딥러닝을 이용한 축구 하이라이트 영상 생성기

(DaSH : Deep Learning and Soccer Highlight)

노신영 이우리 윤나라 길현영
한국항공대학교 소프트웨어학과

Soccer Video Highlight Extraction using Deep Learning

Shinyeong Noh Woori Lee Nara Yoon Hyunyoung Kil
Departmnt of Software, Korea Aerospace University

요 약

최근 축구 경기 영상은 짧게 요약된 경기 하이라이트 영상의 조회수가 전체 경기 영상보다 3배에서 많게는 20배까지 월등히 높은 것을 알 수 있다. 이러한 현상에 맞추어 더 효율적으로 더 다양한 하이라이트 영상을 생성할 수 있도록 하는 시스템을 개발하였다. 본 논문에서는 축구 경기 영상에서 사용자가 하이라이트 장면을 직접 선택할 수 있는 모델을 제안한다. 우리의 모델은 통상적인 하이라이트 외에 돌발 상황인 하이라이트를 검출하기 위해 추가적으로 오디로의 폭발적인 증가 구간을 검출하도록 했다. 직접 수집한 K리그 축구 경기 영상을 통해 하이라이트 예측 성능을 평가한다.

Abstract

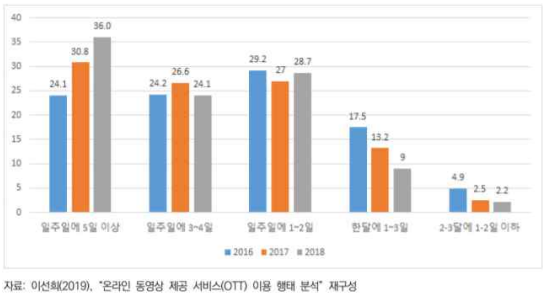
Recently soccer game video shows that the view count of the summurized game highlight video is three to twenty times higher than the entire game video. In response to this trend, we develop a system that allows more efficiently generating a variety of highlight images. In this paper, we propose a model that allows users to directly select highlight scenes from soccer game images. In addition to conventional highlights, our model further allows detection of explosive increasing segments of the audio to detect sudden highlights. It will evaluate the performance of the K-League's prediction through the video clips of the K-League soccer matches collected directly.

KeyWords: Video Highlight, Action Spotting, Event Classification, Audio Highlight

I. 서 론

한국을 비롯한 전 세계의 스마트폰의 보급률은 매해 증가하고 있다. 소프트웨어와 통신 기술의 발달 또한 빠르게 이루어지고 있다. [그림1]에서 볼 수 있듯이, 이로 인해 'YouTube'나 'Naver TVcast' 등의 OTT 동영상 서비스 플랫폼의 수요 또한 증가하고 있다. 이처럼 바쁜 현대인들의 미디어 소비 환경은 시대에 맞추어 이동시간이나 짧은 여가시간 활용에 용이하도록 변화하고 있다.

다양한 콘텐츠들 중에서도 축구 경기 영상은, 10분 내외로 주요 부분만 편집한 경기



자료: 이선희(2019), "온라인 동영상 제공 서비스(OTT) 이용 행태 분석" 재구성

그림 1.온라인 동영상 제공 서비스 이용 형태

하이라이트 영상의 조회 수가 전체 경기 영상보다 3배에서 많게는 20배가량까지 월등히 높은 것을 확인할 수 있다.

그림 2. 전체경기영상과 하이라이트 영상 조회수

또한 해외에서 활약하는 선수들이 많아짐에 따라, 다양한 국가의 다양한 리그의 스포츠가 인기를 끌고 있다. 단순히 해당 선수의 경기뿐만 아니라, 해당 리그에 관심을 가지고 리그의 한 시즌 결과를 예측하며 모든 팀의 결과를 찾아보는 시청자들 또한 적지 않다. 전통적으로 인기를 끄는 축구, 야구, 농구가 아니더라도 League Of Legend와 같은 e-스포츠 또한 스포츠 경기로 인정받으며 많은 인기를 끌고 있다. 이러한 사회 현상에 맞추어 공급자들은 더 빠르게, 더 많은 스포츠 하이라이트 콘텐츠를 제작하여 이익을 창출해야 한다.

하지만 이렇게 제공해야 할 콘텐츠는 점점 증가하는 반면에 투입될 수 있는 콘텐츠 제공자의 인력은 한정적이다. 현재 스포츠 하이라이트 영상이 제작되기까지 모든 단계는 수작업으로 이루어지고 있고, 주요 영상 구간을 선정하고 편집하는 과정에서 시간적, 경제적 소모가 상당하다. 통상적으로 10분짜리 동영상을 편집하는데 편집 시간은 평균 4시간에서 그 이상의 시간이 걸린다고 한다.

이에 우리는 로봇 기자 '사커봇'에서 아이디어를 얻게 되었다. 인간이 작성한 기사를 딥러닝을 통해 학습하여 로봇 알고리즘이 기사를 써서, 데이터 수집부터 최종 기사 생성까지 전 과정을 지연 없이 처리할 수 있어 빠르고 정확하게 정보를 제공할 수 있다는 점에서 영감을 받았다. 방대한 양의 데이터를 보다 빠르게 처리할 수 있는 AI 기술이 필요하다고 생각했다.

본 연구는 딥러닝을 이용해 스포츠, 그중에서도 축구의 경기 하이라이트 영상 제작 과정을 자동화하는 것을 목표로 하였다. 전 세계인의 스포츠라고 불리는 축구 경기는 우리나라에서도 최고의 인기를 끌고 있다. 또한 한 종목당 즐기는 리그가 1~2개로 한정된 야구, 농구와 달리 국가대표 경기, 케이리그, EPL, 분데스리가 등 다양한 리그의 많은 경기들이 소비되고 있다. 때문에 축구 경기의 하이라이트 영상 제작을 자동화한다면 영상 편집자 및 제공자에게 적지 않은 효율성의 증가가 될 것이다.

딥러닝 네트워크를 통해 스포츠 전체 경기 영상에서 하이라이트 이벤트 시점을 자동으로 찾아서 하이라이트 영상 생성까지 만들어지도록 했다. 이를 통해 제공자들이 인력과 시간을 효율적으로 활용할 수 있도록 하고, 더 나아가 일반 사용자들 또한 자신만의 선호 하이라이트 영상을 만들 수 있도록 했다.

II. 관련 연구

현재 증가하고 있는 수요에 따라 긴 영상을 짧게 요약하는 하이라이트 추출 기술들의 많은 연구가 진행되고 있다.

Seonghun Yoon은 e-sport의 종목인 'League of Legends' 경기 영상의 자동 하이라이트 추출을 위해 이미지 혹은 텍스트 등 한 가지 데이터 정보만 사용하는 대신, 영상 이미지와 오디오 데이터 두 가지를 사용하여 하이라이트를 추출하는 오토인코더 기반의 기술을 제안하였다.[1]

Dongmahn Seo는 모바일 환경에서 스포츠 경기를 관전할 수 있는 실시간 중계 시스템과 하이라이트 동영상 서비스와 소셜 미디어 요약 서비스를 제안하였다.[2] 실시간성의 장점이 있는 중계 시스템은 전체적인 경기의 흐름보다 경기의 핵심 내용에만 국한되는 정보만을 제공하는 단점이 있어 그것을 개선하기 위해 소셜 네트워크 서비스와 문자 중계 시스템을 함께 사용하여 풍부한 정보를 통합하는 새로운 형태의 시스템을 설명하였다.

Hansol Lee는 e-sports나 야구 경기 영상의 하이라이트 검출을 위해 하나의 이벤트 구간을 여러 개의 소 구간으로 나누어 하이라이트 결정에 영향을 많이 미치는 구간 내의 특징들을 최대로 추출하기 위한 앙상블 모델을 제안하였고 앙상블 모델의 학습에 필요한 효과적인 데이터

수를 설명하였다.[3]

그리고 스포츠 영상에서 자동으로 하이라이트를 예측할 수 있는 모델을 제안하였다.[4] 이미지 정보만을 활용하는 다른 모델들과는 달리 관중들의 호응과 해설자의 목소리 크기도 경기의 이해를 돕는다는 점을 활용해 영상의 흐름을 파악하면서 오디오와 이미지 정보를 활용하는 B-MTIM을 설명하였고 추가로 특징벡터를 추출하기 위해 GAN을 결합하는 방법을 설명하였다.

III. 시스템 모델

하이라이트를 자동으로 검출하기 위해 이미지 정보와 오디오 정보를 사용하는 두 개의 모델을 제안한다. 이미지 정보를 이용해 Action spotting 으로 4개의 Highlight Event를 검출하고, 오디오 볼륨 크기를 이용해 Highlight 장면을 검출할 수 있다.

3.1. 데이터 셋 구축

우선 경기에서 각 이벤트를 학습시키기 위해 데이터 셋을 구축하였다. 초기에 우리는 기존에 축구 경기영상에서 이벤트 발생 지점과 이벤트 종류를 검출하기 위해 만들어진 SoccerNet 데이터 셋을 이용했다.[5] 해당 데이터 셋은 유럽의 7개 리그(Europe_europa-leage, England_epl, France_ligue-1, Germany_bundesliga, Europe_uefa-champions-league, Italy_serie-a, spain_laliga)에 대해서 500개의 경기영상으로 구성되어있다. 각 경기 영상마다 Goal, Substitution, Card 3개의 이벤트 클래스에 대해 annotation 되어 있으며 모두 6000여개의 annotation 이 있다. 이벤트에 대해 각 annotation 형식은 클래스 명, 이벤트 발생 팀, 이벤트 발생 시작 시간으로 구성된다. 이벤트 발생 시작 시간은 FIFA의 이벤트 발생 정의를 참조하였다. 골은 골라인을 통과한 순간, 선수 교체는 선수가 경기장을 들어온 순간, 경고는 심판이 카드를 꺼내드는 순간이다.

우리는 3개의 이벤트 클래스만이 하이라이트 영상을 구성하기 부족하다 생각하여 추가적인 이벤트로 코너킥을 학습시키기로 결정했다. 코너킥은 세트피스 상황이고, 또 곧바로 골로 자주 연결되는 주요 상황이기에 하이라이트 장면으로 선정하였다. 우리는 새롭게 2020년도 K리그 경기영상 50개에 대해 4개의 이벤트 클래스(골, 선수 교체, 경고, 코너킥)를 뽑아 1500여개의 annotation을 생성하여 자체 데이터 셋을 구축했다.

3.2. 이미지 정보를 이용한 하이라이트 추출 모델

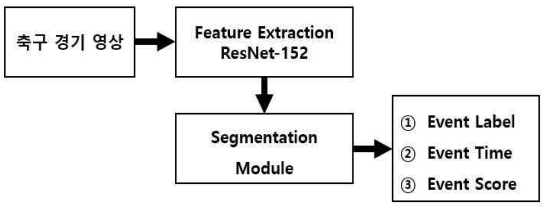


그림 4. 이미지정보를 이용한 하이라이트 추출 모델

이미지 정보를 이용한 하이라이트 추출을 위해 기존 신경망에서 이미지 처리 방식인 Convolutional Neural Network(CNN)을 사용한다. Action Spotting Feature 추출을 하기 때문에, Kinetic data 특성에 맞는 ResNet-152 네트워크를 사용한다. 경기 영상으로부터 ResNet-152 계층을 거쳐 나온 특징들을 이용해 Segmentation Module을 수행한다. Segmentation Module은 Feature Extraction 모듈에서 ResNet을 사용하여 추출된 특징들을 입력 값으로 가지며, 각 event 별로 Action spotting 을 통해 각 장면의 label, 이벤트 발생 시간, 이벤트 score 를 추출한다.

3.3 오디오 정보를 이용한 하이라이트 추출 모델

시스템의 정확도 향상을 위해 오디오 볼륨 분석 기능을 사용했다. 정해진 이벤트만을 찾아 하이라이트로 포함하는 방식에서는, 예기치 못한 돌발 상황으로서 이루어진 하이라이트를 포함하지 못한다. 스포츠 경기에서는 보통 중요한 부분이 된다면 관중의 환호 혹은 야유 소리가 커지기 때문에 이를 활용했다.

전체 축구 경기 영상에서 오디오를 추출하여 볼륨의 크기를 수치화해 행렬로 만들었다. 그 후 행렬내의 시퀀스에서 볼륨이 급격히 커지는 구간을 찾아냈다. 이 볼륨의 증가 수치가 일정 임계 값을 넘는다면 하이라이트로 판단했다. 이렇게 찾아낸 하이라이트 시점이 이벤트 검출로 판별된 시점에 포함되지 않는다면 추가하여 하이라이트 영상이 제작되도록 한다.

3.4. 전체 하이라이트 추출 시스템

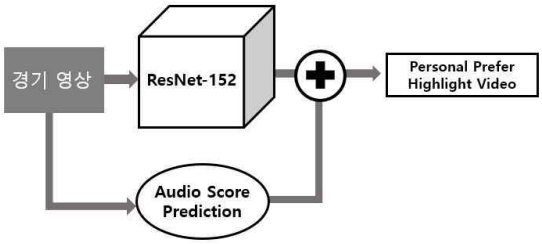


그림 5. 전체 하이라이트 추출 시스템

3.2의 이미지 정보를 이용한 하이라이트 생성 모델과 3.3의 오디오 정보를 이용한 하이라이트 생성 모델의 결과로 나온 하이라이트 이벤트 발생 시간들의 정보를 병합해 전체 하이라이트 영상을 제작한다.

Ⅳ. 시스템 구현 및 결과 화면

프로그램을 실행했을 시 User Interface 이다. 그림 6과 같이, 사용자는 하이라이트 생성을 원하는 영상을 선택한다. 그 후에 원하는 하이라이트 장면 Goal, Card, Substitution, Cornerkick, 오디오 하이라이트를 체크박스를 통해 선택하면 해당 하이라이트 장면들만 추출되어 하이라이트 영상이 제작된다. 생성된 하이라이트 영상은 그림 6의 축구 이미지가 나타나있는 부분에 재생되어 확인할 수 있다. 저장하기 버튼을 눌러 생성된 하이라이트 영상을 저장하고, 다시하기를 눌러 다른 경기영상에 대해 하이라이트를 생성할 수 있다.

그림 6. 프로그램 실행 화면

Ⅴ. 실험 및 결과

실험을 위해 2020년도 K리그 경기 영상 5개를 수집하였다. 실험 시간을 단축하기 위해 모든 경기 영상에서 Feature Extraction 모듈을 통해 미리 특징 벡터를 추출하여 실험을 진행하였다.

5.1. 평가 방법

하이라이트 추출 모델의 성능 평가를 위해, 경기 영상에서 수작업으로 하이라이트 영상에 대해 레이블링(Labeling) 작업을 하였다. 이를 통해 레이블링 된 경기 영상과 본 논문의 모델이 만든 하이라이트 영상의 각 이벤트 별 시작 시간, 장면 개수 등을 평가해 정확도를 측정하였다.

5.2. 실험 평가

표 1은 축구 영상에서 C3D와 ResNet-152 두 개의 네트워크에 대해 실험 한 결과 이다. 각 네트워크 별로 Mean Pooling, Max Pooling, NetVLAD 3가지의 Pooling 방법을 적용해 정확도를 측정하였다. Goal,Card, Substitution, Cornerkick 4가지의 이벤트에 대해 test 경기 5개를 통해 실험을 진행하였다. 가장 성능이 좋았던 ResNet-152 와 NetVLAD를 이용해 최종 하이라이트 생성 모듈에 사용하였다.

표 1. 축구 영상에서 네트워크 별 Pooling 방법에 따른 정확도 비교 실험 결과 (mAP)

	C3D	ResNet -152
Mean Pooling	34.7	35.3
Max Pooling	42.4	43.2
NetVLAD	48.9	51.2

Learning rate : 0.01, Epochs: 20, Training Dataset 40경기, Validation Dataset 5경기, Test Dataset 5경기에 대해 실험을 진행 하였다.

표 2. 기존 SoccerNet과 본 논문의 모델 성능 비교

	SoccerNet	DaSH
Class	3개	4개
Accuracy	0.647	0.412
Epochs	200	20
Goal Acc	0.643	0.509
Card Acc	0.556	0.503
Substitution Acc	0.744	0.512
Cornerkick class	x	0.524

표 2는 기존의 SoccerNet의 축구 하이라이트 장면 추출기와 본 논문의 하이라이트 생성기를 비교하였다. class 수는 기존 SoccerNet은 3개, DaSH는 4개의 class 에 대해 하이라이트 생성이 가능하다. 각 클래스 별 정확도는 표 2를 통해 확인 할 수 있다. 본 논문의 하이라이트 생성기(DaSH)는 적은 데이터 셋과 비교적 부족한 학습량에도 불구하고 유의미한 성능을 달성하였다.

VI. 결론

우리의 연구는 스포츠 하이라이트 영상 제작, 그 중에서도 축구에 대해서 모든 과정들을 자동화 할 수 있게 했다. 기존의 수작업으로 영상을 제작할 때는 1시간 분량 영상 편집이 평균 5시간가량 소요되고, 상당한 수준의 전문 인력들이 필요했다. 이에 반해 DaSH는 축구 하이라이트 영상 제작에 있어 시간 소모가 줄어들었으며, 인력 비용이 상당히 절감된다는 이점이 있다.

현재 우리의 연구는 축구라는 종목에 한해 4개의 이벤트와 오디오 정보만을 포함한다. 그러나 데이터 셋을 구축함에 따라 여러 다른 종류에 스포츠에 적용이 가능하다. 또한 축구에 대해서도 다양한 이벤트들을 지정해 검출하도록 할 수 있다. 뿐만 아니라 이벤트 종류에 대한 명확한 명시와 오디오 정보를 이용해 하이라이트를 생성할 수 있어 사용자들이 자신이 원하는 이벤트 장면만을 모아 소장할 수 있다는 점에서 차별성이 있다.

참 고 문 헌

[1] 윤성훈, 이승진, 김경중, 2018, “오토인코더 모델을 이용한 리그 오브 레전드 게임 동영상 하이라이트 추출”, 한국정보과학회

[2] 서동만, 김수현, 박호건, 고희동, 2012, “소셜 미디어와 중계영상을 활용한 실시간 문자 중계 시스템”, 한국정보과학회

[3] 이한솔, 이계민, 2020, “하이라이트 검출을 위한 구간 분할 앙상블 모델”, 방송공학회논문지

[4] 이한솔, 이계민, 2019, “GAN을 이용한 하이라이트 영상 예측 모델의 성능 개선”, 한국방송미디어공학회

[5] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, Bernard Ghanem, 2018, “SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos”, CVPR

[6] Anthony Cioppa, Adrien Deli`ege, Silvio Giancola, 2020, “A Context-Aware

Loss Function for Action Spotting in Soccer Videos", CVPR
[7] Rockson Agyeman, Rafiq Muhammad and Gyu Sang Choi, 2019, "Soccer Video Summarization using Deep Learning", MIPR
[8] Hao Tang, Vivek Kwatra, Mehmet Emre Sargin, Ullas Gargi, 2011, "DETECTING HIGHLIGHTS IN SPORTS VIDEOS: CRICKET AS A TEST CASE", IEEE
[9] Jun-Ting (Tim) Hsieh, Chengshu (Eric) Li, Wendi Liu, 2017, "Spotlight: A Smart Video Highlight Generator", CS231n
[10] Kun Liu, Wu Liu, Chuang Gan, Mingkui Tan, Huadong Ma, 2018, "T-C3D: Temporal Convolutional 3D-Network for Real-Time Action Recognition", AAAI
[11] Kensho Hara, Hirokatsu Kataoka, Yutaka Satoh, 2018, "Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?", CVPR

==========저자소개==========

이름: 노신영 1997년 5월 26일생
2021년 2월 : 한국항공대학교
소프트웨어학과 (공학사)
관심 분야 : 머신러닝, computer vision
특 기 : 보드게임

이름: 이우리 1997년 1월 8일생
2021년 2월 : 한국항공대학교
소프트웨어학과 (공학사)
관심 분야 : 영화, 농구경기 관람
특 기 : 없는데요

이름: 윤나라 1997년 10월 8일생
2021년 2월 : 한국항공대학교
소프트웨어학과 (공학사)
관심 분야 : 머신러닝, 안드로이드
특 기 : 식물 기르기

==========

감사의 글

유독 짧게만 느껴졌던 2020년의 일 년은 저희 팀에게 큰 의미였습니다. 처음 프로젝트를 시작할 때 어디서부터 어떻게 시작할지 많이 막막했었는데 팀원들의 협력과 화합 그리고 교수님의 훌륭한 지도의 결과로 무사히 프로젝트를 마무리할 수 있었습니다.
저희 프로젝트가 방향성을 잃어갈 때 문제점을 함께 찾아주시며 아낌없는 조언을 해주신 길현영 교수님의 지도가 큰 도움이 되었습니다. 진심을 담아 감사의 말씀 전하고 싶습니다.
비록 처음 원대한 꿈을 않고 계획했던 것만큼의 완성도라고 하기엔 아쉬운 점이 있지만 그래도 마지막까지 저희는 최선을 다했기에 만족스러운 결과를 얻은 것 같아 감사합니다.
데이터 셋 구축부터 모델의 설계, 구현까지 팀원 모두가 밤잠을 줄여가며 프로젝트를 진행하느라 힘든 점이 많았지만 그런 과정에서 저희는 많이 배웠고 한층 더 성장할 수 있게 된 계기였습니다.