

축구 동영상 요약을 위한 재연 장면 자동 추출 알고리즘

정진국, 김주영, 낭중호, 김정수*, 하명환*, 정병희*
서강대학교 컴퓨터학과, *KBS 기술 연구소

(jjguk, rayzel)@mlneptune.sogang.ac.kr, jhnang@ccs.sogang.ac.kr

A Replay Shot Detection Algorithm for the Soccer Video Abstraction

Jin-Guk Jeong Ju-Young Kim Jong-Ho Nang Gyung-Su Kim* Myung-hwan Ha* Byung-Hui Jeong*
Dept. of Computer Science, Sogang University, *KBS Technical Research Institute

요 약

최근 디지털 비디오 데이터의 사용이 급격히 증가하면서 저급 수준의 정보를 이용하여 고급 수준의 내용 정보를 자동으로 추출하는 기술이 필요하게 되었다. 축구와 같은 분야에서는 그 중에서도 골, 프리킥, 파울 장면 등의 고급 수준 내용 정보가 중요한 의미를 갖게 되는데 특히, 이러한 장면 중 중요하다고 여기는 장면은 재연 장면을 통하여 다시 시청자에게 보여지게 되며, 축구 비디오에 대한 요약에서는 이런 장면들이 꼭 포함되어야 한다. 본 논문에서는 이러한 축구 비디오 데이터에서 재연 장면을 자동으로 추출하는 방법을 제안한다. 기본적으로는 축구 고유의 특징들을 이용하는데 첫 번째 특징은 샷의 길이가 너무 짧거나 너무 길지 않다는 것이고, 두 번째 특징은 재연 장면이라는 것은 장면이 느리게 다시 재생되는 것이기 때문에 움직임 특징이 일반적인 장면과는 다르다는 것이다. 본 논문에서는 오브젝트의 움직임을 구분하기 위하여 재연 장면을 두 가지 종류로 나누었다. 하나는 확대 상태의 재연 장면이고 다른 하나는 축소 상태의 재연 장면이다. 본 논문의 알고리즘을 적용하여 실험한 결과 Recall과 precision 모두 70% 이상 나오는 것을 알 수 있었다.

1. 서론

네트워크, 저장 매체의 발달, 이외에 여러 압축 기술 등과 같은 관련 기술의 발달로 인하여 디지털 비디오의 사용이 보다 일반화되어졌다. 또한 이러한 사실로 인해 비디오 데이터 안의 정보를 추출하는 기술의 연구가 더욱 중요한 의미를 갖게 되었다. 비디오 데이터를 분석하는 기술은 여러 가지 방향으로 이루어져 왔는데 그 중 가장 기본이 되는 것은 자동 인덱싱 기술[7,8]이다. 자동 인덱싱 기술이라는 것은 샷과 신을 자동으로 검출하여 비디오의 구조를 알아내는 것으로 비디오 데이터를 분석하는 데 있어 가장 기본이 되는 기술이다. 하지만 이러한 기술만으로는 실제 응용 프로그램에 이용하는 데 한계가 있다. 일반적으로 실제 응용 분야에서 사용자가 원하는 정보는 저급 수준 내용 정보가 아니라 고급 수준 내용 정보이기 때문이다. 이와 같은 이유로 인해 고급 수준 내용 정보를 자동으로 추출하는 기술[1,2,3,4]이 필요하다.

각 응용 분야에 따라 필요한 고급 수준 내용 정보는 서로 다르게 된다. 그 중에서도 축구와 같은 스포츠 분야에서는 특징이 나오는 장면이 중요할 것이다. 하지만 비디오를 요약하는 응용 분야에 이용할 때 축구의 특징 장면만을 이용한다면 너무 짧은 요약이 나오게 되어

사용자의 다양한 욕구를 충족시켜주지 못하게 될 것이다. 이 문제를 해결하기 위해 이용할 수 있는 장면이 재연 장면이다. 각 개인이 중요하다고 여기는 장면은 모두 다르겠지만 재연 장면은 방송국에서 방송을 하는 도중에 중요하다고 여기는 장면이기 때문에 보다 객관적인 기준이 될 수 있다. 이와 같은 이유로 인해 본 논문에서 제시하는 재연 장면 자동 추출 알고리즘은 다양한 응용 분야에서 유용하게 이용이 될 수 있다.

서로 다른 분야의 비디오 데이터들은 각각 다른 특성을 가지고 있다. 특히 스포츠 분야에서는 이러한 특징들이 더욱 잘 나타나는데 이는 운동장의 특성, 선수들의 움직임 등이 다른 분야보다 정형화되어 있기 때문이다. 본 논문에서는 이러한 정형화된 특징을 이용하여 알고리즘을 구현하였다. 본 논문에서 이용한 재연 장면의 특징은 다음과 같다. 첫 번째는 샷의 길이가 너무 짧거나 길지 않다는 것이고 두 번째는 재연 장면이라는 것은 장면이 느리게 다시 재생되는 것이기 때문에 움직임 특징이 일반적인 장면과는 다르다는 것이다. 본 논문의 알고리즘을 적용하여 실험한 결과 Recall과 precision 모두 70% 이상 나오는 것을 알 수 있었다.

2. 재연 장면 추출 알고리즘

2.1 축구 동영상 특성

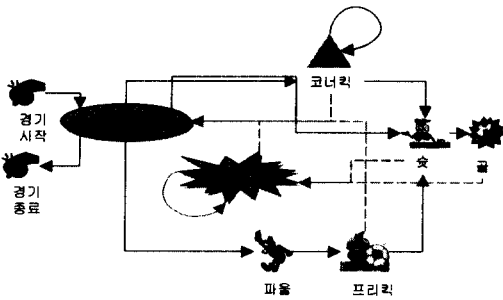


그림 1 축구 동영상 이벤트 상태도

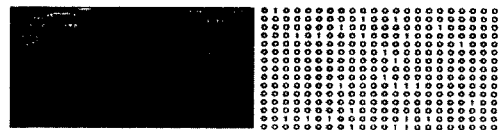
<그림 1>은 축구 동영상의 이벤트 상태도를 나타내는 그림이다. 그림에서 보면 알 수 있듯이 축구 동영상의 진행은 여러 가지 이벤트들 - 코너킥, 슈팅, 골 등 - 이 심판의 경기 진행 혹은 재연 장면으로 연결이 되어 있다. 특히 재연 장면으로 연결이 되는 이벤트는 골 장면이나 심한 파울 장면 등 방송에서 경기 내용상 중요하다고 여기는 장면들이기 때문에 축구 동영상을 요약하는 데 있어 중요한 정보가 된다.

2.2 축구 동영상에서의 샷의 분류

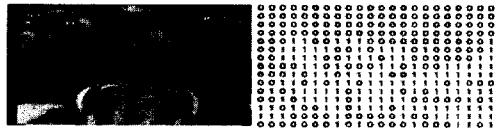
본 논문에서는 축구 동영상의 샷의 종류를 다섯 가지로 나누었다. 확대 장면, Close-Up 장면[5], 축소 장면과 재연 장면은 축소 장면에서 재연이 되는 경우와 확대 장면에서 재연이 되는 경우로 나누었다. 확대 장면은 경기 상태를 확대하여 보여준 장면이고 Close-Up 장면은 경기가 잠깐 중단되었을 때 선수의 모습을 보여주는 장면이다. 그리고 축소 장면은 경기 상태를 축소하여 보여준 장면이다. 재연 장면을 두 가지로 나눈 이유는 화면내의 오브젝트의 크기에 따라서 특징이 다르게 나타나기 때문이다. 즉, 확대 장면에서의 재연 장면은 화면내의 오브젝트가 크기 때문에 축소 장면에서의 재연 장면보다 움직임이 크게 나타나는 것이다.

재연 장면이라는 것은 경기 장면이 느리게 다시 재생되는 것을 얘기한다. 이러한 특징 때문에 재연 장면은 다른 장면과는 다른 움직임 특성을 보여준다. 움직임 특성은 두 가지로 나타낼 수 있는데 하나는 프레임내의 움직임이 어느 정도인가를 나타내는 수치이고 다른 하나는 프레임내의 움직임이 어느 정도로 밀집해있는가를 나타내는 수치이다. 이러한 수치를 얻어내는 방법은 Optical Flow[6]를 이용하였다. Optical Flow라는 것은 여러 프레임을 입력으로 취한 후 프레임 안에서 움직임이 있는 오브젝트가 있는 지 혹은 카메라의 이동이 있는 지를 알아내는 데 이용되는 함수이다. <그림 2>는 본 논문에서 분류한 각 장면의 움직임 특성을 보여주는 그림이다. 각 프레임을 8x8 블록으로 나눠 Optical Flow를 적용했으며, 1이라고 표현이 된 부분이 움직임

이 있다고 추출된 블록이며, 움직임 크기는 이러한 블록들의 에너지 평균값을 나타낸다. 그림을 보면 알 수 있듯이 축소 장면이 다른 장면들에 비해 움직임의 크기도 작고 밀집도 또한 작음을 알 수 있다. 이에 비해 Close-Up 화면은 보통 경기가 중단되어 있는 상태이기 때문에 움직임의 크기가 크지는 않지만 오브젝트의 크기가 크기 때문에 움직임의 밀집도는 크게 나타난다. 축소 장면에서의 재연 장면은 축소 장면 다음으로 움직임의 크기와 밀집도가 작게 나타나고 확대 장면에서의 재연 장면은 움직임의 크기와 밀집도가 크게 나타난다. 이러한 움직임 특징을 이용하여 재연 장면을 찾아낼 수 있는 것이다.



(a) 축소 장면 (1의 개수 : 27, 움직임 크기 : 2.77)



(b) 확대 장면 (1의 개수 : 130, 움직임 크기 : 3.84)



(c) 확대 장면 (1의 개수 : 89, 움직임 크기 : 7.14)



(d) 축소 재연 장면 (1의 개수 : 42, 움직임 크기 : 2.95)



(e) 확대 재연 장면 (1의 개수 : 113, 움직임 크기 : 5.98)

그림 2 장면에 따른 움직임 특성

2.3 축구 동영상에서의 재연 장면 추출 알고리즘

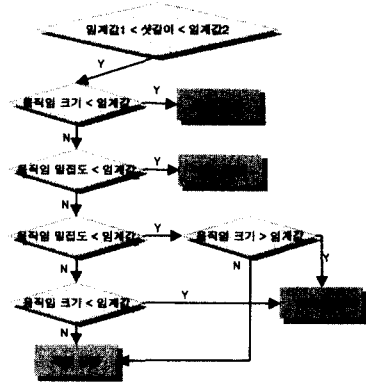


그림 3 재연 장면 추출 알고리즘

<그림 3>는 재연 장면 추출 알고리즘을 나타내는 그림이다. 우선 샷의 길이에 의해 재연 장면 추출하는 범위를 줄여주고 그 이후에 재연 장면의 움직임 특징에 의해 재연 장면을 추출하게 된다.

샷의 길이라는 것은 샷내의 프레임 개수를 의미한다. 축구에서 샷내의 프레임 개수가 아주 적은 경우는 확대 장면이 대부분이고 프레임 개수가 아주 많은 경우는 경기 위주로 보여주는 축소 장면이 대부분이다. 이러한 특성을 이용하여 재연 장면을 찾아내는 범위를 축소시킬 수 있다.

3. 성능 평가 및 분석

표 1 실험결과

	재연 장면	확대 장면	축소 장면
Precision	49% (25/51)	100% (59/59)	81% (87/107)
Recall	93% (25/27)	48% (59/122)	96% (87/91)

<표 1>은 실험 결과를 보여주는 표이다. 실험에 이용한 비디오 데이터는 MPEG-1 형식의 데이터이다. 표를 보면 알 수 있듯이 재연 장면의 검출율이 비록 Precision이 낮기는 하지만 Recall 이 높음을 알 수 있다. 검출율이 100%가 아니면 어떤 수정 도구가 필요하기 때문에 Recall 이 높은 경우가 Precision 이 높은 경우보다 더 도움이 된다. 본 논문의 알고리즘을 사용한 결과 재연 장면 검출율의 Recall 이 93%정도로 좋은 성능을 보임을 알 수 있었다. 재연 장면의 Precision과 확대 장면의 Recall의 성능이 좋지 않은 이유는 경기가 중단되었을 때의 확대 장면과 재연장면의 특성이 비슷하기 때문이다. 이것은 경기가 중단 되었을 때의 확대

장면의 움직임이 적기 때문이며, 이 두 종류의 장면사이에서 검출 성능이 낮아지게 되었다.

4. 결론 및 앞으로의 연구 방향

최근 많은 멀티미디어의 사용으로 인해 멀티미디어 데이터를 분석하기 위한 여러 가지 알고리즘이 제안되어 왔다. 특히 최근에는 저급 수준 내용 정보를 이용하여 고급 수준 내용 정보를 추출하는 연구 또한 활발히 진행되고 있는데 본 논문에서는 여러 가지 영역 중에서 축구 고유의 특징 - 샷의 길이, 장면의 종류에 따른 움직임 특성 - 을 이용하여 재연 장면을 찾아내는 알고리즘을 제안하였다. 실험 결과 본 논문에서 제안한 알고리즘은 recall이 93%, precision이 49% 정도의 성능을 보이는 것을 알 수 있었다. 실험 결과를 통해 본 논문의 알고리즘은 축구 동영상 요약에 유용하게 적용할 수 있음을 알 수 있었다. 카메라의 이동이나 오브젝트의 움직임이 예측할 수 없는 형태로 나타내는 경우로 인해 재연 장면을 잘못 추출하거나 못 찾는 경우가 발생하였는데 앞으로의 연구에서는 오브젝트의 추출 등을 더 보완하여 보다 견고한 알고리즘을 제안하도록 할 것이다.

5. 참고 문헌

[1] Nitta, N., Babaguchi, N. and Kitahashi, T., " Extracting Actors, Actions and Events from Sports Video -A fundamental approach to story tracking," *Proceedings of 15th International Conference on Pattern Recognition, Volume 4*, p. 718 - 721, 2000.
 [2] Noboru Babaguchi, " Towards Abstracting Sports Video by Highlights," *Proceedings of IEEE International Conference on Multimedia and Expo, Volume 3*, p.1519 -1522, 2000.
 [3] V. Kobla, and D.S. Doermann, " Detection of Slow-motion replays for Identifying Sports Videos," *Proceedings of IEEE on Multimedia Signal Processing*, p.135-140, 1999.
 [4] Pan, H., van Beek, P., and Sezan, M.I. , " Detection of slow-motion replay segments in sports video for highlights generation," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing., Volume 3*, 2001.
 [5] P. Xu, L. Xie, S.-F. Chang, A. Divakaran, A. Vetro, and H. Sun, "Algorithms and Systems for Segmentation and Structure Analysis in Soccer Video," *Proceedings of IEEE International Conference on Multimedia and Expo*, 2001.
 [6] D.Zhong and S.-F. Chang, " Long-Term Moving Object Segmentation and Tracking Using Spatio-Temporal Consistency," *Proceedings of IEEE International Conference on Image Processing*, 2001.
 [7] H. J. Zhang, C. Y. Low and S. W. Smoliar, "Video Parsing and Browsing using Compressed Data," *Proceedings of Multimedia Tools and Applications, Vol. 1*, pp. 89-111, 1995.
 [8] H. C. Liu and G. L. Zick, "Scene Decomposition of MPEG Compressed Video," *Proceedings of SPIE on Digital Video Compression: Algorithm and Technologies*, pp.26-37, 1995.