

다중 카메라로 관심선수를 중심으로 촬영한 농구 경기 동영상에서

베스트 뷰 선택방법

홍호탁⁰¹ 송진하² 낭중호¹ 엄기문³ 조기성³서강대학교¹ 순천대학교² 한국전자통신연구원³

hotakhong@sogang.ac.kr, the_jn@sunchon.ac.kr, jhngang@sogang.ac.kr, gmum@etri.re.kr, chokis@etri.re.kr

A Best View Selection Method for Basketball Videos focusing on Interested Player with Multi-Cameras

Hotak Hong⁰¹ Jinha Song² Jongho Nang¹ Gimun Um³ Keeseong Cho³Sogang University¹ Suncheon National University² Electronics and Telecommunications Research Institute(ETRI)³

요 약

최근 스포츠 장면을 다양한 각도로 생생하게 전달하기 위해 동원되는 카메라 대수가 증가함에 따라 수많은 카메라 화면에서 실시간으로 편집을 수행할 때 순간적으로 최고의 화면을 고르는데 어려움이 있다. 지금까지 여러 대 카메라로 스포츠 경기를 촬영한 영상들에서 자동으로 최고의 화면을 선택하는 방법들이 연구되어 왔지만 배경이 고정된 영상들만을 고려해 배경이 움직이는 영상들을 고려하는 연구가 필요하다. 본 논문에서는 다중 카메라로 관심선수를 추적하면서 촬영된 동영상들에서 매 프레임을 추출해 관심선수를 대상으로 최고의 화면을 고를 수 있는 척도를 세워 이를 정량적으로 나타내어 같은 시점의 카메라 화면 중 최고의 화면을 선택하는 방법을 제안한다. 실험을 통해 제안한 방법으로 선택된 최고의 화면과 실제 사람들이 주관적으로 판단한 최고의 화면이 일치하는지 비교 확인한다. 실험 결과 사람들이 제일 선호하는 베스트 뷰와 일치율이 23%로 낮았지만 반대로 제일 선호하지 않은 베스트 뷰와 일치율이 14%로 확실히 사람들이 선호하지 않는 화면은 선택하지 않는 것을 알 수 있었다.

1. 서 론

최근 스포츠 장면을 생생하게 전달하기 위해 다양한 각도에서 경기를 중계하고 이를 위해 많은 카메라가 동원되고 있다. 스포츠 경기에 동원되는 카메라 대수가 증가함에 따라 스포츠 경기 특성상 실시간 편집을 수행할 때 수많은 카메라 화면에서 순간적으로 최고의 화면을 고르는데 어려움이 있다.

지금까지 여러 대 카메라로 스포츠 경기를 촬영한 영상들에서 자동으로 최고의 화면을 선택하는 방법들이 많이 연구되어 왔다. 하지만 기존 방법들은 배경이 움직이는 영상들은 고려하지 않고 배경이 고정된 영상들만을 고려하여 최고의 화면을 선택했다. 따라서 배경이 움직이는 영상들을 대상으로 최고의 화면을 자동으로 선택하는 방법에 대한 연구가 요구된다.

배경이 고정된 영상들에서 최고의 화면을 선택하는 연구는 배경 제거 알고리즘을 사용하여 화면 안에 존재하는 선수들의 실루엣을 추출하여 최고의 화면을 선택하기 위한 요소로 사용하고 있다. 하지만 배경이 움직이는 영상에서는 배경이 고정된 영상과 달리 배경을 모델링하기 어려워 배경 제거 알고리즘을 사용하여 선수들의 실루엣을 추출하기 어렵다. 또한, 배경이 고정된 영상들에서는 카메라의 설치위치를 토대로 선수들의 경기장 내 위치를 알 수 있어 이를 최고의 화면을 선택하기 위한 요소로 사용했지만 배경이 움직이는 화면에서는 같은 방법으로 선수들의 경기장 내 위치를 알 수 없다.

기존 연구에서 사용된 방법을 배경이 움직이는 영상에서는 사용하기 부적합하다. 기존 방법들을 배경이 움직이는 영상에

서도 사용 가능하도록 변경하고 배경이 움직이는 영상이 가지는 특징을 고려하여 최고의 화면을 선택하기 위한 요소들을 추가한다면 어떠한 스포츠 영상에서도 최고의 화면을 선택할 수 있을 것이다.

본 논문에서는 서로 다른 위치에 설치된 4개의 PTZ(Pan-Tilt-Zoom) 카메라로 관심선수를 추적하면서 촬영된 동영상들에서 매 프레임을 추출해 관심선수를 대상으로 최고의 화면을 고를 수 있는 척도를 세워 이를 정량적으로 나타내고 같은 시점의 4개의 카메라 화면 중 최고의 화면을 선택하는 방법을 제안하고 실험을 통해 실제 사람이 주관적으로 판단한 최고의 화면과 일치하는지 비교 확인한다.

2. 관련 연구

다양한 스포츠 분야에서 최고의 화면을 선택하기 위한 방법들이 연구되었다. H. Jiang 외[1]는 스쿼시 경기를 촬영한 영상들에서 선수들을 객체 추적 알고리즘을 통해 추적하여 선수들의 크기가 상대적으로 가장 큰 화면을 최고의 화면으로 선택한다. F. Daniyal 외[2, 4]는 농구 경기를 촬영한 영상들에서 화면 내 존재하는 선수들의 인원수, 활동량, 경기장 내 위치를 정량적으로 나타내어 최고의 화면을 선택한다. F. Chen 외[3]는 농구 경기를 촬영한 영상들을 이벤트 단위로 나누어서 선수들의 가시성, 경기장 내 위치를 정량적으로 나타내어 이벤트 단위로 최고의 화면을 선택한다. X. Wang 외[5]는 축구 경기를 촬영한 영상들에서 화면 안에 선수들의 존재 여부와 경기장 내 위치에 가중치를 두어 사용자가 선호하는 최고의 화면을 선택할 수 있게 한다.

기존 논문들 마다 최고의 화면을 정의하고 있다. 다음 표 1

본 연구는 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (R0101-16-293, 방송용 영상 인식 기반 객체 중심 지식융합 미디어 서비스 플랫폼 개발).

은 기존 논문들에서 정의하고 있는 최고의 화면을 나타낸다.

표 1 최고의 화면 정의

논문저자이름	최고의 화면 정의
H. Jiang 외[1]	사람 크기가 상대적으로 가장 큰 프레임
F. Daniyal 외[2]	경기내용에 대한 정보량이 많이 포함된 프레임
F. Chen 외[3]	사람이 겹쳐있지 않고 해상도가 좋은 프레임
F. Daniyal 외[4]	경기내용에 대한 정보량이 많이 포함된 프레임
X. Wang 외[5]	경기내용에 대한 정보량이 많이 포함된 프레임

기존 연구에서는 배경이 고정된 화면에서 최고의 화면을 고르다 보니 최고의 화면을 ‘경기내용을 한눈에 볼 수 있는 화면’으로 정의하는 논문들이 많다. 본 논문에서는 관심객체 중심으로 촬영된 영상에서 최고의 화면을 고르는 문제로 최고의 화면을 ‘관심선수의 모습을 가장 잘 보여주는 품질이 좋은 프레임’으로 정의하고 앞으로 ‘베스트 뷰’라 말한다.

3. 다중 카메라 영상들에서 베스트 뷰 선택방법

본 논문에서 제안하는 베스트 뷰 선택 과정은 크게 카메라 영상 별 관심선수 추적, 카메라 영상 별 베스트 뷰 판단 척도 계산, 베스트 뷰 선택으로 구분된다. 그림 1은 베스트 뷰 선택 과정을 도식화한 것이다.

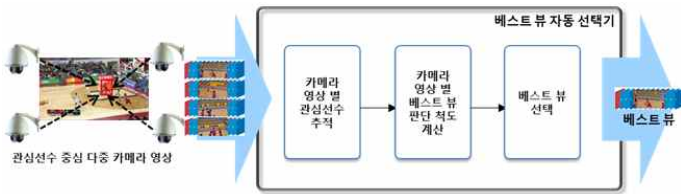


그림 1 베스트 뷰 선택 과정

농구 경기장 내 관심선수의 위치를 추적하면서 실시간으로 4대의 PTZ 카메라를 제어하여 관심선수 중심으로 촬영된 다중 카메라 영상들을 얻는다. 획득한 다중 카메라 영상 별로 화면 내 관심선수 위치를 추적하여 관심선수 영역을 표시한다. 표시된 영역을 대상으로 영상 별로 매 프레임 마다 베스트 뷰 판단 척도를 계산한다. 계산된 값을 기반으로 같은 시점 매 프레임 마다 베스트 뷰를 선택한다.

다중 카메라로 관심선수를 추적하면서 촬영할 때 관심선수가 화면 중앙에 일정한 크기를 유지하도록 촬영되었다. 이러한 특징을 바탕으로 관심선수 영역을 추출하기 위해 영상 내에서 관심선수를 추적한다.

3.1 카메라 영상 별 관심선수 추적

다중 카메라 영상에서 관심선수가 화면 중앙에 위치한다는 특징과 Faster R-CNN[6], KCF(Kernelized Correlation Filters)[7]를 이용하여 카메라 영상 별로 영상 내 관심선수를 추적한다. 각 영상 별로 처음 프레임에서 Faster R-CNN을 이용하여 영상 내 선수들을 검출하고 화면 중앙에서 가장 가까운 선수를 관심선수로 설정하여 KCF를 통해 관심선수를 추적한다. Faster R-CNN과 영상특징만을 이용하여 관심선수를 추적하는 것보다 KCF를 결합하여 관심선수를 추적하는 것이 정확도가 6% 더 좋았다.

다중 카메라 영상 별로 매 프레임 마다 관심선수를 추적하여

영상 내 관심선수의 영역을 표시한다.

3.2 카메라 영상 별 베스트 뷰 판단 척도 계산

각 카메라 영상 별로 매 프레임마다 획득한 관심선수 영역을 대상으로 베스트 뷰 판단 척도를 계산하여 프레임을 정량적으로 표현한다. 베스트 뷰 판단 척도는 관심선수를 식별하는데 중요한 요소인 관심선수의 실루엣과 얼굴 그리고 이미지 품질에 영향을 주는 블러 현상의 정도를 척도로 삼는다.

3.2.1 실루엣 추출을 통한 관심선수의 가시성 정량화

MNC(Multi-task Network Cascades)[8]에서 제시한 방법으로 매 프레임마다 관심선수의 영역 안에서 사람 실루엣을 추출한다. 추출한 실루엣으로 관심선수의 가시성 정도를 알 수 있고 일정 구간동안 추출한 실루엣을 겹쳐 관심선수의 활동량을 알 수 있다. 그림 2은 관심선수의 활동량을 표현한 그림이다.



그림 2 관심선수의 활동량 관심선수의 영역에서 일정 구간 추출한 실루엣은 회색으로 현재 프레임에서 추출한 실루엣은 흰색으로 표현한다. 관심선수의 가시성을 정량화하는 식은 다음과 같다.

$$S_t^i = (1 - \frac{N_w}{N_g + N_w}) * 100 \quad (1)$$

N_w 은 흰색 픽셀 개수, N_g 은 회색 픽셀 개수를 나타낸다.

3.2.2 얼굴검출

haar feature를 사용하여 관심선수의 영역 안에서 사람 정면 얼굴을 검출한다. 검출 여부에 따라 다음과 같이 점수를 할당한다.

$$F_t^i = \begin{cases} 100 & \text{Frontal Face Detection} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

3.2.3 이미지의 블러 현상 정도를 이용한 프레임 정량화

인접한 두 프레임의 차를 통해 프레임에 나타난 블러 현상의 정도를 알 수 있다. 그림 3은 인접한 두 프레임의 차를 이미지로 나타낸 그림이다.

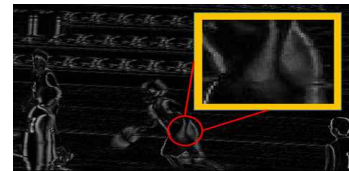


그림 3 인접한 두 프레임의 차 블러 현상은 영상 내 급격한 변화가 일어날 때 생기 때문에 그림 3에서 알 수 있듯이 급격하게 변화가 일어난 정도에 따라 흰색이 두드러진다. 이미지 내에서 블러 현상 정도를 정량화하는 식은 다음과 같다.

$$B_t^i = \frac{D_t^i}{D_t^1 + D_t^2 + D_t^3 + D_t^4} \quad (3)$$

D 는 인접한 두 프레임의 차 이미지에서 흰색 픽셀의 개수, t 는 시간, i 는 카메라 번호를 나타낸다.

위에서 정의한 식 (1),(2),(3)을 통해 각 카메라 영상 별로 매

프레임을 정량적으로 표현한다.

3.3 베스트 뷰 선택

3.2절에서 세운 식 (1), (2), (3)을 바탕으로 베스트 뷰를 선택하기 위한 식은 다음과 같다.

$$FS_i^i = S_i^i + F_i^i - B_i^i \quad (4)$$

t는 시간, i는 카메라 번호를 나타낸다.

영상들의 같은 시점의 프레임들에서 식 (4)으로 계산한 값이 가장 큰 프레임을 베스트 뷰로 선택한다.

4. 실험 및 검증

본 실험에서는 베스트 뷰 판단 척도에 의해 자동 선택된 베스트 뷰와 실제 사람들이 주관적으로 판단한 베스트 뷰가 일치하는지 비교 확인한다. 실험에는 다양한 각도에서 4개의 PTZ 카메라로 관심선수를 추적하면서 촬영된 농구경기 동영상에서 관심선수 추적이 잘 된 구간(200프레임)을 추출하여 실험 데이터로 사용한다.

일반인 16명에게 추출한 구간을 보여주면서 매 프레임마다 4개의 화면 중 베스트 뷰를 선택하게 한다. 그림 4은 매 프레임마다 4개의 카메라를 대상으로 사람들의 선호도를 나타낸 그림이다.

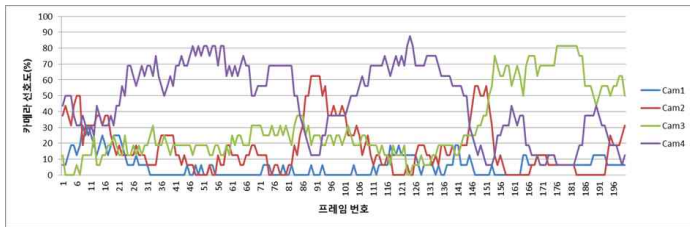


그림 4 매 프레임에서 사람들의 카메라 선호도

전체 200프레임 중 57%에서 4번 카메라 화면의 선호도가 제일 높았고 다른 카메라들과 선호도 차이도 분명하게 구분되었다. 4번 카메라 화면의 선호도가 제일 높은 구간의 프레임들을 분석해보면 공통적으로 4번 카메라 화면에만 관심선수의 얼굴이 식별 가능했고 관심선수의 전신이 전부 보였다. 여기서 베스트 뷰 판단 척도에 얼굴과 실루엣 요소가 적합하다는 것을 알 수 있었다.

사람들에 의해 선택된 베스트 뷰와 베스트 뷰 판단 척도에 의해 자동 선택된 베스트 뷰를 비교한다. 그림 5은 매 프레임마다 선택된 베스트 뷰 카메라 번호를 나타낸 그림이다.

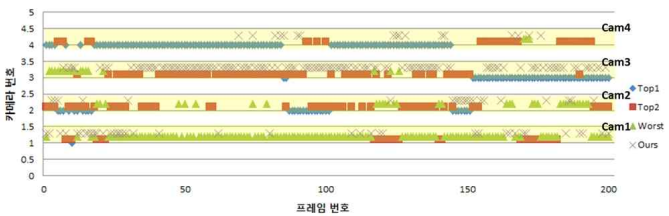


그림 5 매 프레임 선택된 베스트 뷰 카메라 번호

매 프레임에서 사람들이 제일 선호하는 베스트 뷰(Top1), 두 번째로 선호하는 베스트 뷰(Top2), 제일 선호하지 않는 베스트

뷰(Worst) 이렇게 3가지 경우와 자동으로 선택된 베스트 뷰(Ours)를 비교했다. 그 결과 Top1은 23%, Top2은 36%, Worst은 14% 로 전체적으로 낮은 일치율을 보였다. 하지만 Top1과 Top2에서 선택된 화면이 모두 베스트 뷰라고 하면 성능은 59%으로 올라갔고 Worst의 일치율이 낮은 것을 볼 때 확실히 사람들이 선호하지 않는 화면은 고르지 않는 것을 알 수 있었다.

5. 결론

본 논문에서는 기존 연구가 고려하지 않은 배경이 움직이는 영상들에서 베스트 뷰를 선택하는 방법을 제안했다. 다중 카메라로 관심선수를 추적하면서 촬영된 동영상들에서 매 프레임을 추출해 관심선수를 대상으로 베스트 뷰 판단 척도를 계산하여 정량적으로 표현된 값을 기반으로 같은 시점의 4개의 카메라 화면 중 베스트 뷰를 선택했다. 실험을 통해 자동 선택된 베스트 뷰와 실제 사람들이 주관적으로 판단한 베스트 뷰의 일치율을 확인했다. 그 결과 사람들이 제일 선호하는 베스트 뷰와 일치율이 23%로 낮았지만 반대로 제일 선호하지 않은 베스트 뷰와 일치율이 14%로 확실히 사람들이 선호하지 않는 화면은 고르지 않는 것을 알 수 있었다.

참고 문헌

- [1] H. Jiang, S. Fels and J.J. Little, "Optimizing Multiple Object Tracking and Best View Video Synthesis," *IEEE Transactions on Multimedia*, 2008.
- [2] F. Daniyal, M. Taj and A. Cavallaro, "Content and Task-Based View Selection from Multiple Video Streams," *Springer J. Multimedia Tools and Applications*, 2010.
- [3] F. Chen and C. D. Vleeschouwer, "Autonomous production of basketball videos from multi-sensored data with personalized viewpoints," *Proceeding of 10th Workshop on Image Analysis for Multimedia Interactive Services*, 2009.
- [4] F. Daniyal and A. Cavallaro, "Multi-camera Scheduling for Video Production," *Proceeding of Conference for Visual Media Production(CVMP)*, 2011.
- [5] X. Wang, Y. Muramatu, T. Hirayama and K. Mase, "Context-Dependent Viewpoint Sequence Recommendation System for Multi-view Video," *Proceeding of IEEE International Symposium on Multimedia*, 2014.
- [6] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Proceeding of Advances in Neural Information Processing Systems(NIPS)*, 2015.
- [7] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI)*, Vol.37, No.2, 2015.
- [8] J. Dai, K. He and J. Sun, "Instance-aware Semantic Segmentation via Multi-task Network Cascades," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2016.