

실시간 필터링을 위한 축구 비디오의 장면 검출

조준호^a, 진성호^a, 노용만^a, 강경옥^b

E-mail: chojunho@icu.ac.kr

^a한국정보통신대학교 공학부, 멀티미디어그룹

^b한국전자통신연구원, 전파방송연구소

Detection Algorithm of Shoot Scenes in Soccer Video for Real-time Content Filtering

Jun Ho Cho^a, Sung Ho Jin^a, Yong Man Ro^a and Kyeongok Kang^b

^aMultimedia Group, Information and Communications University (ICU)

^bElectronics and Telecommunications Research Institute, Korea (ETRI)

요약

축구 비디오의 슛 장면과 같은 이벤트 검출에는 많은 연구가 있었으나 기존의 방법은 정확성과 실시간으로 필터링하는 부분에서는 문제가 있다. 이에 대해서 본 논문은 현재 실시간으로 가장 많이 방송되는 스포츠 장르 중 축구 비디오에 대한 효율적인 슛 장면 검출 알고리즘을 제안하고자 한다. 제안한 방법은 우선 입력 프레임으로부터 뷰 타입을 결정한다. 슛 장면에 대한 뷰 타입 패턴 변화, 리 플레이의 검출과 오디오 파워의 변화를 이용하여 필터링을 수행한다. 실험 결과를 통하여 제안한 방법이 효율적이고 효과적으로 슛 장면을 검출하였고 다채널의 디지털 방송환경에서 실시간으로 콘텐츠를 필터링하는 서비스가 가능함을 보였다.

1. 서론

다가올 디지털 방송환경은 방대한 양의 프로그램을 시청자에게 제공함으로써, 특정 시간대에 선택할 수 있는 프로그램의 수가 증가하는 양적인 발전을 한다. 그러나 다채널 다매체 환경에서 특정 시간대에 하나의 프로그램만을 선택해서 시청하는 것은 비효율적이며 동시에 여러 채널의 정보를 시청할 수 있는 질적인 발전이 요구된다. 이에 대해 실시간 방송 필터링은 시청자가 한 채널에서 방영되고 있는 선호 프로그램을 시청할 때, 타 채널들에서 실행 중계로 방송되고 있는 프로그램 중 시청자가 선호하는 장면들에 대한 정보를 실시간으로 시청자에게 제공할 수 있다. 일반적으로 실시간으로 중계되는 콘텐츠들은 콘텐츠 제공자가 해당 메타데이터를 제작할 시간적인 여유가 없다. 따라서, TV 단말에서 직접적으로 해당되는 콘텐츠에 대한 분석 알고리즘이 필요하다.

축구 비디오는 가장 많이 실시간으로 방송되고 있는 스포츠 장르 중 하나이며 시청자가 두 시간의 경기 동안 보고 싶어하는 부분은 한정되어 있다. 그래서 축구 비디오에 대해서 특정 이벤트와 같은 시청자가 관심 있는 부분만을 검출해주는 연구가 많이 있었다. 그러나 기존의 축구 비디오 분석방법은 이벤트에 대한 요약 서비스를 위한 것으로서 실시간으로 슛 장면을 검출함에 있어서 문제가 있다. Assfalg는 은닉 마코프 모델을 이용하여 프리킥과 페널티 킥 등 축구경기의 하이라이트를 검출하는 방법이 제안하였다[1]. Chen은 비디오와 오디오 특징 값들을 이용하여 골 이벤트를 검출하는 방법을 제

시하였다[2]. 그러나 제안된 방법들은 전처리 과정을 필요로 하여 실시간으로 이벤트를 검출하지 못하는 문제가 있고 또한 실시간에 근접한 필터링이 가능하더라도 슛 장면 검출에 대해서 정확도가 낮은 문제점이 있다.

본 논문에서는 검출하고자 하는 슛 장면에 대해서 그림 1과 같이 모델링한다. 슛 장면은 골대 근처에서 발생하므로 먼저 골대 영역이 포함된 글로벌 뷰가 나타나고, 이어서 선수나 코치, 또는 관중 등 인물을 중심으로 표현되는 클로즈업 뷰가 나타나는 패턴이 발생한다. 클로즈업 뷰에 이어서 같은 장면을 다른 각도의 카메라로 촬영한 리플레이 샷이 나타나면서 해당 슛 장면에 대한 비디오 패턴은 끝나게 된다. 한편 오디오 프레임은 비디오 프레임중 글로벌 뷰에서 나타나는 슛 장면에 대해 관중의 함성소리나 해설자의 흥분된 소리로 인해 오디오 샘플값의 파워가 증가하는 패턴이 발생한다.

이렇게 오디오와 비디오 프레임으로부터 시간에 따른 패턴을 모델링을 하므로써 실시간 또는 실시간에 근접하게 슛 장면을 검출할 수 있게 된다.

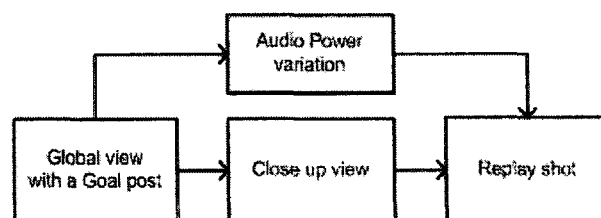


그림 1. 슛 장면 검출 블록도

본 논문의 구성은 II에서 슛 장면의 모델링을 바탕으로 특징 값들을 추출하는 알고리즘을 제안하고 III에서 제안한 방법의 유효성을 검증하기 위한 실험 결과를 제시하며, IV에서 결론 및 향후 연구에 대해 설명한다.

II. 특징 값 추출

스� 장면을 검출하기 위한 특징 값 추출을 위하여 우선 입력 비디오 프레임으로부터 4가지의 뷰 타입을 결정하는 알고리즘을 제시하고, 리플레이를 검출하는 알고리즘을 설명한다. 마지막으로 오디오 프레임으로부터 슛 장면에 대한 오디오 파워의 변화를 분석하고 특징 값으로부터 검출의 정확도를 높이는 알고리즘을 제시한다.

1. 뷰(view) 타입 결정 알고리즘

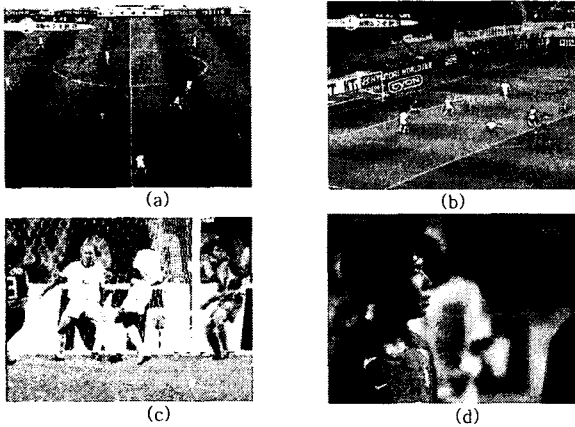


그림 2. 뷰(view) 타입 종류

(a) 골대가 없는 글로벌 뷰, (b) 골대가 있는 글로벌 뷰, (c) 미디엄 뷰, (d) 클로즈업 뷰

프레임의 타입들은 글로벌 뷰(global view), 미디엄 뷰(medium view)와 클로즈업 뷰(close-up view)로 나누어 진다. 글로벌 뷰는 스포츠 게임에서 보여주는 줌아웃(zoom-out)된 장면으로 운동장을 기준으로 하여 표현되며, 전체의 흐름을 판단할 수 있게 보여주는 이미지들을 의미하며, 미디엄 뷰는 특정 이벤트에 초점을 맞추어 찍어내는 이미지들로 선수들의 활동에 초점을 맞추는 이미지들이다. 클로즈업 뷰는 선수들, 코치, 심판 등의 얼굴을 중심으로 줌인(zoom-in)하여 가깝게 보여주는 이미지들이다.

그림 3은 비디오 프레임으로부터 4가지의 뷰 타입을 분류하는 알고리즘을 나타낸다. 디코딩된 비디오 프레임으로부터 RGB 도메인을 HIS 도메인으로 변환하여 칼라 맵(color map)정보를 계산한 후 지역화된 칼라 맵 정보를 이용하여 클로즈업 뷰를 검출할 수 있다. 다른 뷰들은 소벨(sobel) 마스크(mask)를 이용한 에지 추출과 thinning 과정을 거친 후 호프 변환(Hough transform)의 각도정보에서 추출된 대각선의 수가 3개 이상이면 골대가 포함된 글로벌 뷰로 판단하고 수직선으로 존재하는 중앙선을 이용하여 골대가 포함되지 않는 글로벌 뷰와 미디엄 뷰를 분류하게 된다.

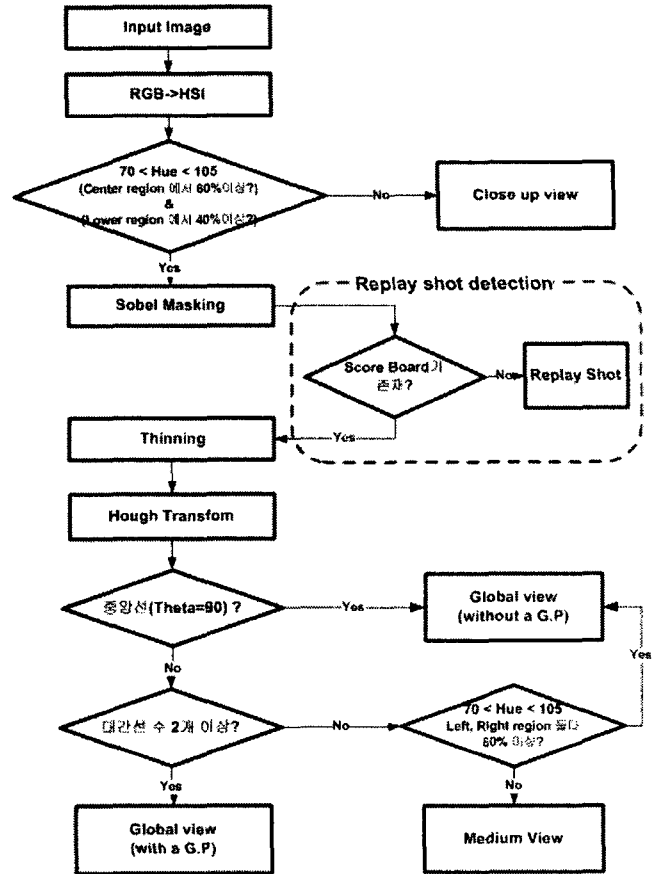


그림 3. 뷰(view) 타입 검출 알고리즘

2. 리플레이 샷(Replay shot) 검출

리플레이 샷은 축구 경기에서 슛 장면이나 파울 장면과 같은 이벤트 후에 발생한다. 제안하는 리플레이 샷의 검출방법은 리플레이 도중에는 경기 중에 항상 존재하는 점수표시 영역이 사라진다는 것을 이용하였다. 리플레이 샷을 검출하기 위해서, 우선 점수표시 영역을 결정하는 것이 필요하다. 점수표시 영역을 검출하기 위해서 경기가 시작되고 5초 동안 입력되는 비디오 프레임에서 에지를 추출하여 같은 위치에 점수표시 영역이 존재하는 것을 이용하였다[5].

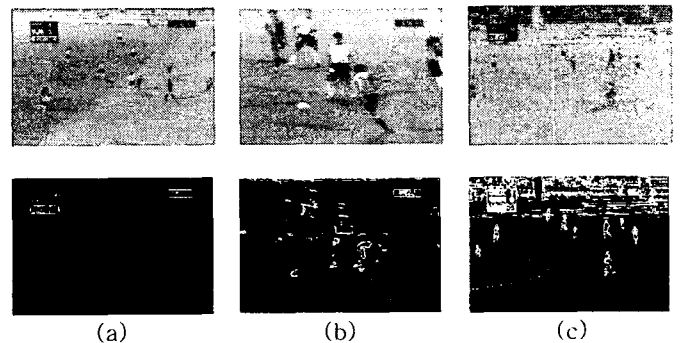


그림 4. Replay shot 검출 이미지
(a) 점수영역 결정 이미지, (b) 리플레이 샷 이미지, (c) 리플레이 샷이 아닌 이미지

동일한 위치에 존재하는 점수표시 영역에서 추출된 에너지로부터 이진화된 밝기 값은 다른 영역 보다 누적되는 속도가 빠르므로 5초 후에 총 누적된 값의 80% 이상인 부분을 필터링 하게 되면 그림 4.(a)와 같이 점수표시 영역에서 높은 응답이 나타나는 것을 확인할 수 있다. 5초 이후에는 결정된 점수표시 영역을 기준으로 입력 프레임을 비교하여 60%이상 일치하지 않으면 (b)와 같이 점수표시 영역이 사라졌다고 판단하여 리플레이 샷으로 판별하게 된다. (c)와 같이 점수표시 영역이 (a)와 60% 이상 일치한다면 점수영역이 존재하므로 리플레이 샷이 아닌 것으로 판단하게 된다.

리플레이 샷 검출 알고리즘은 독립된 모듈이 아닌 뷰 타입 검출 알고리즘에 포함된다. 그러므로 입력되는 비디오 프레임이 글로벌 뷰와 미디엄 뷰를 판별하기 전에 리플레이 샷인지 판별하게 하고 리플레이 샷이라고 판별 될 경우 글로벌 뷰나 미디엄 뷰를 결정 할 필요가 없게 된다. 기존의 연구[6]에서 개선된 점은 리플레이 샷 판별 알고리즘을 추가함으로써 미디엄 뷰와 글로벌 뷰의 발생확률이 줄어들게 됨으로써, 뷰 검출에서 가장 많은 시간이 걸리는 호프 변환의 횟수가 줄어들게 되어 좀 더 실시간으로 처리가 가능하게 된다는 것이다.

3. 오디오 파워(Audio Power)

비디오 프레임에서 점수표시 영역을 추출할 때와 마찬가지로 경기 시작 후 5초 동안에 1초를 한 주기로 하여 오디오 샘플 값들의 파워를 평균한 값을 임계 값으로 정한다. 이 후에 입력되는 오디오 프레임으로부터 1초의 주기로 구한 파워의 평균 값을 임계 값과 비교되고 그 이상의 값이 검출되면 관중의 함성이나 해설자의 흥분된 소리가 있다고 판단하게 된다. 그림 5는 오디오 프레임으로부터 추출된 샘플 값의 크기를 나타낸다. 타원으로 표시된 영역이 실제 슷 장면이 발생한 부분이고 그 영역 내에 오디오 평균 파워 값이 크게 나타나는 것을 확인할 수 있다. 이렇게 검출된 프레임은 비디오 특징 값으로부터 얻어진 슷 장면 검출에 대해서 좀 더 높은 정확도를 얻을 수 있다.

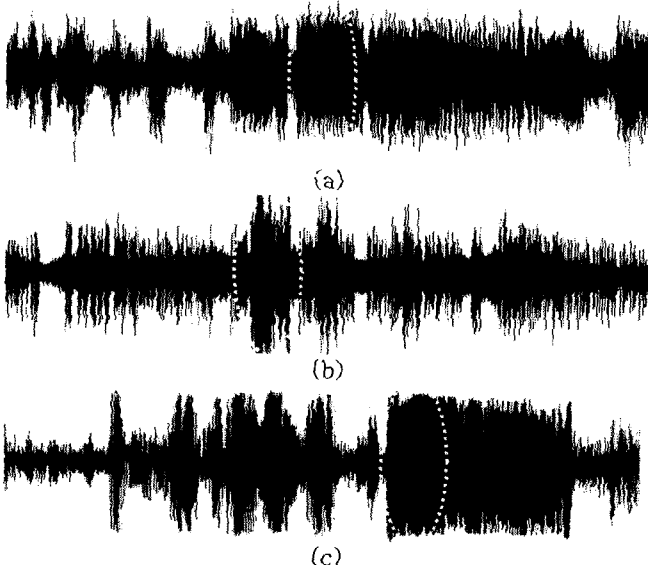


그림 5. 오디오 프레임으로부터 추출된 샘플 크기

V. 실험 결과

제안하는 방법의 유효성을 검증하기 위해서, 본 논문의 실험에서는 하나의 채널의 비디오를 플레이 하면서, 동시에 다른 채널로부터 비디오를 전송 받아서 콘텐츠 필터링 작업을 수행하였다. 실험은 2가지 형태로 나누어서 실시하였다. UDP를 통해서 전송되는 비디오에 대해서, 하나는 PC에서 필터링 알고리즘 자체를 실행하는 것이고, 다른 하나는 하나의 비디오를 플레이 하면서 다른 비디오에 대해 필터링 알고리즘을 수행하는 것이다.

슷 장면 검출 결과 평가를 위해서 수식 (1)의 recall, precision measure가 이용되었다. 수식 (1)에서 N_c 는 올바른 검출 결과의 개수, N_m 는 검출하지 못한 결과의 개수, N_f 는 잘못 검출된 결과의 개수를 나타낸다.

① 펜티엄 IV, 3 GHz CPU, 2GByte 메모리, ② 펜티엄 IV, 1.7 GHz CPU, 512MByte 메모리, ③ 펜티엄 III, 650 MHz CPU, 128MByte 메모리 사양을 가지는 3대의 PC들을 이용하여 실험을 수행하였다. 실험에 사용된 축구 비디오는 2006년 독일 월드컵 예선전, 2004 그리스 올림픽, 2002년 한일 월드컵 등 전 후반 90분 길이의 서로 다른 축구 비디오 3개의 영상으로 352x240의 MPEG-1 비디오로 인코딩 되었다.

$$precision = \frac{N_c}{N_c + N_f} \times 100\% \quad (1)$$

$$recall = \frac{N_c}{N_c + N_m} \times 100\%$$

샘플링 되는 프레임을 그림 3에서 제안하는 뷰 판별 알고리즘에 의해서 결정되는 성능은 그림 6과 같다. 뷰를 결정하는 리콜율은 평균 93.5%이고 정확도는 평균 약 89.5%의 성능을 보였다.

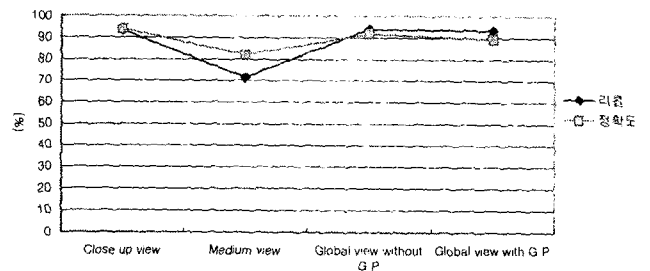


그림 6. 뷰 검출 결과

각각의 뷰 타입을 결정하는 데 걸리는 시간은 그림 7과 같다. 제안하는 콘텐츠 필터링하는 소요되는 시간은 대부분 샘플링된 프레임의 뷰 타입을 결정하는 데 걸리는 시간이다. 그림 7은 서로 다른 사양을 가진 PC 3대를 이용하여, 앞서 설명된 2가지 형태의 실험을 실시하여 획득한 결과이다. 실험을 통해서 알 수 있는 결과는 실시간 콘텐츠 필터링의 경우 10 ~ 15 frame/sec 간격으로 샘플링을 할 경우 사양이 가장 낮은 650MHz의 PC에서도 충분히 잘 동작할 수 있는 것을 알 수 있다. 이는 현재 700MHz의 CPU를 타겟으로 개발되는 TV 셋탑 박스의 사양에도 충분히 충족하는 조건이 된다.

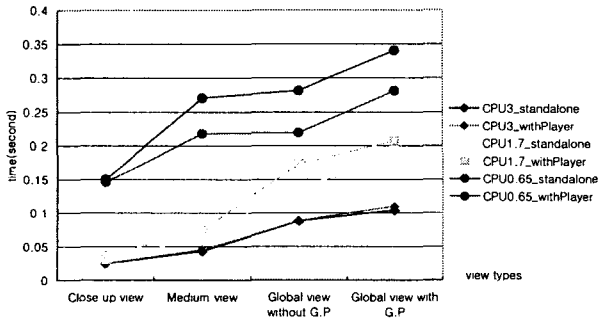


그림 7. 뷰 타입 결정을 위해 소요되는 시간측정 결과

본 논문에서 제안한 슛 장면 검출에 대한 결과는 표 1에 나타난다. 이전 연구결과[6]와 비교했을 때 제안한 슛 장면 검출 알고리즘의 결과는 리콜율은 81.5%이고 정확도는 76.4%로 성능이 향상된 것을 확인할 수 있다. 새로 제안하는 알고리즘은 좀 더 정확한 슛 장면 검출을 위해서 두 프레임의 패턴만을 매칭하는 것이 아닌, 오디오 파워 특징 값과 뒤에 이어지는 리플레이의 패턴을 매칭시킨 후 필터링을 수행하므로 슛 장면이 발생하고 리플레이 샷이 검출되기 전의 시간 차이만큼 지연 시간이 생기게 된다.

표 1. 슛 장면 검출 결과

| | Soccer A | Soccer B | Soccer C |
|-------------------|----------|----------|----------|
| Correct (N_c) | 15 | 27 | 24 |
| Missed (N_m) | 6 | 5 | 3 |
| False (N_f) | 4 | 10 | 7 |
| Precision | 78.9% | 72.9% | 77.4% |
| Recall | 71.4% | 84.3% | 88.8% |

VI. 결론

다채널 디지털 방송환경에서 시청자에게 특정 시간에 하나의 채널에서 방송되는 프로그램의 시청만이 아닌 타 채널에서의 이벤트 장면도 필터링하여 제공하는 기능은 질적인 방송 서비스의 발전을 가져올 것이다. 이에 대해서 본 논문은 실시간에 근접하여 필터링이 가능한 슛 장면 검출 알고리즘을 제안하였고, 실험 결과를 통해 축구경기의 슛 장면에 대해서 효과적이고 효율적으로 필터링 하였음을 보였다.

보다 정확한 슛 장면 검출을 위하여는 보다 세분화된 뷰 결정 알고리즘이 필요하나 처리시간과의 Trade-off 관계에 있다. 그리고 앞으로 오디오의 정보를 효과적으로 이용하는 방안에 대한 추가 실험이 필요하다.

참고문헌

[1] Assfalg, J.; Bertini, M.; Del Bimbo, A.; Nunziati, W.; Pala, P, " Soccer highlights detection and recognition using HMMs" Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE

International Conference on , Volume: 1 , 26-29 Aug. 2002 pp.825 - 828 vol.1

- [2] Shu-Ching Chen, et al., "Detection of Soccer Goal Shots Using Joint Multimedia Features and Classification Rules," Proceedings of the Fourth International Workshop on Multimedia Data Mining, pp. 36-44, 2003.
- [3] Ahmet Ekein et al "Automatic Soccer Video Analysis and Summarization," IEEE Tran. On Image Processing, Vol. 12, No. 7, pp. 796-807, 2003.
- [4] A. Ahmet Ekein and A. Murat Tekalp, "Generic Play-break Event Detection for Summarization and Hierarchical Sports Video Analysis," ICME 2003, pp. 27-29 March, 2000.
- [5] 김의진, 이광국, 허영식, 정철근, 김상균, 김지연, 김희율 "경기 내용에 기반한 농구 비디오 요약 방법" 제 17회 신호처리합동학술대회, Vol. 17, No. 1, pp.67
- [6] 진성호, 조준호, 배태년, 노용만, 강경옥 "다 채널에서 콘텐츠 필터링 알고리즘" 제 17회 신호처리합동학술대회, Vol. 17, No. 1, pp.68

Acknowledgements

본 연구는 한국전자통신연구원(ETRI)의 지원받은 수행중인 SmartTV과제의 연구 결과 중 일부임