

미디어 샷 검출에 관한 연구

이 형^o

^o대전보건대학교 방송콘텐츠과

e-mail: hyung@hit.ac.kr^o

A Study of Medium Shot Detection

Hyung Lee^o

^oDept. of Broadcasting Contents, Daejeon Health Institute of Technology

● 요약 ●

본 논문에서는 장편의 드라마나 영화에서 스토리 기반의 축약된 요약본을 자동으로 제작하기 위해 미디어 샷(media shot) 크기의 샷(shot)들을 추출하기 위한 방법을 고려했다. 미디어 샷 정도의 크기는 보통 인물에 중심을 둔 샷들로 인물들 간의 관계에서 특히 대사나 표정으로 내용을 전달하기 위한 목적으로 적극 권장된다. 비디오 검색을 위한 인덱싱에서 신(scene) 전환 검출 및 샷 경계 검출, 그리고 이미지에서 심도와 초점 기반의 화질 및 피사체 추출 등을 위해 전통적인 신호/영상처리 기법의 활용에서부터 최근의 기계학습 접근 등 다양한 연구들이 진행되고 있다. 영상문법에 근거하여 편집된 영상물에서 미디어 샷 정도 크기의 샷들을 추출하여 배열한다면 어느 정도 원본 내용을 충실히 전달할 수 있는 축약된 요약본을 제작할 수 있다는 가정 하에 해당 샷들을 블러(blur) 기반으로 검출하기 위해 이와 관련된 키워드들을 기반으로 기존 연구들을 살펴보고 적용 방법을 모색한다.

키워드: 미디어 샷(Medium Shot), 샷 검출(Shot Detection), 블러 검출(Blur Detection)

I. Introduction

유튜브 등에서 장편의 드라마나 영화 등의 ‘몰아보기’ 형태의 축약된 요약 편집본을 자주 접할 수 있다. 이는 원본의 저작권 문제도 있었지만 전편을 모두 시청하기에는 시간적인 측면에서 혹은 시청자들의 성향이 나 혹은 여러 가지 여건 및 상황 등에 기인한다고 볼 수 있겠다. 이러한 요약본은 재편집본으로써 비록 시간을 단축시켰지만 원본의 전달 내용을 충분히 전달할 수 있어야 한다. 결국, 원본의 내용을 충분히 반영하기 위한 샷들을 추출하여 이를 재배열함으로써 원본의 압축본을 제작해야 한다.

영상제작은 일반적으로 기획(pre-production), 제작(production), 후반작업(post-production) 등 3단계로 구분될 수 있는데, 이들 중 후반작업에서 편집을 진행하게 된다. 편집의 주요 원칙 중의 하나는 스토리의 연속성을 고려하여 샷들을 선택하고 배열하는 것으로 연속성, 복잡성, 내용 등을 충분히 고려해야 한다[1]. 물론 이러한 원칙은 절대적이기 보다는 관례적이긴 하지만, 편집의 연속성 확보는 명확한 스토리 전달에 있어서 가장 기본이라고 할 수 있으며, 이를 확보하기 위해서는 등장인물의 동일성, 배치, 벡터, 움직임, 색상, 음향 등을 고려해야 한다.

편집의 유형 중 연속편집은 주요 사건의 설명에 목적을 둔 편집방식

으로 많은 정보가 의도적으로 누락되더라도 스토리가 유연하게 진행되도록 어떤 사건의 세부 사항에서 다음으로 자연스럽게 전환이 되게 하는, 앞뒤로 이어지는 화면 내에서 시간, 장소, 내용적인 연속성이 중요시 되는 편집방법이다[2].

샷(shot)은 영화나 텔레비전 프로그램의 기본 구성 단위로 연극에서 장이나 막, 관현악에서는 파트나 바로 구분하는 것과 마찬가지로 촬영의 가장 작은 단위로 간주된다. 이러한 샷은 영상제작의 구성에서 동기, 정보, 구성, 사운드, 카메라 앵글, 연속성 등 6가지의 중요한 요소를 갖추어야 한다[3].

샷의 크기에 따른 용도 측면에서 대략 9가지 정도로 구분할 수 있는데 이들 중 미디어 샷에서는 피사체 눈의 움직임이 확실히 보이고 배경보다는 피사체에 더 주의를 가게 되어 피사체의 시선이 시청자의 시선을 끌게 한다. 추가적으로 근경, 중경, 원경을 적절히 배치하여 입체감과 깊이감을 고려함과 동시에 화면 내의 피사체의 크기를 고려하여 촬영 시 초점과 심도를 조정하게 된다. 특히 미디어 클로즈업(close-up) 샷은 얼굴에 확실한 주의를 끌리게 되며 배경은 중요하지 않아 포커스 아웃시키는 것이 일반적이다.

본 논문에서는 원본에서 이러한 샷들을 추출하여 원본 샷들의

배열과 동일한 순서로 배열한다면 원본 스토리의 내용을 충분히 반영한 요약본을 제작할 수 있을 것이라는 취지에서 이들 샷들을 추출하기 위해 필요한 다양한 방법들을 살펴본다.

II. Preliminaries

1. Scene Detection

영상 작품은 신(scene) 단위로 촬영되기 때문에 시나리오는 신 단위로 구성된다. 또한, 신은 최소 단위인 샷을 시간과 장소를 기준으로 하나로 모은 것이라고 볼 수 있는데 이는 장소 구분에 따라 나누는 것이 기본이다. 그리고 하나의 영상물을 프로그램이라고 본다면 프로그램은 여러 개의 시퀀스들로, 하나의 시퀀스는 여러 개의 신으로, 하나의 신은 여러 개의 샷들로 구성되며, 물리적으로 하나의 샷은 여러 개의 프레임(frame)으로 구성된다.

이러한 신을 검출하기 위한 다양한 도구와 알고리즘들[4-8]이 소개되었는데, 이들은 결과적으로 비디오 데이터를 자동으로 구조화하여 다양한 용도로의 활용을 목적으로 한다.

2. Shot Detection

[9]에서는 연속적인 샷들 사이에서 편집점이나 장면전환을 검출하는 다양한 방법론이 기술되었다. 즉, 편집점과 장면전환을 검출함으로써 샷들을 구분하는 것이다. 이를 위한 접근방법으로 화소 기반, 히스토그램 기반, 엣지 기반, 주파수 변환 기반, 움직임 기반, 통계적 특성 기반 등과 이들의 혼용이 기술되었다.

III. Medium Shot Detection with Blur

서론에서 언급한 바와 같이 장편의 드라마나 영화에서 스토리 기반의 축약된 요약본을 자동으로 제작하기 위한 방법으로 미디어 샷 크기의 샷들을 추출하는 것이다. 즉, 미디어 샷과 미디어 클로즈업 샷, 달리 표현하면 비스트 샷과 웨이스트 샷을 검출하는 것이다. 일반적으로 이들 샷들은 배경을 포커스 아웃시키는 것이 일반적이기에 프레임별로 흐림정도(blurri ness)의 연속성을 파악함으로써 해당 샷들을 검출하는 것이다.

[10-12]에서 이와 관련된 연구들이 소개되었는데 일반적으로 Fast Fourier Transform, Haar- Wavelet Transform, Laplacian Operator, Sobel Operator 등의 방법이 적용되었다. [12]의 경우는 Blur를 Motion Blur, Defocus Blur, Gaussian Blur, Box Blur 등 4가지로 분류하고 CNN(Convolutional Neural Network)을 적용하여 Blur 기반의 객체 분할에 활용하였으며 97% 정도의 정확도를 보였다.

실제적인 연구를 진행해야 하겠지만, 그리고 분명히 장르별로 다르겠지만, 일반적으로 사건과 인물 중심의 드라마나 영화에서 해당 샷들의 특수성을 고려하여 프레임별로 흐림정도를 추출하고 이를 기반으로 샷들을 재구성한다면 원본의 내용을 충실히 반영한 요약본을 자동으로 제작할 수 있을 것으로 보인다.

IV. Conclusions

본 논문에서는 장편의 드라마 혹은 영화에서 스토리 기반의 축약된 요약본을 자동으로 제작하기 위해 미디어 샷 크기의 샷들을 추출하기 위한 블러 검출 방법을 제안하였다.

영상문법에 근거하여 편집된 영상물에서 미디어 샷 정도 크기의 샷들을 추출하여 배열시킨다면 어느 정도 원본 내용을 충실히 전달할 수 있는 축약된 요약본을 제작할 수 있다는 가정 하에 비디오 검색을 위한 인덱싱에서 신 전환 혹은 샷 경계 검출 방법과 이미지에서 심도와 초점 기반의 화질 및 피사체 추출 방법을 고려하여 블러 검출기반의 미디어 샷 정도 크기의 샷들을 추출하기 위한 방법을 살펴보았다.

보통 미디어 샷 정도의 크기는 인물에 중심을 둔 샷들로 인물들의 관계에서 특히 대사나 표정으로 내용을 전달하기 위한 목적으로 적극 활용된다. 물론, 영상 제작자 혹은 촬영 및 편집 시에 이를 준수하지 않을 수 있고 설정샷 혹은 커버샷 등을 위한 다양한 샷들을 배제 시키고 이러한 크기만의 샷들을 배열함으로써 원본 내용을 충분히 전달할 수는 없겠지만 유튜브 등에서 몰아보기 등의 축약된 요약본을 자동으로 제작할 때 도움이 될 수 있을 것이라고 판단된다.

REFERENCES

- [1] Hervert Zettl, "Television Production Handbook," Thomson, pp. 401-408, 2004.
- [2] E. J. Choi, "Theory of Film Making", Communication Books, 2020.
- [3] Roy Thompson, "Grammar of the Shot," Focal Press, pp. 26-27, 66-97, 2021.
- [4] PySceneDetect (<https://scenedetect.com/en/latest/>)
- [5] Scene Detection Algorithms ([https://pyscenedetect.readthedocs.io/en/latest/reference/detection methods/](https://pyscenedetect.readthedocs.io/en/latest/reference/detection%20methods/))
- [6] Dongwook Shin, et al, "Video Scene Detection using Shot Clustering based on Visual Features," Journal of Korea Intelligent Information Systems Society, Vol. 18, No.2, pp.47-60, 2012.
- [7] Enqiang Guo, et al, "Learning to Measure Change: Fully Convolutional Siamese Metric Networks for Scene Change Detection,"(cite as: <https://arxiv.org/abs/1810.09111>), last revised 12 Nov. 2018.
- [8] Ken Sajurada, et al, "Weakly Supervised Silhouette -based Semantic Scene Change Detection," (cite as: <https://arxiv.org/abs/1811.11985>), last revised 5 Sep. 2022.
- [9] S. H. Adhulhussain, et al, "Methods and challenges in Shot Boundary Detection: A Review," Entropy, 2018, 20, 214, 23 Mar. 2018.
- [10] <https://pypi.org/project/blur-detector/>

- [11] P. A. Pagaduan, et al, “iBlurDetect: Image Blur Detection Techniques Assessment and Evaluation Study,” CESIT, pp. 286-291. 2020
- [12] <https://medium.com/nerd-for-tech/review-a-blur-classification-approach-using-deepconvolution-neural-network-blur-classification-93edf4552fd7>