단말 적응적 미디어 화면비 변환 시스템

*이승호 **정진우 ***김성제
한국전자기술연구원 지능형영상처리연구센터
*seunghl@keti.re.kr **jw.jeong@keti.re.kr ***sungjei.kim@keti.re.kr

Device Adaptive Video Resolution Transform System

*Lee, Seungho **Jeong, Jinwoo ***Kim, Sungjei
Korea Electronics Technology Institute
Intelligent Image Processing Research Center

요약

언제 어디서든 한 손으로 미디어 콘텐츠를 소비할 수 있게 해주는 모바일 기기들이 기존 전통적 미디어 콘텐츠 단말기였던 TV나 데스크톱 PC들을 대체하게 되면서 세로형 영상 콘텐츠에 대한 수요가 나날이 높아져 가고 있다. 이와 더불어 모바일 단말기 제조사들은 서로 간의 경쟁에서 앞서기 위해 제품 차별화 전략을 수립하고 모바일 사용자들의 요구 사항을 세세하게 맞추기 위한 결과, 저마다 다른 디스플레이 해상도 규격을 가진 모바일 기기들이 생산되고 있는 상황이다. 이에 미디어 콘텐츠 제작자들은 기존 가로형 영상 콘텐츠와 더불어 새롭게 요구되는 세로형 영상 콘텐츠들을 저마다 다른 해상도 규격에 맞추는데 많은 시간과 비용을 투자하고 있다. 더 나아가 모바일 단말기 해상도 규격과 맞지 않는 영상 콘텐츠를 시청하게 될 경우, 모바일 사용자 입장에서는 디스플레이 전체 영역을 뷰포트로 잡을 수 없어 시청 만족도가 떨어질 수 있다.

이에 본 논문은 한 번의 콘텐츠 제작을 통해서도 추가 비용 없이 다양한 디스플레이 규격을 가진 단말기들에 대해 맞춤형 콘텐츠 서비스 제공을 가능하게 하여 미디어 콘텐츠 소비자들에게 충분한 시청 몰입감을 제공해줄 수 있는 단말 적응적 미디어 화면비 변환 시스템을 제안한다. 단말 적응적 미디어 화면비 변환 시스템은 딥러닝 네트워크 모델과 이미지 관련 라이브러리를 기반으로 하여 설계한 시스템이며, 사용자가 시청하기 원하는 영역을 판단하고, 사용자가 원하는 뷰포트 종횡비에 따라 해당 영역을 잘라내어 사용자가 원하는 세로형 영상 콘텐츠를 제공해준다.

1. 서론

20세기 중후반부터 등장한 TV, 데스크톱 PC 등은 그 시대를 대표하는 미디어 콘텐츠 단말기의 핵심으로서 전통적인 형태의 콘텐츠를 제공해주던 신문과 라디오를 대체하며, 멀티미디어 서비스 산업에서 매우 큰 비중을 담당하게 되었다. TV와 데스크톱 PC의 특징으로는 일반적으로 가로 길이가 세로 길이보다 더 긴 디스플레이 종횡비를 사용한다는 점이 있다. 이러한 단말기 환경 덕분에 자연스럽게 미디어 콘텐츠 또한 가로 길이가 세로 길이보다 더 긴 가로형 영상 콘텐츠가 주류를 이루게 되었다.

그러나 2010년을 전후로 하여 한 손을 이용해 언제 어디서든 사용할 수 있는 스마트폰과 같은 모바일 단말기들이 대중화되면서, 미디어 콘텐츠 단말기 시장을 주도하던 TV와 데스크톱 PC는 점차 모바일 단말기에게 그 주도권을 넘겨주게 되었다. [1]

디스플레이 환경 측면에서 기존 미디어 콘텐츠 단말기와 모바일 단말기의 가장 큰 차이점은 바로 디스플레이 회전 가능 여부이다. 모바일 단말기는 기기 자체를 회전시켜 디스플레이 회전을 매우 쉽게 할 수 있기 때문에, 기존 가로형 영상 콘텐츠와 더불어 세로 길이가 가로 길이보다 더 긴 세로형 영상 콘텐츠도 불편함 없이 시청할 수 있다.

이러한 모바일 특성 덕분에 많은 모바일 사용자들이 점차 세로형 영상 콘텐츠에 점차 익숙해지고 있는 상황이다. Mobile Overview Report에 의하면 모바일 사용자들이 단말기를 세로 방향으로 사용하는 시간이 전체 사용 시간의 94%를 차지하는 것으로 보고되었다. [2] 또한, Mezzo Media의 디지털 동영상 이용 행태 조사 분석 자료에 의하면 작은 영역을 차지하는 가로형 광고보다 전체 화면의 세로형 광고가 더 기억에 남는다고 느끼는 것으로 나타났다. [3]

오늘날, 많은 콘텐츠 서비스 기업들이 이러한 모바일 사용자들의 사용 패턴과 요구 사항을 분석하고 이에 대응할 수 있는 맞춤형 서비스들을 선보이고 있다. YouTube 및 Twitch와 같은 미디어 콘텐츠 플랫폼과 Instagram과 같은 SNS 플랫폼에서 세로형 영상 콘텐츠 전용 서비스를 제공하기 시작했고, 기존 가로형 영상 콘텐츠를 제작 및 공급해왔던 지상 3사를 비롯한 방송국에서도 모바일 환경에 최적화된 세로형 영상 콘텐츠를 추가적으로 제공하기 시작했다.

이 밖에도 스마트폰 제조사 또한 차별화 전략과 더불어 저마다 다른 사용자들의 요구 사항을 만족시키고자 다양한 디스플레이 종횡비를 갖는 스마트폰 단말기들을 출시하고 있으며, 더 나아가 접었다 펼칠 수 있는 플렉서블 디스플레이를 채택한 폴더블 스마트폰이 등장하기에 이르렀다. 2021년 1월 기준, 출시된 스마트폰들의 해상도 규격은 모두 14가지이며, 스마트폰 종류만 무려 151종에 달한다. [4]

이처럼 미디어 콘텐츠 시장 환경이 급격하게 변하면서 콘텐츠의 품

질 관리 이슈가 점차 부각되고 있다. 기존 가로형 영상 콘텐츠 외에 추가로 세로형 영상 콘텐츠까지 공급해야하는 미디어 콘텐츠 기업 입장에서는, 각기 다른 모바일 단말기들의 디스플레이 규격까지 고려하면서 콘텐츠를 생산하기에는 많은 시간과 비용을 소모하게 되어 매우 큰 부담으로 작용하고 있다. 아울러 모바일 단말기 해상도 규격과 맞지 않는 영상을 시청하는 사용자 입장에서는 디스플레이 전체 영역을 뷰포트로 설정할 수 없기 때문에 시청 만족도가 떨어질 수 있다.

본 논문은 가로형 영상 콘텐츠로부터 사용자가 주로 시청하고 싶어하는 영역을 예상 및 추적하고, 사용자가 원하는 디스플레이 종횡비에 맞게 뷰포트 맞춤형 영상을 출력하는 단말 적응적 미디어 화면비 변환 시스템을 소개한다. 이 시스템은 딥러닝 네트워크 모델과 이미지 관련 라이브러리 등을 기반으로 하여 설계한 시스템이며, 각기 다른 종횡비의 단말을 가진 미디어 소비자들에게 충분한 시청 몰입감을 제공할 수 있도록 사용자가 원하는 영역을 판단하고, 해당 영역을 사용자가 원하는 뷰포트 종횡비에 따라 잘라내어 사용자가 원하는 해상도의 영상 콘텐츠를 제공한다.

본 논문은 2장에서 단말 적응적 미디어 화면비 변환 시스템을 구성하고 있는 모듈들에 대해 소개하고, 3장에서 시스템 구현 환경과 화면비 변환 결과 및 분석에 대해 설명하며, 4장에서 결론을 제시하며 마무리를 맺는다.

2. 단말 적응적 미디어 화면비 변환 시스템

단말 적응적 미디어 화면비 변환 시스템은 <그림 1>과 같이 장면 분석기(Scene analyzer), 객체 검출기(Object detector), 객체 추적기(Object tracker), 뷰포트 추출기(Viewport extractor), 초해상도 생성기(Super-resolution generator), 이렇게 총 5개의 모듈로 구성되어 있다.

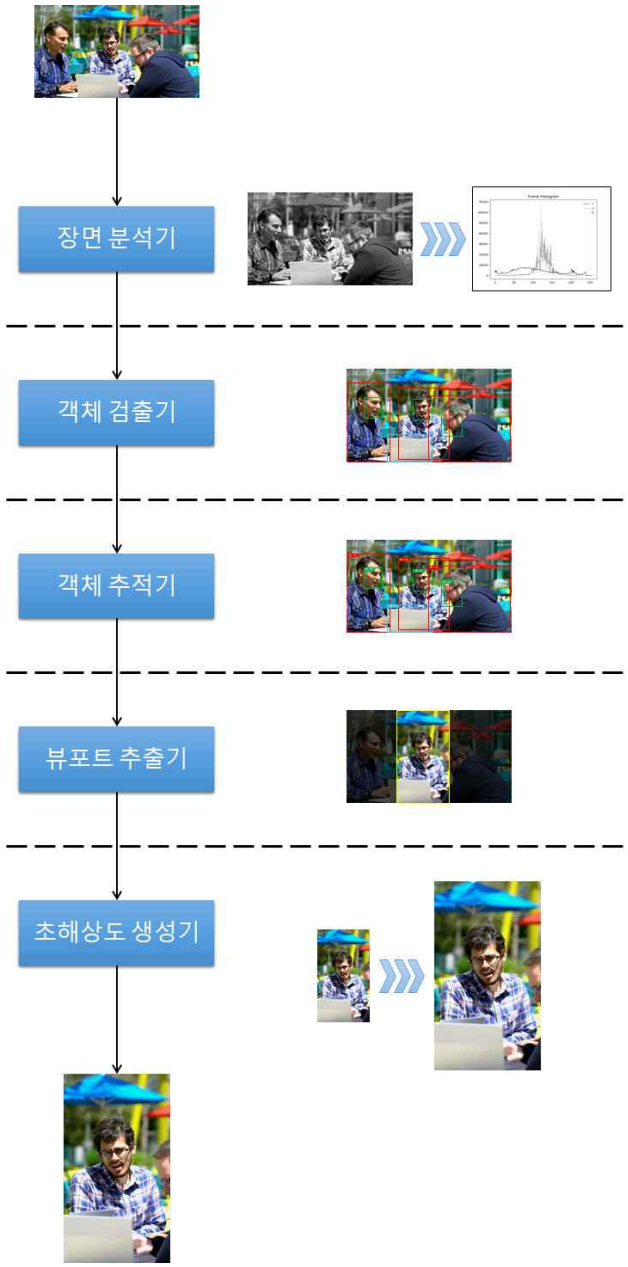
장면 분석기는 2가지 기능을 갖추고 있다. 하나는 영상 테두리 영역의 주요 색상을 검출하여 영상의 테두리 영역을 계산 및 추출하는 기능이며, 다른 하나는 영상 전체의 픽셀 값으로 히스토그램을 만들고 이를 일정 크기의 큐에 저장하면서, 현재 프레임에서 장면 전환이 이루어졌는지를 판별하는 기능이다.

객체 검출기는 널리 사용되고 있는 YOLOv5를 채택하였으며 [5], 80개 클래스를 대상으로 하여 객체를 검출하는 일반적인 YOLOv5(m) 모델과 특별히 사람 얼굴만을 검출하는 YOLOv5(s) 모델을 통합하여 최종적으로 81가지 종류의 객체를 검출할 수 있도록 구현하였다.

객체 추적기는 ByteTracker를 미디어 화면비 변환 시스템에 맞게 수정하여 채택하였다. [6] 기존 ByteTracker는 카메라 자체의 움직임이 거의 없고 장면 전환이 없는 환경에서 다수의 움직이는 사람들만 추적하는 것에 초점을 맞추어 구현되었기 때문에 이를 미디어 화면비 변환 시스템에 그대로 적용하기에는 적합하지 않다고 판단하였다. 따라서 기존 ByteTracker에서 구현되지 않았던 Multi class tracking 기능을 새로이 추가하였고, 장면 분석기에서 장면 변환이 감지될 경우에 Tracking 기능이 동작하지 않도록 Tracking 버퍼 비우기 기능을 추가하였다.

뷰포트 추출기는 장면 분석기, 객체 검출기, 그리고 객체 추적기로부터 얻은 프레임 분석 결과를 활용하여 사용자가 가장 보고 싶어할만한 영역을 추정한다. 구체적으로 장면 전환 여부, 객체 검출 여부, 이전 프레임의 뷰포트 위치, 이전 프레임으로부터 추적된 객체의 유무 등 다양한 정보를 바탕으로 적절한 뷰포트의 위치를 선정한다.

초해상도 생성기는 입력 이미지와 뷰포트 추출기로부터 획득한 뷰포트 ROI (Region of Interest)를 활용하여 단말기 디스플레이에 띄울 이미지를 잘라내고, 해당 이미지의 해상도 및 화질을 향상시키는 역할을 담당한다.



<그림 1. 단말 적응적 미디어 화면비 변환 시스템 구조도>

3. 구현 환경 및 동작 결과

단말 적응적 미디어 화면비 변환 시스템의 구현 및 테스트는 데스크톱 환경에서 진행되었으며, 자세한 하드웨어 및 OS 환경은 다음 <표 1>과 같다.

항목	환경
CPU	Intel(R) Core(TM) i9-9980XE CPU @ 3.00GHz
RAM	64 GB
GPU	NVIDIA GeForce RTX 3090
OS	Windows 10 21H2 Build 19044.1682

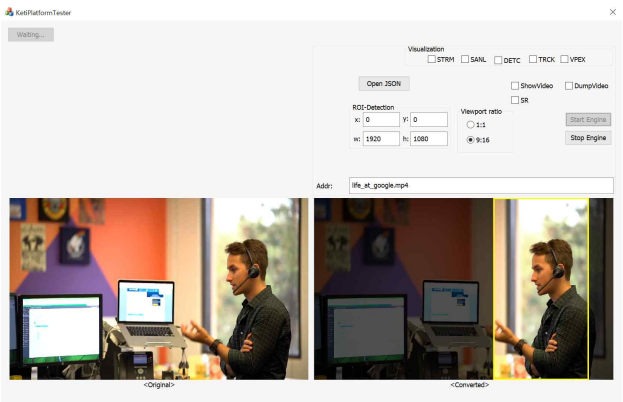
<표 1. 단말 적응적 미디어 화면비 변환 시스템 하드웨어 환경>

또한, 미디어 화면비 변환 시스템을 구현하기 위해 여러 라이브러리를 사용하였다. 이미지 입력 인터페이스와 장면 분석을 구현하기 위해 OpenCV 라이브러리를 사용하였고, 객체 추적 등 딥러닝 모델을 활용한 추론의 고속화를 구현하기 위해 NVIDIA에서 제공하는 CUDA, CUDNN, TensorRT 라이브러리를 사용하였다. 그 외 컴파일 도구 환경 및 자세한 라이브러리 버전은 다음 <표 2>와 같다.

항목	버전
CUDA	11.6
CUDNN	8.3.3.40
TensorRT	8.4.0.6
OpenCV	4.5.3
Tool	Visual Studio 2019 Version 16.11.13

<표 2. 단말 적응적 미디어 화면비 변환 시스템에 사용된 라이브러리 및 도구>

다음 <그림 2>는 미디어 화면비 변환 시스템을 구현한 소프트웨어의 동작 화면이다. 입력 영상은 Google에서 배포하고 있는 영상이며 [7], 9:16 종횡비로 뷰포트를 실시간으로 추출하는 모습을 보여주고 있다. 노란색 박스 UI는 미디어 화면비 변환 시스템이 사용자가 설정한 종횡비에 맞추어 추출한 뷰포트 영역이다.



<그림 2. 단말 적응적 미디어 화면비 변환 시스템 동작 화면>

다음 <표 3>은 미디어 화면비 변환 시스템의 각 모듈에 대한 동작 속도를 측정한 결과이다. 입력 영상의 해상도는 1920x1080 (FHD)로 설정하였고, 각각 1:1 (1080x1080), 9:16 (608x1080) 종횡비로 뷰포트를 추출하였으며, 이미지 초해상도의 배율은 2배로 설정하였다.

모듈	1:1	9:16
장면 분석기	148 fps (≈6.712 msec)	150 fps (≈6.627 msec)
객체 검출기	83 fps (≈11.906 msec)	100 fps (≈9.936 msec)
객체 추적기	20,491 fps (≈0.048 msec)	18,427 fps (≈0.054 msec)
뷰포트 추출기	1,469,546 fps (≈0.000 msec)	1,549,469 fps (≈0.000 msec)
초해상도 생성기	27 fps (≈36.264 msec)	46 fps (≈21.632 msec)

<표 3. 단말 적응적 미디어 화면비 변환 시스템의 각 모듈에 대한 동작 속도 측정 결과>

FHD 영상을 기준으로 전체 시스템 모듈에 대한 동작 속도 측정 결과, 초해상도 생성기를 제외한 나머지 모듈에 대해서는 60 fps를 넘는 동작 속도를 보여주었다. 초해상도 생성기는 1:1 뷰포트 추출의 경우, 일반적인 fps 기준인 30 fps에 미치지 못한 27 fps이었지만, 9:16 뷰포트 추출의 경우에는 46 fps의 동작 속도를 보여줬다.

초해상도 생성기에서 발생하는 프로세스 오버헤드의 비중이 상당히 높은 것으로 파악되었다. 따라서 이를 해결하기 위해서는 초해상도 딥러닝 모델 고속화에 대해 지속적으로 연구할 필요성이 있다고 판단된다.

4. 결론

본 논문에서는 딥러닝 네트워크 모델과 OpenCV를 비롯한 여러 라이브러리와 결합하여, 가로형 영상 콘텐츠로부터 사용자가 시청하기 원하는 영역을 추출하여 세로형 영상으로 보여주는 단말 적응적 미디어 화면비 변환 시스템에 대해 서술하였다. 현재 이 시스템은 최고급 사양의 데스크톱 환경에서 구현되었으며, 향후에는 모바일 환경에서 구현할 예정이다. 그러기 위해선 지속적으로 각 시스템 모듈에 대한 최적화 및 고도화 연구를 진행할 것이며, 시스템 구축에 필요한 하드웨어 사양을 점차 낮출 수 있을 것으로 기대한다. 또한, 모바일 사용자들에게 쾌적하고 만족도 높은 세로형 영상 콘텐츠를 제공할 수 있을 것으로 전망한다.

Acknowledgement

본 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (No.2021-0-00802, 속성을 유지하는 지능적 미디어 화면비 변환 기술 개발, 100%)

참고문헌

[1] 정용찬, 김윤화, “2017년 방송매체 이용행태조사 주요 결과”, KISDI STAT Report 18-03 (정보통신정책연구원, 2018)

[2] “Mobile Overview Report October-December 2014”, Scientia mobile.com (2014), Accessed from: https://data.wurfl.io/M

OVR/pdf/2014_q4/MOVR_2014_q4.pdf
[3] “2018 디지털 미디어 트렌드”, Mezzomedia.co.kr (2018), Accessed from: http://www.mezzomedia.co.kr/api/download?file_no=720&preview=1
[4] “Screen Sizes”, Accessed from: https://screensiz.es
[5] Ultralytics, “Yolov5”, Accessed from: https://github.com/ultralytics/yolov5
[6] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang, “ByteTrack: Multi-Object Tracking by Associating Every Detection Box”, arXiv preprint arXiv:2110.06864, 2021.
[7] Google, “life at google”, Accessed from: https://drive.google.com/corp/drive/u/0/folders/1KK9LV--Ey0UEVpxssVLhVl7dypgJSQgk