형태로 분류하여 자막정보에 의해 추출된 샷들과 합성하여 하이라이트를 생성

하도록 한다.

<표 1> 시각정보에 의한 샷의 분류

구분	종류	특징	설명
VS _i (Visual Shot)	U_{ply}	연속경기(play)	운동장에서 경기장면
	U_{plr}	선수초점(player)	특정 선수 확대장면
	U_{flt}	반칙(fault)	반칙 선수 확대장면
	U_{chg}	선수교체(change)	선수교체 확대장면
	U_{gol}	골(goal)	골 넣은 선수 확대장면
	U_{vwr}	관람석(viewer)	관람석 장면
	U_{bct}	경기중계(broadcasting)	캐스터와 해설자장면
	U_{bch}	벤치(bench)	벤치장면
	U_{ref}	심판(referee)	심판 확대장면

이렇게 분류된 샷들 사이의 관계를 보면 (그림 2)와 같이 가장 많이 나타나는 샷 사이의 시간적 흐름에 따른 구성을 알 수 있다.



(그림 2) 샷 사이의 흐름도

(그림 2)에서 굵은 원의 노드(○)와 굵은 화살표(→)는 각각 시작 노드(샷)와 다음 노드로의 연결을 나타내며, 가는 원의 노드(○)와 가는 화살표(→)는 각각 하위 노드와 다른 노드와의 연결을 나타낸다. 또한 실선의 화살표(→)는 다른 종속된 노드들이 하위에 연결될 수 있음을 나타낸다. 예를 들

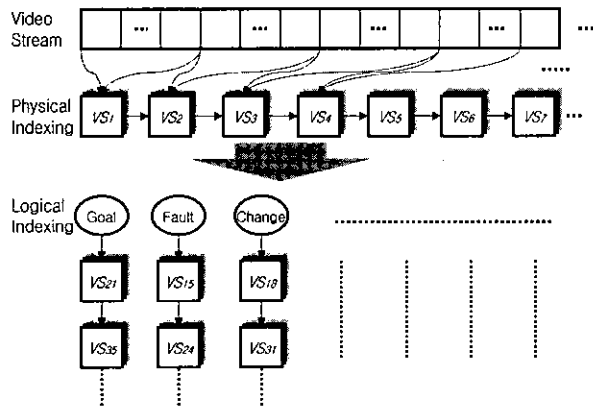
어 경기 도중 골을 넣은 경우(그림 2b) 그 다음에 나타나는 샷은 골을 넣은 선수, 벤치 또는 관람석 중 어떠한 하나의 샷이 나타날 수 있으며, 이 세 샷들은 서로 임의의 순서대로 등장할 수 있고 마지막에는 다시 경기를 수행하게 된다.

3.3 시각정보에 의한 키 프레임 인덱싱

샷들의 키 프레임들은 추출된 순서에 따라 장면 변화가 큰 프레임들 중심으로 자동적으로 일련번호에 의해 물리적으로 인덱싱 된다. 하나의 스포츠 비디오 스트림을 V 라 하면 시각정보에 의한 인덱싱은 식 (6)과 같이 V 는 각 샷들의 키 프레임 VS_i 들로 인덱싱 되어 구성되며, 이 VS_i 는 샷을 구성하는 각각의 프레임 VF_j 들로 구성된다. 여기서 샷은 식 (1)에서 제안한 방법에 따라 장면변화가 큰 키 프레임과 이에 종속된 프레임들의 집합을 말하며, 프레임은 이 샷들을 구성하는 최소 단위를 말한다.

$$V = \sum_{i=1}^n VS_i, VS_i = \sum_{j=1}^m VF_j \quad (6)$$

이들에 대하여 분야별 또는 사건별로 의미를 부여하여 논리적인 구조로 인덱싱 할 수 있다. 시각적 샷의 물리적/논리적 인덱싱의 구성 형태는 (그림 3)과 같다.



(그림 3) 시각적 샷의 물리적/논리적 인덱싱 구조

(그림 3)에서 논리적인 인덱싱은 사용자가 원하는 이벤트를 중심으로 물리적으로 인덱싱된 키 프레임 번호를 연결하여 다양한 형태로 구성이 가능하며, 스포츠 비디오의 어떤 장면이라도 직접 접근이 가능하다.

4. 자막에 의한 키 프레임 추출

4.1 자막의 분석

비디오 데이터에서 자막은 시청자에게 중요한 내용을 전달하는 대표적인 수단이다. 스포츠비디오에서 자막은 거의 일정한 위치에 존재하게 되며 이벤트가 발생 할 경우 화면에 표시되어 일정하거나 변화된 내용을 전달한다. 이 논문

에서는 실험 대상인 축구 경기 비디오에서 자막을 다음 <표 2>와 같이 10가지 종류로 한정하여 분류하였다. 여기서 구성의 팀이나 아나운서, 해설자, 감독, 코치, 선수 리스트 그리고 선수는 각각 팀 또는 사람의 이름들을 나타낸다.

<표 2> 자막의 분류

구분	종류	특징	설명
CSI (Caption Shot)	C_{scb}	점수판(score board)	팀과 점수 및 시간표시
	C_{got}	골(goal)	점수와 선수 및 시간표시
	C_{chr}	선수교체(player change)	교체되는 양 선수 표시
	C_{fr}	반칙(fault)	팀과 선수 반칙표시
	C_{bch}	벤치(bench)	감독, 코치 또는 대기선수표시
	C_{bgn}	경기시작(game begin)	양 팀표시
	C_{end}	경기종료(game end)	양 팀과 점수표시
	C_{lst}	선수명단(player list)	팀과 선수 리스트표시
	C_{cst}	중계석(caster)	아나운서와 해설자표시
	C_{plr}	선수이름(player name)	선수표시

자막의 분석에 의해 자막들은 다음과 같은 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.

4.2 자막에 의한 키 프레임 추출

자막에 의한 키 프레임 추출은 자막들이 갖는 명확한 특성을 바탕으로 템플릿 매칭에 의한 자막 등장 인식과 지역적 차영상을 통한 자막의 내용 변화 인식을 이용하여 추출하는데, 다음과 같은 경우 키 프레임으로 설정한다.

① 자막이 나타난 첫 번째 프레임

자막이 나타난 첫 번째 프레임은 무조건 키 프레임으로 설정하게 되는데, 자막 등장의 인식은 최소 사각 영역을 설정하여 자막의 위치와 크기 및 컬러 정보를 바탕으로 식 (7)과 같은 방법을 이용하여 추출한다. 여기서 $MI_j(p, s, c)$ 는 사전지식을 바탕으로 생성된 자막의 위치(p), 크기(s) 그리고 컬러(c)값을 갖는 자막의 템플릿을 말하며 $CI_i(p, s, c)$ 는 입력되는 프레임들의 비교될 자막 정보를 나타낸다.

$$Similarity = CI_i(p, s, c) - MI_j(p, s, c) \quad (7)$$

, where $i = 1 \dots m, j = 1 \dots n$

② 존재하는 자막의 내용이 변한 첫 번째 프레임
자막이 존재하고 있는 상태에서 자막 내의 내용이 변하는

경우에는 자막 내의 유동 영역을 설정하여 식 (8)과 같은 지역적 차영상을 이용하여 자막 내의 내용 변화를 추출하고 내용이 변한 첫 번째 프레임에 키 프레임으로 설정한다.

$$LD(x, y) = \sum_{y=0}^N \sum_{x=0}^N |I_a(x, y) - I_b(x, y)| \quad (8)$$

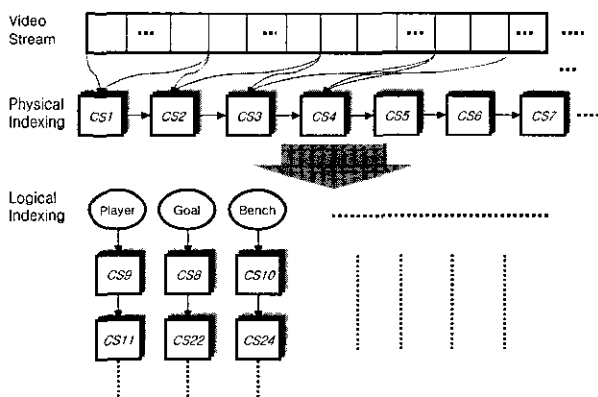
4.3 자막에 의한 키 프레임 인덱싱

자막에 의한 물리적 인덱싱은 식 (9)와 같이 하나의 비디오 스트림을 V 라 하면, V 는 자막에 의해 추출된 각 샷들의 키 프레임 CS_i 들로 인덱싱되어 구성되며, 이 CS_i 는 샷을 구성하는 각각의 프레임 CF_j 들로 구성된다. 여기서 샷은 템플릿 매칭과 차영상을 이용하여 추출한 자막의 키 프레임과 이에 중속된 프레임들의 집합을 말한다.

$$V = \sum_{i=1}^n CS_i, CS_i = \sum_{j=1}^m CF_j \quad (9)$$

논리적인 인덱싱 또한 시각적 샷의 논리적 인덱싱과 마찬가지로 사용자가 원하는 다양한 형태로 구성이 가능한데, 물리적으로 인덱싱된 키 프레임들을 연결하여 원하는 이벤트에 해당하는 자막 샷에 접근이 가능하다.

자막 샷의 물리적/논리적 인덱싱의 구성 형태는 (그림 4)와 같다.



(그림 4) 자막 샷의 물리적/논리적 인덱싱 구조

(그림 4)에서 물리적 인덱싱은 자막 정보에 의해 추출된 키 프레임들을 중심으로 구성되고, 논리적 인덱싱은 <표 2>의 자막의 종류에 따라 이벤트를 중심으로 구성된다.

5. 하이라이트 생성

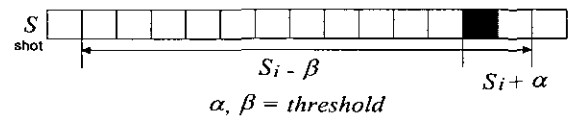
5.1 샷 설정

샷 설정이란 키 프레임을 중심으로 그에 부속되는 프레임들을 할당하여 하나의 샷을 구성하는 것을 말하며 시각적 샷과 자막 샷 모두 내용의 흐름에 맞게 설정해 주어야 한다. 먼저 시각정보에 의해 추출된 키 프레임을 보면, 이 키

프레임들은 특정 이벤트가 발생한 후 갑작스런 장면 전환이 발생하여 추출되었다. 따라서 V_{ply} 와 V_{bct} 를 제외한 모든 샷들은 키 프레임 이전의 이벤트를 포함하여야 한다.

또한, 자막의 표시도 간헐적인 점수판의 표시를 제외하고는 항상 이벤트 발생과 동반하여 등장하게 된다. 예를 들어, 점수판 자막(C_{sco})은 골 이벤트가 발생하지 않더라도 경기 도중 시청자들에게 점수를 간헐적으로 나타내어 점수를 알려준다. 하지만 반칙을 범한 선수와 덤을 나타내는 반칙 자막(C_{ft})의 경우에는 항상 선수가 반칙을 범하는 이벤트가 발생한 뒤에 나타나게 된다.

따라서 하이라이트 추출을 위한 샷 설정은 (그림 5)와 같이 키 프레임을 기준으로 실제 이벤트가 발생한 정도만큼의 영역(β)과 다시 경기가 재개되는 정도만큼의 영역(α)을 합하여 하나의 샷으로 설정하도록 하는데, 여기서 β 는 최소한 α 이상으로 설정하여야 한다.

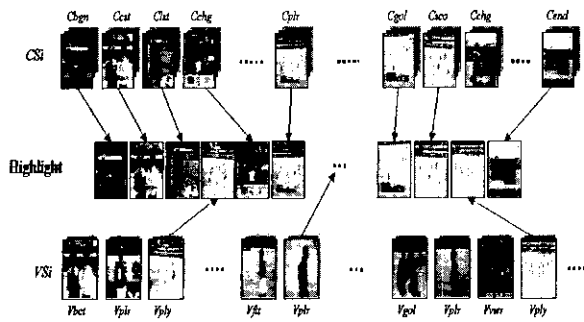


(그림 5) 키 프레임에 의한 샷 영역

5.2 하이라이트 생성

하이라이트 생성은 자막 샷 CS_i 를 중심으로 구성하고 시각적 샷 VS_i 중 일부를 중복되지 않는 범위에서 적당한 위치에 추가하여 생성하는데, 20개의 샷을 하나의 하이라이트로 설정하여 구성한다. 기본적인 자막 샷과 다른 자막 샷 및 시각적 샷과의 결합을 통한 하이라이트 생성 규칙은 다음과 같으며, (그림 6)은 하이라이트 생성의 예이다.

- (1) 기본적인 하이라이트로 구성될 샷은 C_{sco} , C_{gol} , C_{bgn} , C_{end} , C_{lst} , C_{bct} 의 6개이다.
 - ① 가장 첫 번째 샷은 C_{bgn} 과 C_{bct} 중 하나이며, 둘 중 하나가 선택되면 다른 하나는 바로 뒤에 따른다.
 - ② 위의 ①이 결정된 다음엔 C_{lst} 샷이 따른다.
 - ③ 골이 있어 득점을 한 경우에는 C_{gol} 과 C_{sco} 샷이 골 득점 수에 맞게 추가된다.
 - ④ 가장 마지막 샷은 C_{end} 와 C_{bct} 중 하나이며, 둘 중 하나가 선택되면 다른 하나는 바로 앞에 추가된다.
 - ⑤ 경우에 따라 C_{bct} 는 나타나지 않을 수 있다.
- (2) 기본적인 하이라이트가 구성된 다음에는 자막 샷 C_{chg} , C_{ft} , C_{plr} 그리고 C_{bct} 와 시각적 샷 중 V_{ply} , V_{plr} , V_{vwr} , V_{ref} 그리고 V_{bct} 를 선택적으로 시간의 흐름에 맞게 사용자가 추가하는 것을 원칙으로 한다.
- (3) 하지만 사용자는 시간의 흐름과 무관하게 추가될 샷들을 선택할 수도 있다.



(그림 6) 하이라이트 생성 예

하이라이트는 (그림 6)의 예와 같이 자막 샷과 시각적 샷이 결합되어 생성되므로 하이라이트 인덱싱 자체는 자막 샷의 기본 하이라이트 인덱스와 자막 샷의 나머지 종류 및 시각적 샷들의 인덱스를 혼합하여 구성한다.

6. 실험 및 분석

이 논문의 실험은 펜티엄 III 400MHz, Windows98 환경에서 Visual C++ 6.0 언어로 프로그래밍 하였으며, 비디오 자료는 2001.4월 펼쳐진 제3회 문화관광부 장관배 고교축구 4강 경기의 전반전을 대상으로 AVI 압축 형태의 비디오를 OSCAR II 캡처 보드로 초당 5프레임을 실험 데이터로 캡처하여, 프레임 크기를 400X300으로 정규화 하여 사용하였다.

실험에 사용된 4 경기의 전반전 평균 시간은 2852초였으며, 이들을 컬러 히스토그램, χ^2 히스토그램 그리고 이 논문에서 제안한 컬러 히스토그램과 χ^2 히스토그램을 혼합한 방법을 사용하여 추출한 시각적 키 프레임 추출 결과는 <표 3>과 같다.

<표 3> 시각적 키 프레임 추출 결과

구 분	키 프레임 추출 방법	실험 추출 수	실제 추출 수
A 경기	컬러 히스토그램	456	387
	χ^2 히스토그램	453	
	제안한 방법	442	
B 경기	컬러 히스토그램	469	372
	χ^2 히스토그램	455	
	제안한 방법	445	
C 경기	컬러 히스토그램	471	367
	χ^2 히스토그램	458	
	제안한 방법	439	
D 경기	컬러 히스토그램	458	359
	χ^2 히스토그램	451	
	제안한 방법	423	

<표 3>의 실험 결과에서 실험 추출 수는 초당 5 프레임으로 캡처한 실험 데이터를 사용하여 추출한 결과이고 실

추출 수는 캡처하지 않은 비디오 시퀀스에서 추출한 결과를 말한다. 어느 방법에서도 실제 잘못 추출된 프레임들이 다수 포함되어 있음을 알 수 있지만, 제안한 방법으로 추출한 키 프레임 수가 가장 적으며 잘못 추출된 키 프레임도 실제 가장 적음을 알 수 있다.

이와 같이 시각적인 키 프레임을 추출하는데 있어서 결과 분석표는 <표 4>와 같은데, 프레임 당 평균 계산 시간은 하나의 프레임을 처리하는데 걸리는 시간을 말하며 평균 에러 추출율은 각 방법별 4경기의 실험 추출수 평균을 실제 추출수의 평균으로 나누어 퍼센트로 나타낸 것이다. 여기서 보면 컬러 히스토그램은 프레임 당 평균 계산 시간은 가장 빠르지만 평균 에러 추출율은 가장 높고, 히스토그램은 평균 계산 시간도 매우 느리고 평균 에러 추출율도 높은 편이다. 제안한 방법은 평균 계산 시간은 컬러 히스토그램보다는 약간 느리지만 어느 정도 빠른 계산 시간을 갖고 평균 에러 추출율은 매우 낮으므로 세 가지 방법 중 가장 우수한 방법임을 알 수 있다.

<표 4> 프레임 당 평균 계산 시간

키 프레임 추출 방법	프레임 당 평균 계산시간	평균 에러 추출율
컬러 히스토그램	0.4	24.8 %
χ^2 히스토그램	1.2	22.7 %
제안한 방법	0.6	17.8 %

자막에 의한 키 프레임 추출 결과는 <표 5>와 같은데, 자막의 정확한 특성 때문에 실제 오류는 발생하지 않았으며, 단지 방송사별 또는 같은 방송사더라도 모양과 형태 및 위치가 서로 다른 경우가 발생하는 경우에는 이들을 다시 설정해 두어야 하는 문제점이 존재한다.

<표 5> 자막 키 프레임 추출 결과

구 분	자막에 의한 추출 수
A 경기	53
B 경기	47
C 경기	57
D 경기	45

<표 3>과 <표 5>의 결과에서 나타난 것처럼 시각정보에 의해 추출된 샷의 길이보다 자막에 의해 추출된 샷의 길이가 비교적 크게 추출되며, 자막에 의해 추출된 전체 샷 자체도 실제 비디오보다 함축된 하나의 하이라이트라고 볼 수 있다.

실제 데이터와 실험을 통하여 생성된 하이라이트의 평균 재현 시간은 <표 6>과 같다. 이 결과는 자막 샷과 시각적 샷을 합성하여 하이라이트를 20개의 샷으로 제한하여 생성한 결과지만, 사용자는 샷의 수를 임의로 조정하는 것도 가능하다.

〈표 6〉 평균 재현 시간

구 분	실제 데이터	하이라이트	생성율
평균 재현 시간	2852초	327초	11.4%

여기서 생성율은 실제 데이터에서 하이라이트를 생성하여 함축된 정도를 나타내는 것으로, 하이라이트 샷의 수 설정과 시각적 샷 및 자막 샷의 수에 따라 다르게 나타난다.

결과에서 알 수 있는 것처럼 자막을 이용한 키 프레임 설정 방법은 하나의 비디오에서 갑작스런 자막 위치, 크기 그리고 컬러 정보가 변하지 않는다면 100% 정확히 추출이 가능하다. 하지만 시각정보를 이용한 방법에서는 프레임간 변화가 크더라도 거의 동일한 컬러 정보를 갖고 있다면 키 프레임을 추출하는데 오류를 동반하게 된다. 따라서 자막 키 프레임과 시각적 키 프레임을 합성하여 하이라이트를 생성하는 방법은 보다 더 효율적임을 알 수 있다.

7. 결론 및 향후 연구과제

이 논문에서는 축구 비디오 데이터에서 중요한 이벤트를 포함하는 하이라이트를 생성하는 새로운 방법 제안하였는데, 하이라이트 생성은 시각적인 정보와 자막에 의한 정보를 바탕으로 이벤트를 중심으로 하여 하이라이트 생성 규칙에 따라 생성하였다.

컬러 히스토그램과 χ^2 히스토그램을 합성한 방법을 제안하여 시각적인 키 프레임을 보다 더 효율적으로 추출하였으며, 자막이 갖는 특성을 바탕으로 문자 인식 과정을 거치지 않고도 지역적 차영상을 통한 유사성 측정으로 정확하게 자막의 키 프레임을 추출하였다.

이 논문은 거의 비슷하게 반복되는 부분이 많은 긴 스포츠 비디오를 중요한 부분만을 추출하여 한눈에 시청 할 수 있도록 하는 새로운 방법으로서 기대효과가 크지만, 실제 비디오상의 시각적인 변화에 잘 적응하는 효율적인 방법에 대한 연구와 자동 인덱싱 방법에 대한 연구가 계속 수행되어야 할 것으로 판단된다.

참 고 문 헌

[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 through 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 Abstract-

ting," *ACM Communication*, Vol.40, No.12, pp.55-62, 1997.
 [4] L. He, E. Sanocki, A. Gupta and J. Grudin, "Auto-Summari- zation of Audio-Video Presentations," *Proceedings of ACM Multimedia '99*, pp.489-498, 1999.
 [5] Sun, X., Kankanhalli, M., Zhu, Y. & Wu, J., "Content-Based Representative Frame Extraction for Digital Video," *Inter- national Conference on Multimedia Computing and Sys- tems*, pp.190-193, 1998.
 [6] Smith, M. A. & Kanade, T., "Video Skimming for Quick Browsing based on Audio and Image Characterization," *Technical Report No.CMU-CS-95-186*, School of Com- puter Science, Carnegie Mellon University, 1995.
 [7] Zhang, H. J., Low, C. Y. & Smoliar, S. W., "Video Parsing and Browsing using Compressed Data," *Multimedia tools and Applications 1*, pp.89-111, 1995.
 [8] W. Chang, G. Sheikholeslami, J. Wang and A. Zhang, "Data Resource Selection in Distributed Visual Information Sys- tems," *IEEE Trans. on Knowledge and Data Engineering*, Vol.10, No.6, pp.926-946, 1998.
 [9] M. Smith and T. Janade, "Video Skimming for Quick Brow- sing based on Audio and Image Characterization," *Tech. Report CMU-CS-95-186*, Computer Science Department, Carnegie Mellon University, July, 1995.
 [10] 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.
 [11] N. Vasconcelos and A. Lippman, "A Spatiotemporal Motion Model for Video Summarization," *CVPR*, Santa Barbara, 1998.
 [12] G. Medioni, R. Nevatia and I. Cohen, "Event detection and Analysis from Video Streams," *DARPA98*, pp.63-72, 1998.
 [13] R. Rosales and S. Sclaroff, "3D Trajectory for Tracking Multiple Objects and Trajectory Guided Recognition of Actions," *CVPR*, June, 1999.
 [14] D. Pfoser and Y. Theodoridis, "Generating Semantics-Based Trajectories of Moving Objects," *International Workshop on Emerging Technologies for Geo-Based Applications*, Ascona, Switzerland, 2000.
 [15] Hampapur, A., R. Jain and T. Weymouth, "Digital Video Indexing in Multimedia Systems," *In Proc. of AAAI-94 Workshop on Indexing and reuse in Multimedia Systems*, 1994.
 [16] Hampapur, A., R. Jain and T. Weymouth, "Production Model Based Digital Video Segmentation," *Multimedia Tools and Applications*, Vol.1, No.1, pp.9-46, 1995.
 [17] E. Ardizzone, M. L. Cascia, "Automatic Video Database In- dexing and Retrieval," *Multimedia Tools and Applications*,

Vol.4, No.1, pp.29-56, 1997.

- [18] J. C. Lee, Q. Li, W. Xiong, "VIMS : A Video Information Management System," *Multimedia Tools and Applications*, Vol.4, No.1, pp.7-28, 1997.
- [19] B. Furht, S. W. Smoliar, H. J. Zhang, "Video and Image Processing in Multimedia System," *Kluwer Academic Publishers*, pp.335-356, 1995.
- [20] S. J. Dennis, R. Kasturi, U. Gargi, S. Antani, "An Evaluating of Color Histogram Based Methods in Video Indexing," *Research Progress Report CSE-96-053 for the contract MDA 904-95-C 2263*, 1995.
- [21] Zhang, H. J., J. Y. A. Wang and Y. Altunbasak, "Content-based Video Retrieval and Compression : A Unified Solution," *Proc. ICIP '97, Int. Conf. on Image Processing*, pp. 113-16, 1997.
- [22] Zhong, D. and S. F. Chang, "Spatio-Temporal Video Search Using the Object Based Representation," *Proc. ICIP '97, Int. Conf. on Image Processing*, pp.121-24, 1997.
- [23] Zabih, R., J. Miller and K. Mai, "Feature-based Algorithms for Detecting and Classifying Scene Breaks," *Proc. ACM Int. Conf. on Multimedia*, pp.189-200, 1995.
- [24] Armen, F., A. Hsu and M. Y. Chiu, "Feature Management for Large Video Databases," *In Storage and Retrieval for Image and Video Databases, Proc. SPIE, 1908* : pp.2-12, 1993.



전 근 환

e-mail : khjeon@kunjang.ac.kr

1993년 군산대학교 전산학과 졸업
 1995년 충북대학교 대학원 전산학과 (이학석사)
 1997년 충북대학교 대학원 전산학과 박사 과정(수료)

1997년~현재 군장대학 컴퓨터응용학부 조교수
 관심분야 : 시간지원 데이터베이스, 전문가 시스템, 멀티미디어 데이터베이스 등



신 성 윤

e-mail : syshin@cs.kunsan.ac.kr

1994년 군산대학교 전산학과 졸업
 1997년 군산대학교 대학원 정보통신공학과 (공학석사)
 2000년~현재 군산대학교 대학원 컴퓨터 과학과 박사과정

1998년~현재 군장대학 컴퓨터응용학부 겸임교수
 관심분야 : 컴퓨터비전, 패턴인식, 인공지능, 멀티미디어 데이터 베이스 등



이 양 원

e-mail : ywrhee@cs.kunsan.ac.kr

1979년 숭실대학교 전산학과 졸업
 1983년 연세대학교 산업대학원 전산전공 (공학석사)
 1994년 숭실대학교 전자계산학과(공학박사)
 1979년~1986년 국방 연구원

1986년~현재 군산대학교 컴퓨터정보과학과 교수
 관심분야 : 영상 데이터베이스, 패턴인식, 인공지능, 멀티미디어 데이터베이스 등



류 근 호

e-mail : khryu@dbtab.chungbuk.ac.kr

1976년 숭실대학교 전산학과 졸업
 1980년 연세대학교 산업대학원 전산전공 (공학석사)
 1988년 연세대학교 대학원 전산전공 (공학박사)

1976년~1986년 육군군수지원사전산실(ROTC장교), 한국전자통신 연구소(연구원), 한국방송통신대 전산학과(조교수) 근무
 1989년~1991년 Univ. of Arizona 연구원(TempIS Project)
 1986년~현재 충북대학교 전기전자컴퓨터공학부 교수
 관심분야 : 시간데이터베이스, 시공간데이터베이스, 객체 및 지식 기반 시스템, Temporal GIS, 지식기반 정보검색시스템, 데이터 마이닝, 데이터베이스 보안 및 Bio-Informatics.