특집논문 (Special Paper)
방송공학회논문지 제25권 제2호, 2020년 3월 (JBE Vol. 25, No. 2, March 2020)
https://doi.org/10.5909/JBE.2020.25.2.143
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

오디오와 이미지의 다중 시구간 정보와 GAN을 이용한 영상의 하이라이트 예측 알고리즘

이 한 솔[a)], 이 계 민[a)‡]

Video Highlight Prediction Using GAN and Multiple Time-Interval Information of Audio and Image

Hansol Lee[a)] and Gyemin Lee[a)‡]

요 약

최근 다양한 매체를 통해 폭발적인 양의 콘텐츠가 업로드 되고 있으며 그 가운데 게임과 스포츠 영상은 상당한 비율을 차지한다. 방송사에서는 시청자 편의를 위해 경기 영상 중 흥미를 끄는 장면을 모아 하이라이트 영상을 만들어 제공한다. 그러나 이는 시간과 비용이 많이 소요되는 문제가 있다. 본 논문에서는 게임과 스포츠 경기에서 자동으로 하이라이트를 예측하는 모델을 제안한다. 기존의 방법들이 이미지 정보만을 주로 이용하는데 반해 우리는 오디오와 이미지 정보를 함께 사용하며, 영상의 단기적 전후관계와 중장기적 흐름을 동시에 파악하는 방법을 제시한다. 또한 더 좋은 특징벡터를 찾아내기 위해 GAN을 결합한 모델을 설명한다. 제안하는 모델들은 e스포츠 경기 영상과 야구 경기 영상을 이용하여 평가한다.

Abstract

Huge amounts of contents are being uploaded every day on various streaming platforms. Among those videos, game and sports videos account for a great portion. The broadcasting companies sometimes create and provide highlight videos. However, these tasks are time-consuming and costly. In this paper, we propose models that automatically predict highlights in games and sports matches. While most previous approaches use visual information exclusively, our models use both audio and visual information, and present a way to understand short term and long term flows of videos. We also describe models that combine GAN to find better highlight features. The proposed models are evaluated on e-sports and baseball videos.

Keyword : Video highlight, Multimodal model, GAN, Multiple time-interval model, Audio information

a) 서울과학기술대학교 일반대학원 미디어IT공학과(Dept. of Media IT Engineering, Graduate School, Seoul National University of Science and Technology)
‡ Corresponding Author : 이계민(Gyemin Lee)
E-mail: gyemin@seoultech.ac.kr
Tel: +82-2-970-6416
ORCID: https://orcid.org/0000-0001-6785-8739
※ 이 논문의 연구 결과중 일부는 한국방송미디어공학회 “2019년 추계학술대회”에서 발표한 바 있음.
※ This work was supported by National Research Foundation of Korea(NRF-2017R1E1A1A03070596).
· Manuscript received December 18, 2019; Revised February 4, 2020; Accepted February 4, 2020.

Ⅰ. 서 론

온라인에서는 현재 다양한 매체를 통해 장르를 불문하고 대량의 콘텐츠가 업로드 되고 있다. 과거와는 달리 스마트폰, 인터넷 등의 발달로 스트리밍 플랫폼 서비스의 접근이 편리해지면서 이러한 콘텐츠들이 큰 인기를 끌고 있으며, 특히 축구와 야구 같은 스포츠 경기나 게임 경기 영상은 꾸준히 높은 수요를 보이고 있다. 경기 영상은 보통 길기 때문에 시청자 편의나 네트워크 효율을 위해 방송국에서는 하이라이트 영상을 제공하는 경우가 많다. 하지만 하이라이트 영상을 제작하는 것은 전문적인 기술과 장비를 요구하기 때문에 시간과 비용 면에서 문제가 발생한다. 이에 본 논문에서는 자동으로 하이라이트를 예측하는 모델을 제안한다.

영상의 하이라이트 추출과 관련된 대부분의 연구는 영상을 이해하는데 이미지 정보만을 이용한다. 하지만 스포츠와 같은 경기 영상에서는 관중들의 호응과 해설자의 목소리 크기 등이 경기를 이해하는데 큰 도움이 된다. 따라서 우리는 오디오와 이미지 정보를 함께 사용하는 모델을 제안한다. 또한 경기 영상은 보통 한 순간의 이벤트만 봐서는 그 이벤트가 득점으로 이어지는가에 대한 판단이 어려우므로 우리는 단기적 전후관계와 중장기적 흐름을 같이 파악하는 다중 시구간 모델을 이용한다. 이 때, 우리는 Generative Adversarial Network (GAN)[1]을 이용하여 더 유용한 특징 벡터를 추출할 수 있도록 하는 모델 개선 방법을 제시한다. 제안하는 모델들은 직접 수집한 e스포츠 경기 영상과 야구 경기 영상을 이용하여 평가하였다.

Ⅱ. 관련 연구

영상을 요약하거나 하이라이트를 찾는 방법에 관한 다양한 연구들이 진행되고 있다. Zhang 등은 LSTM과 Determinantal Point Process (DPP)를 결합한 모델을 제안하였다[2]. 또한 Mahasseni 등은 CNN과 LSTM을 이용한 기본 구조에 GAN을 결합한 비지도 학습 알고리즘을 소개하였다[3]. 한편 Zhang 등은 기본적인 encoder-decoder 알고리즘에 또 다른 retrospective encoder를 추가한 계층적 구조의 모델을 설명하였고[4], Zhou 등은 모델이 다양성과 대표성을 가지는 프레임들을 선택하도록 유도하기 위해 그에 따른 reward를 부여하는 강화학습을 이용한 알고리즘을 제시하였다[5].

앞에서 나열된 연구들은 영상의 시각적 정보만을 이용한 모델들이며, 오디오 또는 텍스트 정보를 이용한 연구도 존재한다. Lee 등은 오디오와 이미지 정보를 같이 활용함과 동시에 adversarial network를 결합하는 방법을 제안하였고[6], 개인방송에서 채팅 데이터를 이용하여 오디오 정보와 함께 영상에서 하이라이트를 검출한 연구 또한 이루어지고 있다[7,8].

Ⅲ. 하이라이트 예측 알고리즘

이 장에서는 하이라이트를 자동으로 예측하기 위해 제안하는 모델들을 설명한다. 먼저 단기적 흐름과 중단기적 흐름을 동시에 이용하는 Multiple Time-Interval Model (MTIM)을 소개한다. 그 다음, GAN을 결합하는 방법을 제시하고, 오디오와 이미지 정보를 모두 사용하는 모델을 제안한다. 마지막으로 이를 모두 결합한 우리의 최종 모델 BiMTIM-GAN을 설명한다.

1. 오디오 다중 시구간 모델 MTIM

영상에서 특정 장면이 중요한 이벤트임을 판단하기 위해서는 해당 장면의 전후 상황이 어떻게 진행되는지를 보는 것이 중요하다. 이를 위해 양방향 LSTM이 많이 이용되고 있다[2]. 하지만 콘텐츠마다 중장기적 흐름이 중요한 경우가 있다. 예를 들어, 축구와 야구 같은 전통적인 경기에서 현재 선수들의 플레이가 이 후 득점으로 이어질지는 직전 직후의 동향만을 보는 것이 아니라 오래 지켜봐야 하는 경우도 있다. 이를 위해 중장기적 흐름을 파악하는 모델 MTIM을 제안한다.

그림 1(a)은 다중 시구간 모델 MTIM의 구조를 보여준다. 첫 번째 층 두 개의 LSTM 중 하나는 영상의 단기적인 전후 관계를 파악하고 나머지 하나는 영상의 중장기적 흐름을 파악하는 역할을 한다. 우선 오디오 데이터로부터 얻은 짧

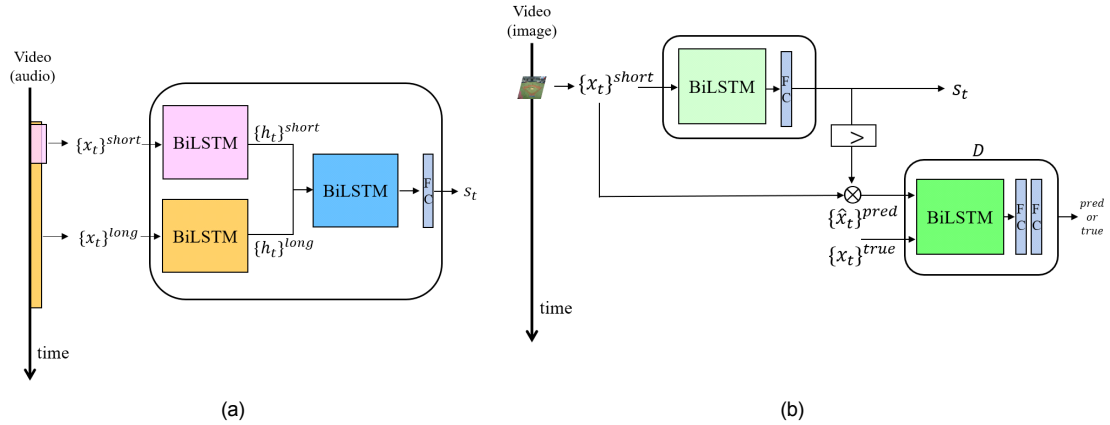


그림 1. (a) 오디오 다중 시구간 모델 (MTIM), (b) GAN을 이용한 하이라이트 특징 추출 (STIM-GAN)
Fig. 1. (a) Multiple time-interval audio model (MTIM), (b) Highlight feature extraction using GAN (STIM-GAN)

은 구간(1초)에 해당하는 특징벡터x^{short}와 긴 구간에 해당하는 특징벡터x^{long}를 각각의 LSTM에 넣어 h^{short}와 h^{long}을 구한다. 두 결과를 함께 두 번째 층의 LSTM과 FC layer의 입력으로 주어 하이라이트 스코어 s_t를 얻는다. 이때의 LSTM은 서로 다른 시구간에서 나온 h^{short}와 h^{long}을 하나로 합쳐주는 역할을 한다. 결국 MTIM은 손실함수

$$L_{CE} = \frac{1}{T}\sum_t \text{cross-entropy}(s_t, y_t) \quad (1)$$

을 최적화 하며, 위 식에서 y_t는 ground truth label을 의미한다. 최종 하이라이트는 스코어 s_t가 높은 프레임들을 모아서 만들 수 있다.

2. GAN을 이용한 하이라이트 특징 추출 STIM-GAN

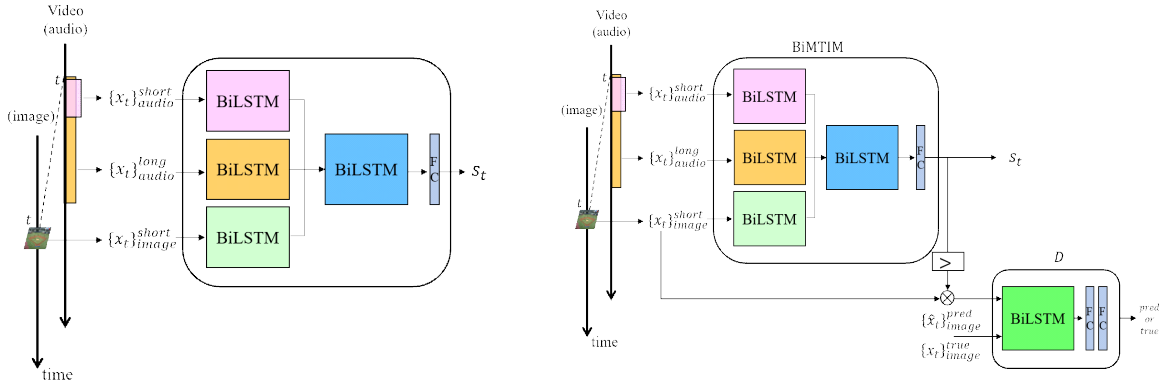
더 나은 하이라이트를 만들기 위해서, 모델은 주요 장면의 특성을 잘 나타내는 좋은 특징 벡터를 추출해 낼 수 있어야한다. 우리는 이러한 기능을 향상시키고자 GAN을 결합한 모델 STIM-GAN을 제안한다. GAN은 generator와 discriminator로 이루어진 알고리즘으로, generator는 discriminator를 속이기 위해 학습이 될수록 실제와 매우 유사한 가짜 데이터를 생성한다. 반면에 discriminator는 generator가 생성한 가짜 데이터와 실제 데이터를 정확히 구분하기 위해 학습하며 generator와 대립 관계를 가진다.

우리의 모델은 generator 대신 본 논문에서 제안하는 하이라이트 예측 모델들을 이용한다. 그림 1(b)는 GAN이 결합된 Single Time-Interval Model (STIM)을 보여준다. 여기서 STIM은 LSTM과 FC layer로 이루어진 기본 예측 모델이다. GAN의 discriminator는 실제 하이라이트와 모델이 만들어낸 하이라이트를 구별하면서 더 중요한 특징을 찾아낼 수 있도록 돕는다. x^{true}는 실제 하이라이트이고 x^{pred}는 STIM으로부터 얻어진 s_t로 선택된 프레임을 나타낸다. STIM은 discriminator D가 잘못 예측하도록 최대한 ground truth와 유사한 하이라이트를 생성한다. 따라서 우리의 모델은 다음의 최적화 문제를 푼다.

$$\min_{STIM} \max_D L_{CE} + \log D(x^{true}) + \log(1 - D(x^{pred})) \quad (2)$$

3. GAN을 이용한 오디오/이미지 다중 시구간 모델 BiMTIM-GAN

이 절에서 우리는 오디오와 이미지 정보를 같이 활용하는 방법을 설명하고 GAN과 결합하여 성능을 향상시킨 모델을 제시한다. 그림 2(a)에 있는 BiMTIM은 MTIM을 확장시킨 구조를 가지며, 짧은 구간의 오디오 특징벡터x_{audio}^{short}, 긴 구간의 오디오 특징벡터x_{audio}^{long}, 그리고 이미지 특징벡터 x_{image}^{short}가 각각의 LSTM을 통과한 후 결합된다. 이어서 두 번째 층의 LSTM를 거쳐 하이라이트 스코어 s_t를 만든다.



(a) (b)

그림 2. (a) 다중 시구간 오디오/이미지 모델(BiMTIM), (b) GAN을 이용하여 확장시킨 최종 모델 (BiMTIM-GAN)
Fig. 2. (a) Multiple time-interval audio/image model (BiMTIM), (b) the proposed GAN-extended model (BiMTIM-GAN)

그림 2(b)에 GAN을 이용한 우리의 최종 모델 BiMTIM-GAN의 구조가 있다. 최종 모델은 위의 STIM-GAN에서 STIM 대신에 BiMTIM을 사용한다. Discriminator D는 앞에서 설명한 것과 마찬가지로 BiMTIM에 의해 얻은 s_t로 선택된 프레임 x^{pred}와 ground truth에 해당하는 프레임 x^{true}을 구분하도록 학습한다. BiMTIM은 discriminator D가 예측된 하이라이트와 실제 하이라이트를 구별해 내지 못하도록 보다 나은 특징벡터를 추출하여 더욱 실제와 근접한 하이라이트를 만든다. 이를 위해, BiMTIM-GAN은 앞에서와 비슷한 다음의 최적화 문제를 푼다.

$$\min_{BiMTIM} \max_{D} L_{CE}+\log D(x^{true})+\log(1-D(x^{pred})) \quad (3)$$

Ⅳ. 실험 및 결과

제안한 모델들을 평가하기 위해 우리는 Twitch[9]와 Kakao TV[10]에서 각각 e스포츠와 야구 경기영상을 직접 수집하였다. 실험 데이터는 사전에 특징벡터를 추출한 후 이용하였다. 정해진 구간(1초 등) 단위로 데이터를 나눈 다음, 오디오는 Mel Frequency Cepstal Coefficient (MFCC)를 이용하여 각 구간 별로 특징벡터 x_{audio}을 추출하였다. 본 실험에서는 40ms에서 추출한 20차원의 MFCC 특징벡터 25개를 결합하여 1초에 500차원을 가지는 특징벡터 x_{audio}를 만들었다. 이미지는 ImageNet[11]에 사전 학습된 ResNet-34[12]를 이용하여 초당 1프레임에서 512차원의 특징벡터 x_{image}을 추출하였다.

우리는 정량적 평가를 위해 F-score를 활용하였다. F-score는 비디오 요약에 많이 사용되며 정밀도(precision)와 재현율(recall)의 조화평균으로 구할 수 있다.

$$P = \frac{|H_{gt} \cap H_{pred}|}{|H_{pred}|}, \quad R = \frac{|H_{gt} \cap H_{pred}|}{|H_{gt}|} \quad (4)$$

$$F\text{-}score = \frac{2PR}{P+R} \times 100\,\% \quad (5)$$

위 식에서 H_{gt}와 H_{pred}는 각각 ground truth와 모델에 의해 예측된 결과이다.

본 실험에서는 제안한 모델들과의 성능 비교를 위해서 2개의 FC layer를 가지는 간단한 MLP모델을 구현하였다. MLP모델은 이벤트들의 전후관계를 파악하지 않고 각 구간의 정보만으로 하이라이트 스코어를 만든다. 각 layer의 크기는 입력 벡터의 크기와 동일하게 구성하였다.

1. e스포츠 데이터

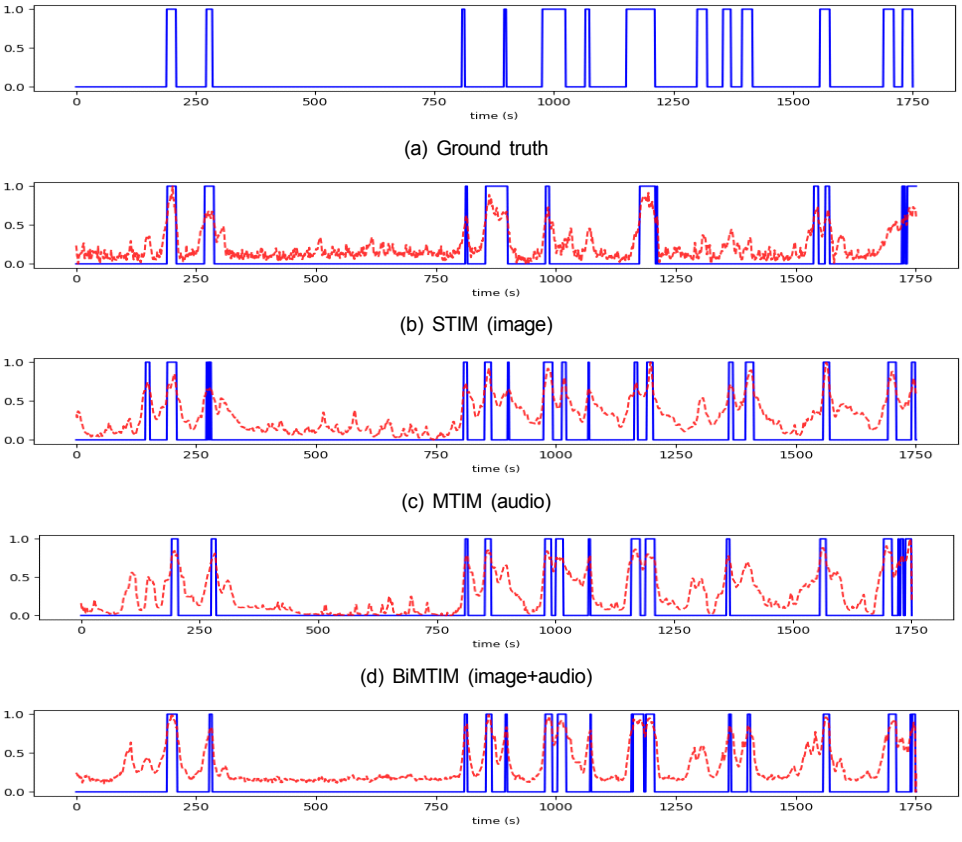
2017년에 Twitch에서 중계된 'League of Legends' 대회 5개(IEM World Championship Katowice 2017, 2017 LoL

표 1. e스포츠와 야구경기 데이터 요약 정보
Table 1. Summary of e-Sports and baseball data sets

Type	Statistics	Video length (sec)	Length of highlights (sec)	Highlight ratio (%)
e-Sports	mean (±std)	2,096.76 (±599.10)	213.27 (±70.99)	10.55 (±3.78)
	max	4,785	469	22.30
	min	1,483	146	9.84
Baseball	mean (±std)	12,175.39 (±1,176.13)	599.25 (±225.34)	4.95 (±1.93)
	max	14,866	1,361	12.59
	min	9,909	76	0.61

World Championship, LoL All Star 2017, 2017 LoL Champions Korea Spring, 2017 LoL Champions Korea Summer)에서 수집한 63개의 경기 영상으로 모델의 성능을 평가하였다. 이 가운데 7개의 경기 영상을 테스트 데이터로 사용하였고 나머지 경기 영상을 학습에 이용하였다. 모든 경기 영상에 대한 ground truth는 e스포츠 전문 채널 OGN[13]에서 제공하는 하이라이트 영상을 활용하였다. 표 1은 데이터에 대한 세부사항을 보여준다. 영상의 평균 길이는 30분, ground truth는 약 3분으로 전체 길이의 10% 비율이다. 본 실험에서는 테스트 영상의 10%를 하이라이트로 선택하였다. 짧은 구간은 1초, 긴 구간은 30초를 기준으로 실험을 진행하였다.

그림 3은 테스트로 사용된 경기 영상 중에서 한 영상에 대한 결과를 시각적으로 보여준다. 그림에서 파란 선은 하이라이트에 해당하면 1, 그렇지 않은 부분은 0으로 구분하여 하이라이트의 유무를 나타내고 있으며, 빨간 점선은 하



(a) Ground truth

(b) STIM (image)

(c) MTIM (audio)

(d) BiMTIM (image+audio)

(e) BiMTIM-GAN (proposed)

그림 3. e스포츠 영상에 대한 모델별 실험 결과 (파란 실선: 하이라이트 구간, 빨간 점선: 하이라이트 스코어)
Fig. 3. Experiment results on an e-Sports test video (blue: highlight labels, red: highlight score)

이라이트 score를 표시한다. 이미지 정보만을 사용하는 STIM은 750초 이후의 결과가 ground truth와 매우 다름을 한눈에 알 수 있다. 오디오 정보만을 사용하는 MTIM은 120초 구간에서 잘못 예측하였다. 이들과 달리 오디오와 이미지 정보를 모두 사용하는 BiMTIM과 BiMTIM-GAN은 ground truth와 매우 비슷한 결과를 보이고 있다.

표 2는 실험에 대한 정량적 결과를 나타낸다. MLP 결과를 보면 50 전후의 F-score로 가장 낮은 성능을 보이는데 이는 이벤트의 전후관계를 파악하는 것이 하이라이트 예측에 중요하다는 것을 보여준다. 단기적 전후관계에 집중하는 STIM은 이미지 정보를 사용할 때 66.55, 오디오 정보를 사용할 때 68.94를 가지는 반면에 중장기적 흐름을 함께 파악하는 MTIM은 70.65로 STIM보다 더 높은 F-score를 가진다. 따라서 단기적 전후관계와 중장기적 흐름을 같이 고려하는 것이 영상을 이해하는데 효과적임을 확인 할 수 있다. 특히 오디오와 이미지 데이터를 모두 사용하는 BiMTIM은 73.46을 가지며 하나의 정보만을 사용하는 모델들 보다 더 높은 F-score를 가진다. 이 결과는 하나의 정보만을 사용하는 것보다 다중 정보를 사용하는 것이 영상을 이해하는데 필요한 정보와 특징을 더 많이 획득하므로 하이라이트를 예측하는데 보다 유용한 것으로 보인다. GAN을 결합한 우리의 최종 모델 BiMTIM-GAN은 74.15로 가장 우수한 결과를 가진다. 즉, GAN을 통해 하이라이트 예측 모델이 더 좋은 특징 벡터를 찾게 되어 모델의 성능이 향상된 것으로 볼 수 있다.

2. 야구 경기 데이터

2018년 4월부터 5월 초까지 기간 중에 Kakao TV에서 중계된 한국 프로 야구 경기영상 28개를 이용하여 모델을 평가하였다. 이 중 5개의 경기 영상을 테스트 데이터로 이용하였고 ground truth는 Naver-sports[14]에서 제작한 하이라이트 영상을 활용하였다. 데이터에 대한 세부사항은 표 1에 나타내었다. 야구 경기의 전체 길이는 평균 3시간 20분, ground truth의 평균 길이는 약 600초로 이는 전체 경기 영상의 대략 평균 5% 비율이며 실험에서도 전체 영상 길이의 5%를 하이라이트로 검출하였다. 짧은 구간은 1초로, 긴 구간은 2분을 기준으로 실험을 진행하였다. 이는 야구 같은

표 2. e스포츠와 야구 데이터에 대한 실험 결과 (F-score)
Table 2. Experiment results(F-score) on e-sports data and baseball data sets

Data type	Model	e-Sports (%)	Baseball (%)
Image	MLP	52.12	24.18
	STIM	66.55	53.65
	STIM-GAN	69.28	57.56
Audio	MLP	48.94	26.45
	STIM	68.94	55.33
	MTIM	70.65	57.57
Image + Audio	MLP	53.88	20.58
	BiMTIM	73.46	61.90
	BiMTIM-GAN	74.15	63.57

전통적인 스포츠의 경우 사람이 직접 움직이는 경기이므로 한 이벤트에 대한 과정이 e스포츠 경기에 비해 길다는 것을 고려한 선택이다.

그림 4는 제안하는 모델들의 일부 결과(5000~7000초)를 시각적으로 보여준다. STIM은 5700초와 6400초 부분의 하이라이트를 잘못 예측 하였으며, MTIM은 6250초에서 6400초 구간을 예측하지 못하였다. 반면에 오디오와 이미지 정보를 모두 이용한 BiMTIM은 위의 모델들이 잘못 예측한 6250초에서 6400초 구간을 제거하였다. 그리고 GAN을 결합한 BiMTIM-GAN은 전체적으로 ground truth와 가장 근접한 결과를 보이며, 특히 5900초에서 6100초 구간을 다른 모델들에 비해 가장 잘 예측하였다.

표 2를 보면, 우선 MLP모델에 대한 결과는 이미지와 오디오 정보를 다 사용하여도 F-score가 30을 넘지 못한다. e스포츠 보다 야구 경기의 길이가 더 길기 때문에 야구 경기의 하이라이트를 예측하는 것이 더 어려움을 보여준다. STIM은 이미지 정보를 사용할 경우와 오디오 정보를 사용할 경우 각각에서 53.65와 55.33을 가진다. 반면에 MTIM은 57.57로 STIM보다 높은 F-score를 갖는다. 더 나아가 BiMTIM은 61.90으로 하나의 정보만을 이용하는 모델들의 F-score보다 훨씬 우수한 결과를 보인다. 이 결과는 e스포츠와 마찬가지로 다중 정보를 사용하는 것이 더 풍부한 정보와 특징을 확보하면서 하이라이트를 예측하는데 보다 유용한 것으로 해석할 수 있다. 그리고 최종 모델 BiMTIM-GAN은 63.57로 가장 높은 F-score를 갖는다. 따라서 GAN

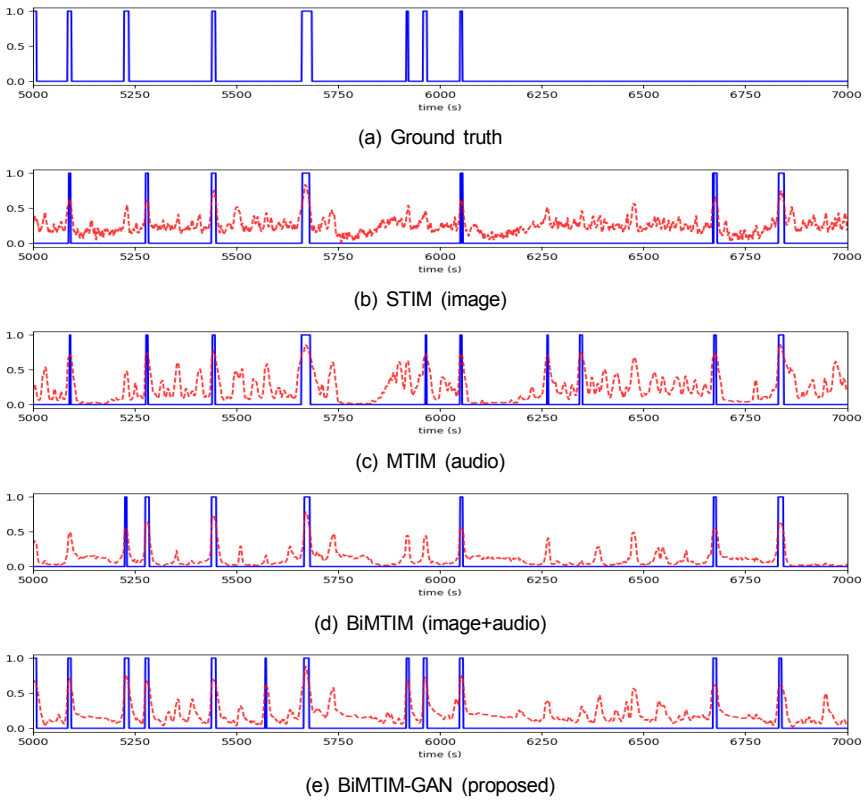


그림 4. 야구 영상에 대한 모델별 실험 결과 (2000~4000초, 파란 실선: 하이라이트 구간, 빨간 점선: 하이라이트 스코어)
Fig. 4. Experiment results on a baseball video (2000~4000sec, blue: highlight labels, red: highlight score)

이 모델이 더 좋은 특징을 찾아낼 수 있도록 도움을 줌으로써 성능 향상에 효과적임을 알 수 있다.

V. 결 론

본 논문에서는 콘텐츠의 단기적 흐름과 중장기적 흐름을 함께 파악하는 MTIM을 제안하였고 영상을 이해하는데 다중 시구간 정보가 도움이 된다는 사실을 보였다. 또한 오디오와 이미지 정보를 함께 활용하여 보다 풍부한 정보와 특징을 확보하는 하이라이트 예측 모델을 설명하였고 실험을 통해 성능이 향상되었음을 확인하였다. 특히 우리의 최종 모델은 다중 데이터와 GAN을 모두 결합한 구조를 가지며 다른 모델들과 비교하였을 때, 가장 높은 성능을 보임을 정량적 결과와 시각적 비교로 확인하였다.

제안된 모델들은 오디오와 이미지 정보만을 사용하는데, 개인방송 플랫폼의 경우에는 채팅 데이터를 획득할 수 있기 때문에 채팅 데이터까지 이용하는 모델이 더 높은 성능을 가질 것으로 기대할 수 있다. 또한 긴 구간의 경우 영상의 장면 전환을 고려하여 특징벡터를 추출한다면 보다 향상된 결과를 얻을 수 있을 것이라 예상한다.

참 고 문 헌 (References)

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” In NIPS, pp. 2672-2680, 2014, http://papers.nips.cc/paper/5423-generative-adversarial-nets.

[2] K. Zhang, WL. Chao, F. Sha, and K. Grauman, “Video Summarization with Long Short-term Memory,” European Conference on Computer Vision, Amsterdam, Netherlands, pp. 766-782, 2016, doi:10.1007/978-3-319-46478-7_47.

[3] B. Mahasseni, M. Lam, and S. Todorovic, “Unsupervised Video Summarization with Adversarial LSTM Networks,” The IEEE

Conference on Computer Vision and Pattern Recognition, pp. 2982-2991, 2017, doi: https://doi.org/10.1109/cvpr.2017.318.
[4] K. Zhang, K. Grauman, and F. Sha, “Retrospective Encoders for Video Summarization,” In ECCV, pp. 383-399, 2018, doi: https://doi.org/10.1007/978-3-030-01237-3_24.
[5] K. Zhou, Y. Qiao, and Tao Xiang, “Deep Reinforcement Learning for Unsupervised Video Summarization with Diversity-Representativeness Reward,” In Thirty-Second AAAI Conference on Artificial Intelligence, pp. 7582-7589, 2018, https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewPaper/16395.
[6] H. Lee, G. Lee, “Summarizing Long-Length Videos with GAN-Enhanced Audio/Visual Features,” In ICCV workshop, 2019.
[7] E. Kim, G. Lee, “Highlight Detection in Personal Broadcasting by Analysing Chat Traffic : Game Contests as a Test Case," Journal of Broadcast Engineering, Vol. 23, No. 2, pp. 218-226, 2018, doi: http://dx.doi.org/10.5909/JBE.2018.23.2.218.
[8] E. Kim, G. Lee, “Video Highlight Prediction Using Multiple Time-Interval Information of Chat and Audio,” Journal of Broadcast Engineering, Vol. 24, No. 4, pp. 553-563, 2019, https://doi.org/10.5909/JBE.2019.24.4.1.
[9] Twitch, https://www.twitch.tv/ (accessed Dec. 23, 2019).
[10] Kakao TV, https://tv.kakao.com/ (accessed Dec. 23, 2019).
[11] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet Classification with Deep Convolutional Neural Networks,” In NIPS, 2012, doi: https://doi.org/10.1145/3065386.
[12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” In CVPR, pp. 770-778, 2016, doi: https://doi.org/10.1109/cvpr.2016.90.
[13] OGN, http://ogn.tving.com/ (accepted Dec. 23, 2019).
[14] Naver-sports, https://sports.news.naver.com/(accepted Dec. 23, 2019).

저 자 소 개

이 한 솔

- 2019년 : 서울과학기술대학교 전자IT미디어공학과 학사
- 2019년 ~ 현재 : 서울과학기술대학교 일반대학원 미디어IT공학과 석사과정
- ORCID : https://orcid.org/0000-0002-1127-976X
- 주관심분야 : 머신러닝, 딥러닝, 신호처리

이 계 민

- 2001년 : 서울대학교 전기공학부 학사
- 2007년 : University of Michigan EECS 석사
- 2011년 : University of Michigan EECS 박사
- 2011년 ~ 2012 년 : University of MichiganResearch Fellow
- 2013년 ~ 현재 : 서울과학기술대학교 전자 IT 미디어공학과 부교수
- ORCID : https://orcid.org/0000-0001-6785-8739
- 주관심분야 : 머신러닝, 신호처리, 의료정보학