

# 축구 비디오 인덱싱을 위한 장면 전환 검출과 시각 정보 분석

신성윤, 강오형, 문 경, 이양원  
군산대학교 컴퓨터정보과학과

## Scene Change Detection and Visual Information Analysis for Soccer Video Indexing

Seong-Yoon Shin, Oh-Hyong Kang, Kyung Moon, Yang-Won Rhee  
Dept. of Computer Information Science, Kunsan Nat'l University  
E-mail : syshin@cs.kunsan.ac.kr, ywrhee@kunsan.ac.kr

### 요 약

비디오 데이터를 인덱싱 하기 위해서는 우선적으로 장면 전환을 검출하여 키 프레임을 추출하고 추출된 키 프레임을 바탕으로 인덱싱 작업을 수행한다. 본 논문에서는 장면 전환을 검출하기 위하여 컬러 히스토그램과  $\chi^2$  히스토그램을 합성한 방법을 이용하여 키 프레임을 추출하고, 축구 비디오가 갖는 특성을 이용하여 샷 사이의 흐름을 파악하여 시각 정보를 분석하며, 이를 바탕으로 축구 비디오를 다양한 방법으로 인덱싱하는 방법을 제시한다.

### 1. 서론

비디오 데이터(video data)는 시각적인 내용과 청각적인 내용, 그리고 텍스트 스트림(text stream)들과 같은 다양한 정보들이 일정한 방법으로 동기화 되어 사용자들에게 제공되는 것으로서 멀티미디어 시스템(multimedia system)에서 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 위한 보다 효율적인 방법들이 개발되어야 할 필요성이 점점 증대되고 있다.

비디오를 시간의 흐름에 따라 중요한 정보를 포함하는 샷(shot)으로 분류하는 장면 전환 검출 및 키 프레임(key frame) 추출 작업은 비디오를 다양한 형태로 인덱싱(indexing) 하기 위하여 가장 먼저 이 수행되어야 할 작업이다.[2,3,4]. 이렇게 생성된 샷들은 비디오 제작자나 편집자 및 일반 사용자들이 쉽게 검색하고 간편하게 편집하며 재생하여 볼 수 있도록 논리적으로 인덱싱(indexing)되고 브라우징(browsing)이 가능하도록 구축되어야 한다.

효율적인 비디오의 인덱싱을 위하여 장면 전환 검출에 의해 분류된 각각의 샷들은 대표 프레임(representative frame)[5] 또는 키 프레임(key frame)[6,7]이라고 하는 샷을 대표하는 하나의 중요한 프레임을 갖게된다.

본 논문에서는 먼저, 장면 전환을 검출하기 위하여 컬러 히스토그램(color histogram)과  $\chi^2$  히스토그램을 합성한 방법으로 시각적인 키 프레임을 추출한다. 이렇게 추출된 키 프레임을 분석하여 축구 비디오가 갖는 전반적인 샷의 흐름을 파악하도록 한다. 또한 추출된 키 프레임을 바탕으로 물리적으로 인덱싱을 수행하며, 샷이 갖는 특성을 바탕으로 다양한 분야별로 논리적으로 인덱싱을 수행하도록 한다.

전체적인 시스템 구조는 그림 1과 같으며 2장에서는 장면 전환 검출에 의한 시각적 키 프레임 추출을 설명하고, 3장에서는 추출된 키 프레임을 바탕으로 시각적 샷을 분석하고 샷의 흐름을 파악한다, 그리고 4장에서는 분석된 샷들에 대한 키 프레임 인덱싱에 대해서 설명하고, 5장에서는 실험을 통한 결과를 분석하며 6장에서 결론을 맺는다.

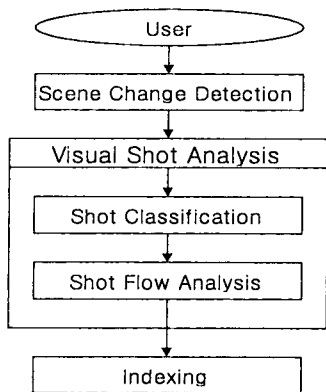


그림 1. 시스템 구조

## 2. 장면 전환 검출

장면 전환 검출은 긴 비디오 스트림을 일정한 기준에 따라 작은 크기의 샷들로 나누고 각각의 샷들을 대표하는 키 프레임들을 추출하는 것이다. 대부분의 시스템에서 장면 전환이 크게 발생하는 지점의 첫 번째 프레임들을 키 프레임으로 설정하여 이들 키 프레임들을 기준으로 비디오를 인덱싱 하여 비디오 편집, 검색 또는 브라우징 등에 유용하게 사용한다.

장면 전환 검출에 의해 장면을 여러 개로 분할하고 키 프레임들을 추출하기 위한 방법에는 입력되는 연속적인 두 프레임간의 차이를 이용하는 방법[9,10], 프레임 사이의 컬러 히스토그램 차이값을 이용하는 방법[11,12,13,14], 프레임 내의 객체 이동을 계산하여 추출하는 방법[15,16], 프레임 에지(edge)를 추출하여 에지 변화를 비교하여 추출하여 이용하는 방법[17] 그리고 전체적인 웨이블릿(wavelet)을 측정하여 이용하는 방법[18] 등 많은 방법들이 이용되고 있다.

가장 보편적으로 사용되는 장면 전환 검출 방법은 식 (1)과 같은 컬러히스토그램의 차이값을 계산하여 사용하는 방법이다[11,12,13,14]. 여기서  $d(I_i, I_{i+1})$ 은 이웃하는 두 프레임 ( $I_i, I_{i+1}$ ) 사이의 유사성을 측정하는 함수이고  $H_i(j)$ 는  $i$ 번째 프레임의  $j$ 번째 히스토그램 값을 나타낸다.

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

일반적인 컬러 히스토그램은 식 (1)의 방법에 따라 사용되지만 좀 더 나은 검출을 위하여 식 (2)와 같이

RGB 컬러 공간대를 따로 나누어 계산하는 방법도 많이 사용하고 있다.

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

성능이 우수한 다른 방법으로는  $\chi^2$  히스토그램 방법을 많이 사용하는데, 이 방법은 컬러 히스토그램이나 템플릿 매칭(template matching) 방법보다 좋은 결과를 나타내기 때문에 장면 전환 검출의 많은 분야에서 사용하고 있다[14].

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

식 (3)에서  $H_i(k)$ 는  $i$ 번째 프레임의 히스토그램에서  $k$ 번째 빈(bin)의 값을 의미한다.

위의 식 (3)은 이용자에 따라 다음의 식 (4)와 같이 변형되어 사용될 수도 있다.

$$d(I_i, I_j) = \sum_{k=1}^n \sqrt{H_{i(k)}^2 - H_{j(k)}^2} \quad (4)$$

본 논문에서 이용하는 장면 전환에 의한 키 프레임 추출 방법은 RGB를 따로 계산하는 컬러 히스토그램과  $\chi^2$  히스토그램의 장점을 합성한 식 (5)의 방법을 이용하는데, 이 방법은 각 프레임 사이의 차이 값을 계산하고 차이 값의 크기에 따라 키 프레임들을 추출하여 샷들을 구분하는 계산 방법이다.

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

### 3. 샷의 분류 및 흐름 분석

장면 전환 검출에 의해 추출된 시각적 샷들은 실제 장면 변화가 큰 부분에서 추출되므로 상당히 많은 수의 샷들이 존재하게 된다. 많은 수의 샷들을 일정한 기준이 없이 서로 다르게 분류하는 것은 상당히 복잡하므로, 본 연구에서는 축구 비디오의 특성을 이용하여 각각의 샷들을 크게 표 1과 같이 9가지 형태로 제한하여 분류하였다.

표 1. 시각정보에 의한 샷의 분류

구분	종류	특징	설명
VS <sub>i</sub> (Visual Shot)	<i>V<sub>ply</sub></i>	연속 경기 (play)	운동장에서의 경기 장면
	<i>V<sub>plr</sub></i>	선수 초점 (player)	특정 선수 확대 장면
	<i>V<sub>flt</sub></i>	반칙 (fault)	반칙 선수 확대 장면
	<i>V<sub>chg</sub></i>	선수 교체 (change)	선수교체 확대 장면
	<i>V<sub>gol</sub></i>	골 (goal)	골 넣은 선수 확대 장면
	<i>V<sub>vr</sub></i>	관람석 (viewer)	관람석 장면
	<i>V<sub>bct</sub></i>	경기 중계 (broadcasting)	캐스터와 해설자 장면
	<i>V<sub>bch</sub></i>	벤치 (bench)	벤치(감독, 코치, 선수 등) 장면
	<i>V<sub>ref</sub></i>	심판 (referee)	심판 확대 장면

표 1과 같이 분류된 샷들 사이의 관계를 보면 그림 2와 같이 샷 사이의 시간적 흐름에 따른 관계와 구성을 알 수 있다.

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

예를 들어, 축구 경기를 진행하던 도중 어느 한 팀이 골을 넣은 경우(그림 2c)에는 연속 경기 샷과 골 샷 다음에 나타나는 샷은 골을 넣은 선수, 벤치 또는 관람석 중 어떠한 하나의 샷이 나타날 수 있으며, 이 세 샷들은 순서가 정해져 있지 않고 서로 임의의 순서대로 등장할 수 있으며 마지막에는 다시 경기 재개하여 연속 경기를 수행하게 된다.

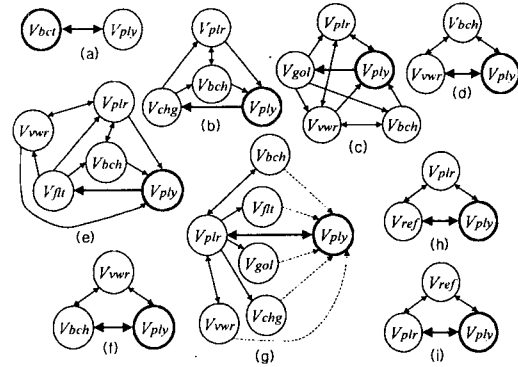


그림 2. 시각적 샷 사이의 흐름도

### 4. 키 프레임 인덱싱

장면 전환 검출에 의하여 장면 변화가 큰 프레임들 중심으로 구성된 샷들의 키 프레임들은 추출된 순서에 따라 자동적으로 일련번호에 의해 물리적으로 인덱싱 된다.

스포츠 비디오에서 하나의 축구 비디오 스트림을 *V*라 하면 추출된 키 프레임들을 바탕으로 한 인덱싱은 식 (6)과 같다. 여기서 *V*는 각 샷들의 키 프레임 *VS<sub>i</sub>*들로 인덱싱 되어 구성되며, 이 *VS<sub>i</sub>*는 샷을 구성하는 각각의 프레임 *VF<sub>j</sub>*들로 구성된다.

여기서 샷은 식 (1)에서 제안한 장면 전환 검출 방법에 따라 장면 변화가 큰 키 프레임과 이에 종속된 프레임들의 집합을 말하며, 프레임은 이 샷들을 구성하는 최소 단위이다.

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

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

또한, 추출된 샷들에 대하여 분야별 또는 사건별로 일정한 의미를 부여하여 논리적인 구조로 인덱싱이 가능하다. 논리적인 인덱싱은 사용자가 원하는 다양한 형태로 구성이 가능한데, 물리적으로 인덱싱 된 키 프레임의 번호를 연결하여 스포츠 비디오의 어떤 장면이라도 직접 접근이 가능하도록 구성할 수 있다. 추출된 샷들의 물리적/논리적 인덱싱의 구성 형태는 그림 3과 같다.

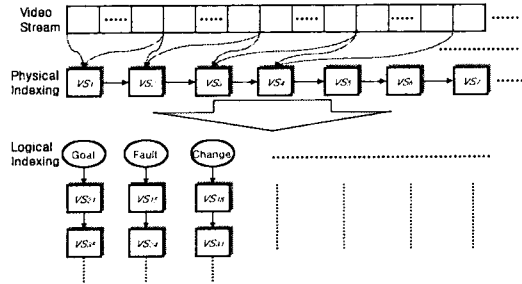


그림 3. 샷의 물리적/논리적 인덱싱 구조

### 5. 실험 및 분석

본 논문의 실험은 펜티엄 III 400MHz PC와 Windows98 환경에서 Visual C++ 6.0 언어로 프로그래밍 하였다.

비디오 자료는 2001.4월 펼쳐진 제3회 문화관광부장관배 고교축구 4 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 2852 초였으며, 이들을 컬러 히스토그램,  $\chi^2$  히스토그램 그리고 본 논문에서 제안한 컬러 히스토그램과  $\chi^2$  히스토그램을 혼합한 방법을 사용하여 추출한 시각적 키 프레임 추출 결과는 <표 3>과 같다.

표 2. 방법별 키 프레임 추출 결과

구분	키 프레임 추출 방법	추출 수
A경기	컬러 히스토그램	456
	$\chi^2$ 히스토그램	453
	제안한 방법	442
B경기	컬러 히스토그램	469
	$\chi^2$ 히스토그램	455
	제안한 방법	445
C경기	컬러 히스토그램	471
	$\chi^2$ 히스토그램	458
	제안한 방법	439
D경기	컬러 히스토그램	458
	$\chi^2$ 히스토그램	451
	제안한 방법	423

표 2의 실험 결과에서 추출 수는 초당 5 프레임으로 캡처한 실험 데이터를 사용하여 추출한 결과이다.

또한, 이와 같이 키 프레임을 추출하는데 필요한 프레임 당 평균 계산 시간은 표 3과 같은데, 여기서 보면 본

연구에 이용한 방법이 키 프레임 추출 결과도 어느 정도 우수 하지만 수행 시간도 비교적 빠르다고 볼 수 있다.

표 3. 프레임 당 평균 계산 시간

키 프레임 추출 방법	계산 시간(sec/frame)
컬러 히스토그램	0.4
$\chi^2$ 히스토그램	1.2
제안한 방법	0.6

### 6. 결론

본 논문에서는 축구 비디오에서 장면 전환 검출에 따라 샷을 분할하고 각 샷들의 키 프레임을 추출하며, 이들을 분석하고 흐름을 파악하여 다양한 형태로 인덱싱하는 방법을 제시하였다.

컬러 히스토그램과  $\chi^2$  히스토그램을 혼합한 방법을 이용하여 키 프레임을 보다 더 효율적으로 추출하였으며, 축구 비디오가 갖는 특징과 흐름을 이용하여 물리적/논리적으로 인덱싱을 수행하였다.

앞으로 장면 전환에 잘 적응하는 보다 효율적인 장면 전환 검출 방법에 대한 연구가 더욱더 진행되어야 할 것으로 판단된다.

### [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] Sun, X., Kankanhalli, M., Zhu, Y. & Wu, J., "Content-Based Representative Frame Extraction for Digital Video", *International Conference on Multimedia Computing and Systems*, 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 Computer 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
- [9] 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
- [10] 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.
- [11] E. Ardizzone, M. L. Cascia, "Automatic Video Database Indexing and Retrieval", *Multimedia Tools and Applications*, Vol. 4, No. 1, pp. 29-56, 1997.
- [12] 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
- [13] B. Furht, S. W. Smoliar, H. J. Zhang, "Video and Image Processing in Multimedia System", *Kluwer Academic Publishers*, pp. 335-356, 1995
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.