

혈흔 감지 기반 동영상 내 잔인한 장면 필터링 시스템

손지원⁰, 박상근

경희대학교 소프트웨어융합학과

sjw0592@khu.ac.kr, sk.park@khu.ac.kr

Blood Detection-Based Gory Scenes Filtering for Videos Streaming Services

Jiwon Son⁰, Sangkeun Park

Department of Software Convergence, Kyung Hee University

요약

온라인 동영상 플랫폼에서 폭력적이거나 불쾌감을 유발할 수 있는 콘텐츠가 증가하면서, 시청자의 정서적 충격을 줄이기 위한 보호 장치의 필요성이 커지고 있다. 특히 혈흔이 포함된 잔인한 장면은 시청자에게 큰 불쾌감을 줄 수 있다. 기존의 자동화된 필터링 방법들은 전체 장면을 불러 처리하여 영상의 중요한 정보나 맥락을 손상시키는 문제가 있으나, 본 연구에서는 유형 장면만을 선택적으로 불러 처리해 콘텐츠의 흐름을 유지하면서도 불쾌감을 줄일 수 있는 시스템을 제안한다. 이를 위해 유튜브 시청 시 유형 장면을 자동으로 감지하고 부분적으로 불러 처리하는 크롬 확장 프로그램을 개발하였다.

1. 서론*

유튜브와 같은 온라인 동영상 플랫폼을 통해 방대한 양의 동영상 콘텐츠가 생산 및 소비되고 있다. 그러나 시청자에게 불쾌감이나 정서적 충격을 줄 수 있는 콘텐츠의 양도 꾸준히 증가하고 있다. 영상 시청자에게 불쾌감을 줄 수 있는 장면으로는 선정적 표현(음란한 묘사, 신체 노출 등), 폭력적인 행위 및 유해 물체(칼, 총 등), 잔인한 장면(상처, 피 등) 등이 있다. 이러한 장면들이 예고 없이 등장할 경우 시청자에게 강한 불쾌감과 정서적 충격을 야기할 수 있다. 특히, 피는 폭력의 심각성을 즉각적으로 인식하게 하는 강력한 시각적 요소로 작용하며, 시청자의 심리에 부정적인 영향을 미칠 수 있다. 예를 들어, 폭력적인 게임에서 피가 포함된 조건에서 게임을 할 때 플레이어가 게임을 더욱 잔혹하게 인식하고 적대감을 크게 느끼는 경향이 있다[1].

폭력적인 영상으로부터 시청자를 보호하기 위해 다양한 방법이 활용되고 있다. 영상 제작자는 별도의 경고 메시지를 추가하거나 불쾌감을 줄 수 있는 장면을 편집하여 시청자의 불쾌감을 최소화할 수 있다. 하지만 이러한 작업에는 편집자의 주관적인 판단이 개입될 뿐만 아니라 상당한

시간과 비용이 소요된다. 이에 따라 SW 기술을 활용한 자동화 연구가 활발히 진행되었으며, 특히 폭력적 장면을 자동으로 감지하고 처리하는 다양한 접근법이 제안되었다. 이러한 연구들은 분류, 객체 감지, 세그멘테이션 등의 기술을 통해 폭력적 장면 전체를 불러 처리하는 방식을 활용하고 있다[2, 5]. 그러나 기존의 방법들은 콘텐츠의 중요한 정보나 맥락을 손상시킨다는 한계를 가지고 있다. 따라서 시청자의 불쾌감을 최소화하면서도 영상의 자연스러운 흐름을 방해하지 않는 효과적인 영상 필터링 방법이 필요하다.

본 연구는 콘텐츠의 맥락을 유지하면서도 시청자가 불쾌감을 느낄 수 있는 특정 장면을 스스로 회피할 수 있는 시스템을 제안한다. 이 시스템은 영상에서 피가 포함된 장면만을 선택적으로 불러 처리하여, 시청자에게 불쾌감을 주는 요소를 제거하면서도 영상의 흐름을 유지한다. 이를 위해, 유튜브 시청 시 유형 장면을 감지하고 해당 부분만을 불러 처리한 영상을 시청할 수 있는 크롬 확장 프로그램을 개발했다.

2. 관련 연구

2.1. 부적절한 콘텐츠 필터링 기법

적절한 콘텐츠 감지를 위한 다양한 방법이 연구되었다. Verma et al.[2]는 특정 장면이 부적절한지 분류하는 방식으로 폭력적인 장면을 감지했다. 이 방식은 폭력성을

* 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00253723).

나타내는 객체를 감지하는 대신 장면 전체로 폭력성을 판단했다. 장면 내에서 폭력성을 나타내는 객체를 감지하기 위해, Tofa et al.[3]은 동영상 내에서 칼, 총, 피 같은 유해 객체를 감지했으며, Islam et al.[4]은 동영상에서 피 객체를 감지하여 이를 기반으로 폭력적인 장면을 분류하는 연구를 수행했다. 하지만 기존 연구는 폭력성과 관련된 객체를 감지하더라도 해당 객체만을 선택적으로 불러 처리하지 않고 전체 장면을 불러 처리하는 한계가 있다. 폭력성과 무관한 다른 객체까지 불러처리를 적용하면 해당 장면을 이해하기 위한 다양한 정보가 손실된다는 문제가 발생한다.

2.2. 동영상에서의 부적절한 콘텐츠 필터링

유튜브와 같은 동영상 스트리밍 서비스에서 부적절한 콘텐츠를 필터링하는 연구도 존재한다. Lee et al.[6]은 유튜브의 썸네일, 제목, 오디오 데이터 내 텍스트를 분석해 유해한 콘텐츠 여부를 판단했으며, Yousaf와 Nawaz[7]는 영상 내 장면을 폭력적 및 선정적 장면으로 분류하여 아동에게 유해한 콘텐츠를 식별했다. Anand et al.[8]은 유튜브 동영상의 텍스트, 이미지, 감성 분석을 통해 부적절한 콘텐츠를 판단하고, 채널 내 부적절한 동영상 비율을 제공하는 시스템을 개발했다. 이 연구는 동영상의 부적절성을 판단하는 데 중점을 두었으나, 특정 객체를 인식하고 이를 필터링하는 단계까지는 이르지 못했다는 한계가 있다.

3. 영상 내 잔인한 요소 필터링 시스템

본 연구에서는 유튜브 영상 시청 중 피가 포함된 장면을 감지하고, 피 부분만 불러 처리하여 수정된 영상을 사용자 웹 브라우저에 별도의 영상 재생 UI로 삽입하는 크롬 확장 프로그램을 개발했다. 시스템의 구조는 [그림 1]과 같다.

3.1. 피 감지 모델 개발

피 감지 모델 학습을 위해 Roboflow[9, 10, 11]에서 수집한 데이터로 커스텀 데이터셋(4,787개 이미지)을 구축했다. 데이터셋은 상처, 혈흔, 고어 장면, 피가 아닌 빨간 천 이미지 등을 포함하며, 각 이미지에 대응하는 객체 라벨 및 좌표 정보가 포함된 텍스트 파일이 있다. 클래스는 단일 클래스(0: Blood)로 설정했으며, 데이터는 학습(80%), 검증(10%), 테스트(10%) 세트로 분할했다. 모델 성능 평가는 Precision, Recall, mAP를 기준으로 하였다.

네 가지 YOLO 모델 (YOLOv5s, YOLOv5s-seg, YOLOv8s, YOLOv8s-seg)을 학습하여 비교했다. 학습 파라미터는 epoch 50, batch size 64, imgsz 416으로 설정했다 모델별 성능 결과는 [표 1]과 같다.

YOLOv5는 안정성과 커뮤니티 지원 측면에서, YOLOv8은 최신 기능과 성능 측면에서 강점을 보였다. 실시간 처리를 고려해 두 모델 모두 s(small) 크기 모델로 실험한 결과, YOLOv8s가 더 우수한 성능을 보였다. 최종적으로, 세그멘테이션이 가능한 YOLOv8s-seg 모델을 epoch 200, batch size 64, imgsz 640으로 추가 학습해 사용했다.

		YOLOv5s	YOLOv5s-seg	YOLOv8s	YOLOv8s-seg	YOLOv8s-seg epoch 200
Box	precision	0.879	0.818	0.799	0.845	0.928
	recall	0.758	0.71	0.645	0.65	0.836
	mAP50	0.847	0.813	0.751	0.776	0.909
	mAP50-95	0.638	0.586	0.585	0.618	0.826
Mask	precision		0.782		0.828	0.922
	recall		0.653		0.621	0.799
	mAP50		0.759		0.754	0.892
	mAP50-95		0.472		0.503	0.697

표1. 모델별 피 감지 성능 평가 결과

3.2. 피 객체 필터링 시스템 개발

객체 감지 후, 감지된 피 부분에 불러 처리를 적용해 영상을 수정했다. 이를 위해 yt-dlp¹를 사용해 유튜브 스트리밍 URL을 추출하고, ffmpeg²를 통해 오디오 분리와 프레임 추출을 수행했다. 추출된 각 프레임에 학습된 모델을 적용해 피 객체를 감지하고, supervision³을 사용해 감지된 박스 영역에 불러 처리를 수행했다.

필터링된 영상을 사용자에게 제공하기 위해 Vanilla JS와 Flask⁴ 서버를 사용해 크롬 확장 프로그램을 개발했다⁵. 프로그램 사용 화면은 [그림 3]에 제시되어 있다. 확장

¹ <https://github.com/yt-dlp/yt-dlp>

² <https://www.ffmpeg.org/>

³ <https://supervision.roboflow.com/latest/>

⁴ <https://flask.palletsprojects.com/en/3.0.x/>

⁵ <https://youtu.be/cl1C7NQ4XJw>

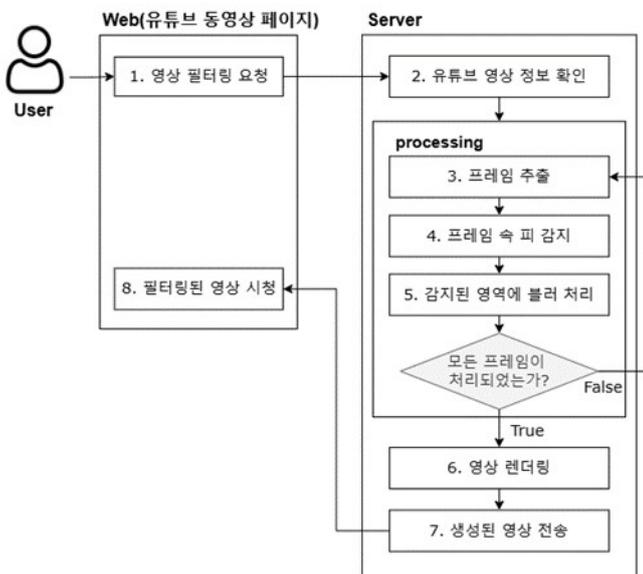


그림 1. 영상 내 잔인한 요소 필터링 시스템 구조도

프로그램을 설치하고 활성화하면, 유튜브의 [구독] 버튼 오른쪽에 아이콘이 생성된다. 이 버튼을 클릭하면 현재 시청 중인 영상이 서버로 전송되고, 서버에서는 피가 포함된 장면을 인식하고 블러 처리를 적용한 새로운 영상을 생성한다. 새롭게 생성된 블러 처리된 영상은 기존의 원본 영상의 하단에 삽입되어, 시청자가 피로 인한 불쾌감을 줄인 영상을 시청할 수 있도록 제공한다.



그림3. 피 객체 필터링을 위한 크롬 확장 프로그램

4. 결론

본 연구에서는 동영상 내 폭력적 장면, 특히 피를 감지하고 해당 객체에 블러 처리를 적용하는 시스템을 제안했다. 기존의 필터링 방식은 장면 전체를 블러 처리해 콘텐츠의 맥락을 훼손하고 불필요한 정보까지 차단하는 한계를 지니고 있었으나, 본 연구는 특정 객체만을 감지하여 필터링하는 방식으로 시청자의 불쾌감을 최소화하면서도 영상의 흐름과 중요한 정보를 유지하는 데 중점을 두었다. 실험 결과, YOLOv8s-seg 모델과 박스 영역 블러 처리

방식이 가장 효과적이었으며, 이를 크롬 확장 프로그램으로 구현하여 사용성을 확인했다.

향후 연구에서는 감지할 수 있는 유해 요소의 범위를 확대하고, 다양한 필터링 옵션을 제공함으로써 시청자가 더욱 자유롭게 콘텐츠를 제어할 수 있도록 시스템을 발전시킬 계획이다. 특히, 세그멘테이션 영역에 블러 처리를 적용하여 기존의 박스 영역 필터링보다 정밀한 필터링이 가능하도록 개선할 예정이다. 또한, 실시간 성능을 개선해 즉각적이고 효율적인 필터링 시스템을 구축해 사용자 경험을 더욱 개선하고자 한다.

참고문헌

- [1] Farrar et al. "Contextual Features of Violent Video Games, Mental Models, and Aggression." *Journal of Communication*, 56(2), 387-405, 2006
- [2] Verma et al. "Automatic video censoring system using deep learning." *International Journal of Electrical and Computer Engineering (IJECE)*, 12(6), 6744-6755, 2022
- [3] Tofa et al. "Inappropriate scene detection in a video stream." *Brac University, Thesis & Report, BSc (Computer Science and Engineering)*, 33-35, 2017
- [4] Islam et al. "Blurring of inappropriate scenes in a video using image processing." *Brac University, Thesis & Report, BSc (Computer Science and Engineering)*, 42, 2018
- [5] William Larocque. "Gore Classification and Censoring in Images." *University of Ottawa*, 2021
- [6] Lee et al. "Implementation and Evaluation of Harmful-Media Filtering Techniques using Multimodal-Information Extraction." *Journal of information and communication convergence engineering (JICCE)*, 21(1), 75-81, 2023
- [7] Yousaf and Nawaz. "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos." *IEEE Access*, vol. 10, pp. 16283-16298, 2022
- [8] Anand et al. "Customized video filtering on YouTube." *arXiv:1911.04013*, 2019
- [9] Joost Bos. "Blood Segmentation Dataset." *Roboflow Universe*, 2023
- [10] Blood. "blood_segmentation Dataset." *Roboflow Universe*, 2023
- [11] backupgeneral. "blood_segmentation_2 Dataset." *Roboflow Universe*, 2023