얼굴인식을 활용한 영상 내 특정인물 기반 대표 이미지 추출 시스템

이현지 이계민
서울과학기술대학교
{wahpopaltee, gyemin}@seoultech.ac.kr

Video Thumbnail Generation Using Character Face Recognition

Lee, Hyunji Lee, Gyemin
Seoul National University of Science And Technology

요약

최근 인터넷 플렛폼이 대중화되면서 영상물을 접하는 횟수가 늘어났다. 영상 선택에 있어서 대표 이미지가 중요한 역할을 하는데, 현재 빅데이터를 이용하여 개인 맞춤 서비스가 활성화 되면서 이를 이용하여 개인 맞춤 서비스로 특정인물 기반 대표 이미지 추출할 수 있게 된다면 영상 선택에 있어 소비자의 편의를 도우며 이목을 끌 수 있을 것으로 예상된다. 이에 본 논문은 영상 산업 기술과 방송 통신 융합 서비스의 일환으로 특정인물 기반 대표이미지를 추출하는 서비스에 대해 연구하였다. 이를 위하여 얼굴 인식을 처리하는 컴퓨터 비전 기술을 이용하여 얼굴 인식 분야를 연구 개발하였다.

1. 서론

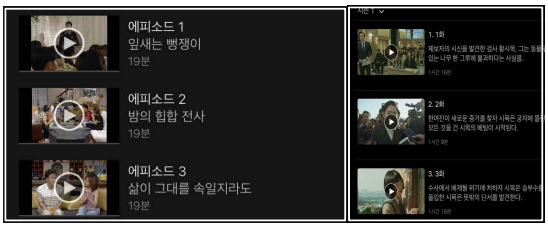
현재 OTT 서비스 (넷플릭스,, 왓챠 등)는 드라마, 예능 등에서 각 회차마다 대표 이미지와 줄거리를 제공한다. [그림1.1] 참고. 하지만 내용이랑 무관하게 무작위로 대표이미지가 추출되는 경우가 많다. 빅데이터를 이용하여 사용자의 시청 기록을 이용해 선호하는 배우와 장르 등 취향 파악이 가능해졌지만, 대표 이미지에 있어서 개인 취향이 전혀 반영되지 않은 것을 알 수 있다. 하지만 OTT시장은 더욱 넓어지고 계속해서 새로운 플랫폼이 나오고 있기 때문에 개인의 취향 반영에 있어 강점을 두어 다른 기업보다 경쟁력을 갖추는 것이 굉장히 중요해졌다.

얼굴인식은 이미 컴퓨터 비전 등 오픈 소스를 통하여 상당한 개발이 되어있다. 얼굴 인식을 이용한 항목에서는 주로 휴대폰 잠금 서비스, 출석 확인 등 보안과 관련된 항목이 많았고 대표 이미지 추출 등 미디어와 관련된 항목은 거의 찾아볼 수 없다. 특히 대표 이미지 추출은 앞서 말한 부분과 같이 무작위로 추출되거나, 사람이 직접 수기로 찾아내는 경우가 많다. 이에 따라 대표 이미지를 자동적으로 추출할 수 있는 시스템이 갖춰진다면, 제작과 공급이 정해져있는 OTT서비스뿐만 아니라 개개인의 참여로 이루어지는 인터넷 플렛폼 (Youtube, TikTok) 등에서도 활용할 수 있어 서비스 분야가 점차 확대 될 것으로 전망된다.

영상 내에서 특정인물 기반 얼굴 인식을 하게 된다면, 결국 영상 내 특정 인물이 출연한 부분을 즉 특정 인물이 나오는 시간까지 파악을 할 수 있게 되는데, 이를 이용하면 영상 내 검색 및 영상 요약 본 추출이 가능해진다. 현재 Youtube와 같은 접근성이 좋은 인터넷 플렛폼이 활성화 됐기 때문에 방송국에서는 시청자들의 이목을 끌기 위해 방송의 일부를 요약하여 요약본으로 제공해주는 횟수가 점 차 늘어나고 있어 이러한 부분에도 적극 활용할 수 있을 것으로 예상한다.

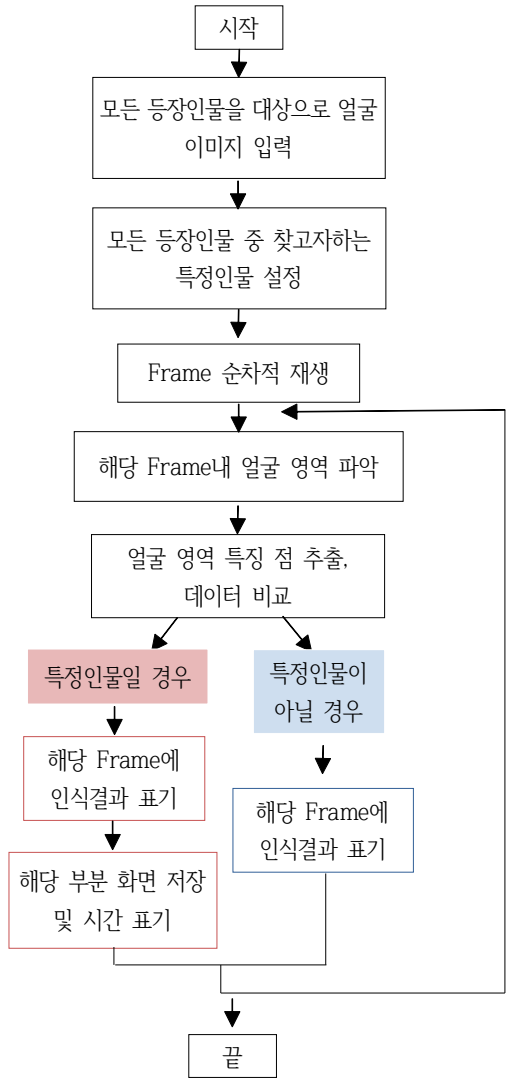
이로 인해 제작 측 에서는 기존 수기로 직접 출연 부분을 찾아야 했던 인력 낭비를 줄여 전보다 쉽게 영상 요약본 추출이 가능해질 것으로 예상된다. 배급 측면에 있어서도 개인의 취향에 맞춘 대표 이미지 제공이 가능해지면서 시청자의 편의를 도울 수 있는 장점이 있다. 또한 영상을 소비하는 소비자도 특정 인물이 나오는 시간대를 파악할 수 있기 때문에 시간 절약의 효과도 누릴 수 있다.

분야는 드라마, 예능에만 국한된 것이 아니다. 2020 도쿄 올림픽을 통해 대중들의 이목을 끄는 선수들이 대거 등장하며 선수의 인지도도 점차 높아지고 있기 때문에, 스포츠 경기 분야에서도 적극 활용할 수 있을 것이다. 이 뿐만이 아니라 한류 중심에 있는 엔터테이너 즉 가수들의 무대 영상에서도 활용할 수 있으며 응용 분야는 점차 많아질 것으로 예상된다.



[그림1.1] (좌)OTT서비스 중 왓챠의 대표 이미지 예시(드라마 논스톱)
(우)OTT서비스 중 넷플릭스의 대표 이미지 예시(드라마 비밀의 숲)

2. 작품 설계 및 구현



[그림2] 이 시스템의 Flowchart

본 논문은 컴퓨터 비전(OpenCV)[1,2]와 Face Recognion[3] 라이브러리를 이용하여 얼굴의 특징점을 찾아낸다.[그림2].

처음 입력으로 사용하고자 할 영상과 함께 모든 등장인물의 이미지를 넣는다. 모든 등장인물의 이미지로 얼굴의 특징 점을 추출한다. 이후에는 모든 등장인물 중 특히 분석을 원하고자 하는 특정인물을 설정한다. 그다음에는 영상의 각 frame을 순차적으로 재생한다. frame 재생이 될 때 각 frame에서 나오는 얼굴 영역을 파악한다. 그 다음 얼굴영역 특징 점을 추출하고, 그 특징 점들을 앞서 입력으로 넣은 모든 등장인물의 데이터와 비교한다. 이때 특정인물과 일치한다면 frame에 인식 결과를 표현하고, 이후 특정인물이 나온 시간을 표기하고, 특정인물이 나온 부분의 Frame을 이미지 파일(jpg)로 저장할 수 있도록 하였다.

이 시스템의 세부적인 사항은 다음과 같다. 모든 등장인물에 대하여 이미지를 넣는 이유는 정확도를 높이기 위함에 있다. 특정인물만 넣는다면 정확도의 차이가 있어 반례로서 다른 등장인물의 사진을 넣어 해결하였다. 특정인물만 넣었을 때 특정인물이 아님에도 특정인물이라고 분류한 경우(오탐)는 육안 상으로도 확인할 수 있다. [그림3]. 수치적으로도 정확도에서 차이가 나는 것을 확인할 수 있는데, 특정인물의 이미지만 넣었을 경우 오탐의 경우는 218회이다. 이때 모든 등장인물의 이미지를 다 넣었을 때의 오탐 횟수는 1회로 기록하였다. 특정인물이 나왔는데 특정인물이라고 도출하지 못한 경우(미탐)에 있어 특정인물만 넣은 경우는 7회, 모든 등장인물을 넣은 경우는 20회를 기록해 미탐할 확률이 줄어 특정인물일 때 특정인물이 맞다고 판단한 경우(정탐)횟수가 늘어났지만, 오탐에서의 결과가 현저히 좋지 않아 모든 등장인물의 이미지를 넣는 방법으로 채택하였다. [표1].

모든 등장인물을 넣었다고 하더라도 미처 입력에 넣지 않은 인물이 등장하였을 때에는 frame에서 결과 값을 'Unknown'으로 표기하여 결과물을 정확히 확인할 수 있게 하였다. [그림4]

또한 영상이 길어짐에 따라 처리할frame이 많아질 수 있기에 FPS조절이 가능하게 하여 처리 시간을 단축시켜 실용성을 더하였다.

	모든 등장인물을 입력으로 넣은 경우	특정인물만 입력으로 넣은 경우
정탐	88	101
미탐	20	7
오탐	1	218
해당 인물이 나온 총 횟수	108	108

[표1] 입력에 따른 정확도 분석 차이 표(단위 초)(사용 영상: 무한도전 -326회 요약본, 11분49초)

[그림 3] 특정인물을 착각한 모습

[그림4] 'Unknown'의 사례

3. 실험 및 결과

	영상 종류	길이
1	한국 드라마(질투의 화신 23화 요약본)	23분06초
2	한국 예능(라디오스타 739회 요약본)	15분07초
3	한국 예능(무한도전 326회 요약본)	11분 49초
4	스포츠 경기(2020도쿄 올림픽-남자 양궁 단체전 결승)	07분20초
5	외국 영화(퀸카로 살아남는 방법 [2004]요약본)	21분01초

[표2] 시연 영상의 정보를 나타낸 표

영상	정탐	오탐	미탐				실제 특정 인물이 나온 횟수
			놓침	작음	옆모습	기타	
1	210	3	11	21	21	4	267
2	275	10	0	12	15	4	306
3	88	1	8	4	7	1	108
4	41	1	1	5	3	0	50
5	345	7	2	30	16	2	395

[표3] 각 영상별로 정확도 분석 한 결과 (단위 초)

영상	경우1			경우2		
	정밀도	재현율	F1 Score	정밀도	재현율	F1 Score
1	0.99	0.95	0.97	0.99	0.79	0.88
2	0.96	1.00	0.98	0.96	0.90	0.93
3	0.99	0.92	0.97	0.99	0.81	0.88
4	0.98	0.98	0.98	0.98	0.82	0.91
5	0.98	0.99	0.99	0.98	0.87	0.92

[표4] 각 영상 별로 F1 Score을 구한 값

본 시스템을 개발하기 위해 사용된 시연 영상은 한국 드라마 (질투의 화신-23화 요약본, 23분06초), 한국 예능(라디오 스타-739회 요약본, 15분07초), 한국 예능(무한도전-326회 요약본, 11분49초), 올림픽 경기 (2020 도쿄 올림픽 양궁 - 남자 단체전 결승, 07분20초), 외국 영화(퀸 카로 살아남는 방법 2004-요약본, 21분01초)를 활용하여 장르가 다른 영상물에서의 정확도를 분석하였다. [표2]

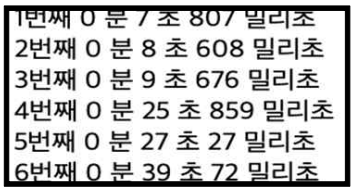
정확도 분석을 위해 정탐, 오탐, 미탐으로 분류할 수 있고 미탐 종류에는 놓침(특정인물이여도 놓치거나 다른 인물로 착각), 작음(얼굴이 작아 파악 불가), 옆모습(옆모습이라 파악 불가), 기타(조명·각도·표정 등의 이유로 파악 불가능 한 경우)로 나누어 보았다. 한국 드라마 영상에선 순서대로 210, 3, 11, 21, 21, 4를 기록했고 한국 예능(라디오 스타)에선 275, 10, 0, 12, 15, 4, 한국 예능(무한도전)에선 88, 1, 8, 4, 7, 1를 기록 하였다. 이어서 올림픽 영상에선 41, 1, 1, 5, 3, 0, 마지막 외국 영화에서는 345, 7, 2, 30, 16, 2를 기록 하였다.[표3]. 미탐 중 작음, 옆모습, 기타에 해당한 사례는 [그림5]로 확인할 수 있다.

이를 이용 하여 정밀도 (결과 값이 참일 때 입력도 참일 확률), 재현율 (입력 값이 참일 때 결과 값도 참일 확률)을 이용해 F1 Score(정밀도와 재현율을 동시에 고려한 지표)라는 결과 값을 도출해내었다. 오류 중에서도 Miss를 두 가지로 분류할 수 있다. 첫 번째는 시스템 자체의 결함이다. 특별한 이유가 없음에도 불구하고 파악을 하지 못한 '놓침'의 경우가 있고, 두 번째는 시스템 자체의 결함이 아닌 작음, 옆모습, 앵글 등과 같이 육안으로도 파악이 불가능한 경우이다. 따라서 F1 Score를 구할 때 Miss에서 '놓침'만 해당하는 경우 1, Miss에 '놓침', '작음', '옆모습', '앵글' 이 모든 경우를 합친 경우 2 , 이렇게 두 가지를 따로 나누어 구해보았다. 한국 드라마에서는 경우 1에 대한 F1 Score 의 값이 0.97, 경우 2에 대해선 0.88을 기록했다. 한국 예능(라디오 스타)에서도 순차적으로 0.98, 0.93을 기록하였고, 한국 예능(무한도전) 에는 0.97, 0.88을 기록, 올림픽 영상에선 0.98, 0.91을 기록, 외국 영상에서는 0.99, 0.92를 기록하였다. 비교적 움직임이 적은 토크쇼 위주의 영상(한국예능 - 라디오스타), 정적인 스포츠 양궁 (올림픽 영상), 원샷 위주의 영상 (외국영화)의 경우보다 유동성이 많은 한국 드라마, 한국 예능(무한도전)의 F1 Score 값이 낮은 것을 확인할 수 있다. 이는 움직임이 많은 영상일수록 경우2에 대한 F1 Score이 낮아지는 것을 의미한다. 하지만 경우1에서 F1 Score 값은 전부 0.97~0.99로 얼굴이 단독으로 크게 나오는 경우엔 정확도가 높아 실용성을 입증하였다.

정확도 분석까지 마친 시연 영상의 결과물은 다음과 같다. 영상 내 특정 인물이 나올 때 그 인물인지 확인할 수 있도록 얼굴인식 박스를 추가하였다. [그림6] 해당 그림에는 얼굴인식 박스가 있으나 이는 정확도 분석과 시연 결과 확인을 위한 것이고 실제 사용에는 이를 제거하여 이미지를 추출할 수 있다. 이에 이어 해당하는 장면을 자동으로 이미지 파일(jpg, png)로 저장할 수 있도록 하였다. 이어서 [그림7]과 같이 특정 인물이 나올 때 영상 내의 시간을 표기하였다. 실시간으로 특정 인물이 나올 때 마다 추가되며, 순서대로 특정 인물이 나온 횟수, 분, 초, 밀리초에 대한 정보를 나타낸다.

[그림5] 인식이 안되는 부분 (좌측부터 작음, 옆모습, 표정)
(한국 예능 2)

[그림6] 특정인물 얼굴인식 장면을 저장한 이미지 파일
(한국드라마)



[그림7] 특정인물이 나오는 장면의 시간 출력한 결과 값

4. 결론

특정인물 기반 대표 이미지 추출 시스템은 대표 이미지 추출을 위해 영상 내의 시간을 활용할 수 있기 때문에 활용도가 매우 높을 것으로 예상된다.

시청자의 시간 단축, 배급사의 경쟁력, 제작사의 인력 낭비 감소 등 영상이 유통되는 과정에서 관련된 이들이 크게 이점을 느낄 수 있을 것이다. 이어 현재 인터넷 플렛폼을 이용하여 영상을 소비함과 동시에 제작하는 개개인이 늘어나고 있어 대중들의 수요가 높아질 것으로 예상된다. 또한 대표 이미지 추출뿐만 아니라 영상 내 검색 기능, 요약본 추출 등 점차 활용할 수 있는 범위가 넓어지며 영상물의 영역도 드라마, 예능에만 국한 된 것이 아닌 스포츠 경기 등 다양한 영상물과 관련한 분야에 적용할 수 있다.

또한 FPS 조절을 통해 영상 내 분석이 짧은 시간 내에 가능하도록 실용 적인 측면도 고려하였다. 이에 그치지 않고 다양한 분야의 영상을 활용하여 정확도 분석을 마쳤기 때문에 실현 가능성과 신뢰성도 고루 갖추어 앞으로 활용 방안이 기대된다.

<참고문헌>

[1]사우랍 카푸『파이썬3로 컴퓨터 비전 다루기』,에이콘 출판, 2018

[2]이세우『파이썬으로 만드는 OpenCV 프로젝트』, 프로그램인사이트, 2019

[3]Adrian Rosebrock,『Face recognition with OpenCV, Python, and deep learning』 (2018.06.18.)(http://www.pyimagesearch.com/2018/06/18/face-recognition-with-opencv-pyython-and-deep-learning/)