미디어 제작을 위한 씬 검출 기법

*송혁 *고민수 * **유지상
*한국전자기술연구원 **광운대학교
*{hsong, kmswqet}@keti.re.kr, **jsyoo@kw.ac.kr

Scene extraction technology on deep learning for media production

*Hyok Song **Min-Soo Ko ***Jisang Yoo
*Korea Electronics Technology Institute **Kwangwoon University

요약

인터넷 환경의 변화에 따라 텍스트 기반의 정보 전달에서 멀티미디어 기반의 스트리밍 방식으로 바뀌어가고 있다. 또한 대용량의 동영상 데이터뿐 아니라 Shorts, Clip 또는 Reels 등 다양한 방식의 동영상 형태로 배포되고 있으며 서비스 플랫폼에서는 손쉽게 편집할 수 있도록 기능을 제공하고 있다. 대용량 콘텐츠, TV, Youtue 콘텐츠를 포함하여 소용량 동영상 편집에 필요한 영상 제작 기술에서 가장 인력과 시간이 많이 소요되는 부분은 편집 단계로 딥러닝 기반 인공지능 기술을 활용하여 자동화하고 있으며 영상편집에서 가장 기본이 되는 단위인 씬검출 기법을 개발하였다. 키프레임 검출 기법과 유사도 기법을 이용하여 씬을 추출하였으며 블록 Cost Function을 이용하여 최적화하여 0.5214의 정확도를 도출하였다.

Keywords: Intelligent Contents Learning, Scene Component Analysis, Scene Recommendation, Object De-identification, Smart Video Editing.

1. 서론

5G를 비롯하여 유무선 통신 속도의 변화 및 인터넷 환경의 개선으로 인하여 과거 텍스트 기반 검색에 기반한 데이터 전송 활용 환경에서 이미지 및 동영상 기반의 스트리밍 환경으로 변화하고 있다. 이러한 변화로 인하여 인터넷에 기반한 다양한 플랫폼 역시 빅데이터에 기반한 서비스로 변화하고 있다. 이와 더불어 멀티미디어 데이터의 다양한 활용방법이 제시되면서 Youtube와 같이 대용량 동영상을 서비스하는 방식에서 벗어나 Tiktok, Instagram 또는 Youtube의 Reels, 또는 Shorts 동영상과 같이 다양한 형태의 편집을 통한 콘텐츠 활용 방향의 확대에 나서고 있다. 기존 동영상 콘텐츠는 개인의 Lifelog를 전달하는 방식에서 벗어나 비즈니스, 쇼핑 등 다양한 분야에서 새로운 시장을 창출하고 있다.

동영상을 제작·배포하기 위해서는 동영상의 취득, 수집, 선정, 편집 등의 처리가 필요하며 인공지능 기술을 통한 반자동 및 자동화가 가능하다. 기존 다양한 촬영 기법의 학습을 통하여 최적의 촬영 위치, 각도를 도출하는 방법, 수많은 영상에서 적절한 씬이나 컷을 선정하는 기술 등의 개발이 진행되고 있다. 그중 가장 인력과 시간의 투입이 많이 필요로 하는 편집 과정을 인공지능을 통하여 개선하고자 한다. 본 논문의 2장에서는 일반적인 영상 제작 기술 및 프로세스를 소개하고 세부적인 편집 절차를 도시하였다. 3장에서는 편집 기술 중 편집을 위한 가장 기본 단위인 씬을 구분하고 편집하는 기술개발 내용 및 결과를 보인다.

2. 영상 편집 기술

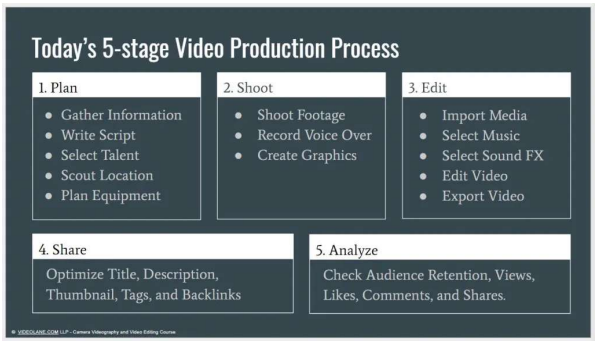

Fig 1. Video production process[1]

그림 1은 영상 제작 순서를 보인다. 제시한 단계는 기획자의 의도 및 제작 방향에 따라 달라질 수 있으나, 크게 촬영계획을 세우는 Plan 단계, 영상을 촬영하는 Shoot 단계, 촬영된 영상을 편집하는 Edit 단계, 최종 마무리 Share 단계 및 결과물의 반응을 살펴 보완하는 Analyze 단계로 구분된다. 그중 인공지능으로 접근하고자 하는 단계는 Edit 단계이다.

Edit 단계는 세부적으로 영상을 가져와 잘라내는 Trimming 단계, 잘라낸 영상을 선택하는 Shot selection 단계, 잘라낸 영상을 흐름에 맞게 선별하는 Bifurcation 단계, 다양한 영상 효과를 삽입하는 Visual effect 단계, 그리고 전체적인 컬러의 통일

및 효과를 입히는 Color correction 단계 등으로 구분할 수 있다. 본 연구에서는 Trimming, Shot selection 및 Bifurcation 단계를 위한 씬 검출 기술을 개발하였다.

3. 씬 검출 기술

비디오 콘텐츠에서 씬 단위를 분리하기 위해서 Key-frame의 이미지 Feature를 이용하여 Similarity Matrix를 생성하였다. 키프레임을 도출하기 위해서 Sequential Grouping을 이용한 Scene 단위 분리 기술을 개발하였다. 또한 다중 프레임의 Feature간 유사도 계산 기반의 화면 변환 지점을 검출하는 기술을 개발하였다.

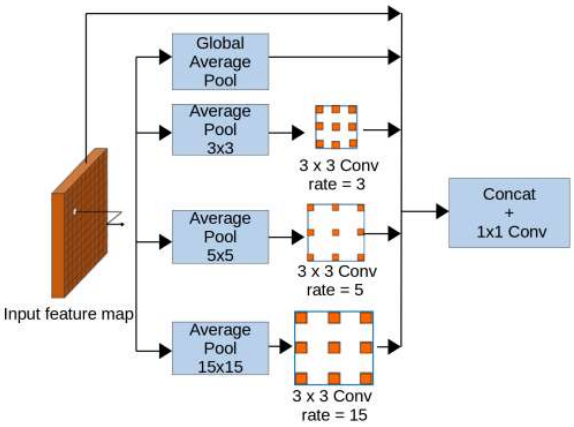

Fig 2. Vortex Pooling architecture[2]

그림 2에서 보는 바와 같이 Vortex Pooling architecture에 기반하여 학습된 EfficientNet에 기반하여 Feature Extractor를 개발하였다[3]. 딥러닝 모델의 여러 Layer에서의 Spatial feature를 융합하여 고품질의 Feature를 추출하는 결과를 보였다. 이는 여러 크기의 Dilation filter를 이용한 CNN Cell을 적용하여 추출하였다.

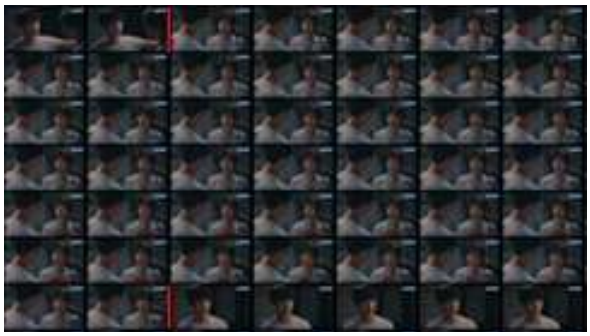

Fig 3. Single Transition detection

그림 3에서 보는 바와 같이 추출된 키프레임의 Feature와 유사한 영역에서 동일 씬으로 인식하였고 변화가 시작된 경우 Transition으로 검출한 결과를 볼 수 있다. 이는 EfficientNet, Dilation filter 및 Spatial filter를 융합하여 도출한 결과이다

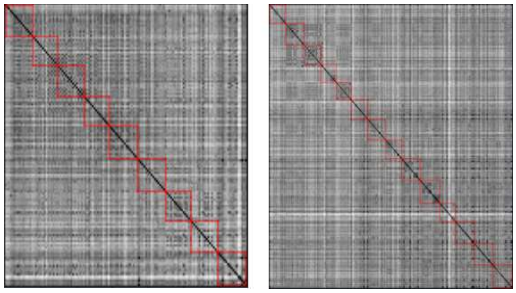

Fig 4. Sequential Grouping 결과

Similarity 값의 누적 계산을 이용한 블록 Cost function 기술을 개발하여 최적화를 통한 Scene 영역을 추출하였으며 최종 결과 표1과 같이 0.5214를 보였다. 그림 4는 Sequential Grouping 결과를 보이며 그림과 같이 중앙부에 집중된 것으로 도출되었다.

Table 1. Test result

Method	Accuracy
Ours	0.5214
OSG[4]	0.46

4. 향후 연구방향

본 연구는 영상 자동·반자동 편집 기술개발의 일환으로 영상 콘텐츠의 씬 추출, 추출된 씬 기반 장르 분석, 분석 결과에 의한 샷 도출, 편집 추천 등을 통한 자동 편집 기술 개발을 위하여 가장 기초가 된 연구이다. 본 연구 결과를 통하여 전체적인 자동편집 기술을 개발할 예정이다.

5. Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2021-0-00804, Media production technology using learning based directing methods)

References

[1] https://www.videolane.com
[2] Chen-Wei Xie, Hong-Yu Zhou, and Jianxin Wu. Vortex pooling: Improving context representation in semantic segmentation. arXiv preprint arXiv:1804.06242, 2018.
[3] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.
[4] Rotman, Daniel, Dror Porat, and Gal Ashour. "Robust and efficient video scene detection using optimal sequential grouping." 2016 IEEE international symposium on multimedia (ISM). IEEE, 2016.