

딥러닝 기반 배우 재인식 기술

임동혁, 서용석, 박지현
한국전자통신연구원(ETRI)

iammoni@etri.re.kr

Actor Re-Identification Based on Deep Learning

Dong-Hyuck Im, Yongseok Seo, Jihyun Park
Electronics and Telecommunications Research Institute

요약

본 논문은 영화/드라마에서 배우를 재인식하기 위한 시스템을 제안한다. 세 개의 이미지를 입력하여 이미지 간의 차이점과 공통점을 동시에 학습할 수 있도록 두 쌍의 Cross-Input Neighborhood Difference 특징 맵을 구성하는 방식으로 네트워크를 확장하였다. 영화, 드라마를 이용하여 생성한 자체 데이터셋을 이용하여 실험한 결과 제안한 모델을 비디오 콘텐츠의 배우 재 식별에 적용할 수 있음을 확인하였다.

I. 서론

사람 재인식이란 복수의 카메라로 이루어진 환경에서 카메라 1에서 사람 A가 인식되었을 경우 이 사람이 다른 시간 혹은 다른 카메라에 등장했을 경우 이 두 인물을 같은 인물이라고 재인식하는 기술을 말한다. 다수의 감시 카메라의 내용을 자동 분석하기 위해 사람 재인식 기술에 대한 연구가 진행되고 있다. 이러한 사람 재인식 기술은 비단 감시 카메라뿐만 아니라 영화, 드라마와 같은 비디오 콘텐츠에도 적용될 수 있다. 본 논문에서는 영화, 드라마의 출연 배우에 대한 사람 재인식 기술을 적용하는 시스템을 제안한다.

사람 재인식 기술은 카메라 각도와 조명의 변화에 많은 영향을 받지만, 최근에는 딥 러닝을 통해 성능이 크게 향상되었다 [1, 2]. Ahmed et al. [3]은 딥 네트워크를 사용하여 두 이미지에서 특징 맵을 추출하고, 추출된 특징 맵 간의 Cross-Input Neighborhood Difference (CIND)를 계산하여 두 이미지 사이의 차이에 대한 특징 맵을 새롭게 구성하였다. 이렇게 만들어진 특징 맵은 다시 네트워크를 통과하고 최종적으로 입력 이미지가 같은 사람인지 아닌지 판단한다.

기존 방법의 문제점은 두 개의 이미지를 사용하여 같은지 다른지 여부를 판별하기 때문에 차이점에 대한 부분만 학습하고 차이점과 공통점을 함께 고려하지 못한다는 점이다. 따라서 본 논문에서는 세 개의 이미지에서 추출된 특징을 두 쌍의 특징 맵으로 재구성한 후 학습에 사용하는 모델을 제안한다.

II. 본론

본 논문에서는 사람 재인식 정확도를 향상시키기 위해 IDLA [3] 아키텍처를 확장하였다. 세 개의 이미지를 이용하여 특징을 추출하고 추출된 특징 사이의 관계를 두 쌍의 특징 맵으로 재구성하도록 네트워크를

확장하였다. 확장된 네트워크는 기존보다 한 쌍의 특징 맵이 추가로 생성된다. 이러한 특징 맵은 마지막에 있는 전결합 레이어를 통해 정보를 공유하고 최종 결론을 내리게 된다.

그림 1은 세 개의 이미지에서 추출된 특징 맵들이 CIND 레이어를 통해 재구성됨을 보여준다. 식 (1)과 (2)는 두 특징 맵 사이의 관계를 계산하는 방법을 보여준다. f_i 는 컨볼루션 레이어를 통해 앵커 이미지로부터 추출된 특징 맵이고, $g_{i,1}$ 및 $g_{i,2}$ 는 각각 비교되는 이미지(image1/image2)로부터 추출된 특징 맵이다. Image1은 앵커 이미지와 동일한 사람에 대한 이미지이고, Image2는 다른 사람에 대한 이미지로 구성된다. $1(5,5)$ 는 1로 구성된 5×5 행렬을 의미하고 $N[g_{i,n}(x,y)]$ 는 $g_{i,n}$ 행렬에서 (x,y) 를 중심으로 하는 5×5 행렬을 의미한다. CIND 레이어는 2개의 이미지에서 한 이미지의 특징 좌표의 특징 값과 다른 이미지의 해당 좌표 주변의 특징 값의 차이를 계산하여, CIND 특징 맵 $K_{i,n}$ 를 생성한다.

$$K_{i,n}(x,y) = f_i(x,y) 1(5,5) - N[g_{i,n}(x,y)] \quad (1)$$

$$K_{i,n'}(x,y) = g_{i,n}(x,y) 1(5,5) - N[f_i(x,y)] \quad (2)$$

CIND 특징 맵 $K_{i,n'}$ 은 $K_{i,n}$ 의 반전시켜서 계산하는 것으로 식 (1)에서 f_i 와 $g_{i,n}$ 의 순서를 바꿔서 얻을 수 있다.

식 (1)과 (2)와 같은 방식으로 생성된 특징 맵은 컨볼루션 레이어를 통과하고 최종적으로 전결합 레이어와 소프트맥스 레이어를 통해 네 개의 클래스로 분류된다. 기존 IDLA 모델에서 사용된 두 개의 전결합 레이어로는 확장된 네트워크의 모든 특징 맵 정보를 완벽하게 처리할 수 없어서, 본 논문에서는 전결합 레이어를 추가하였다.

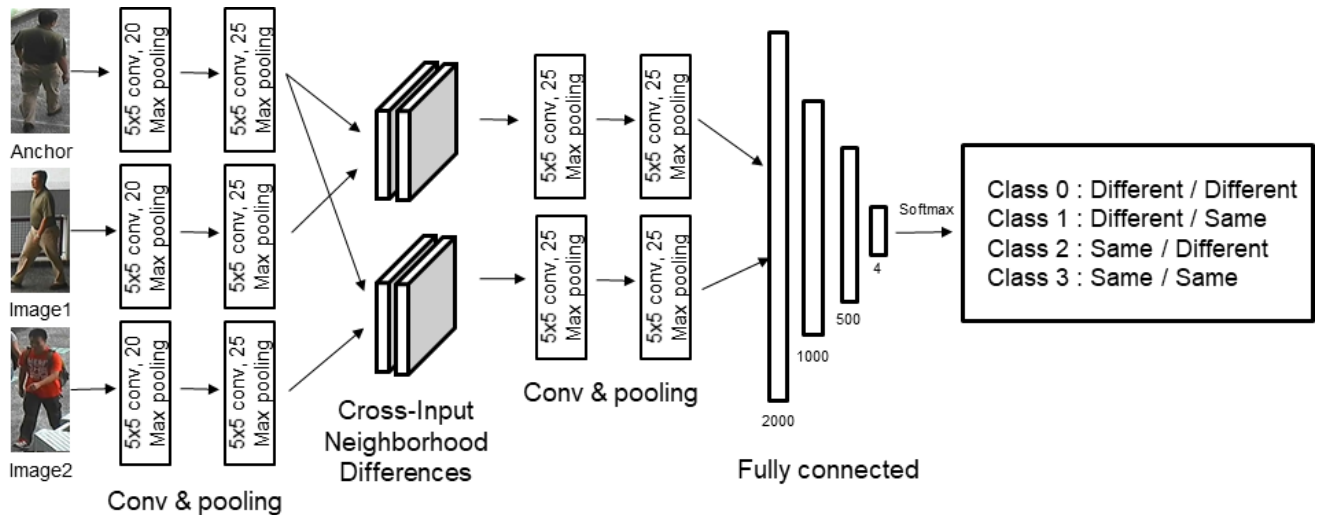


그림 1 제안하는 배우 재인식 모델

제안한다. 실험 결과 제안 모델이 영화/드라마와 같은 비디오 콘텐츠에도 적용 가능함을 확인하였다.

III. 실험결과

우리는 영화, 드라마를 대상으로 배우 재인식을 수행하기 위한 자체 데이터셋을 생성하였다. 60 여개의 영화와 드라마를 대상으로 자체 개발한 얼굴 인식 기술과 상체/의상 검출 기술을 같이 적용하여 데이터셋을 빠른 시간에 생성할 수 있었다. 상체 영역 데이터와 의상 영역 데이터는 각각 48,000 개 이미지 (2,235 클래스), 26,000 개 이미지 (1,994 클래스)의 구성되었다.



그림 2 영화/드라마로 구성된 자체 데이터셋

표 2 는 영화/드라마 의상 데이터세트에 대한 제안 시스템의 정확도이다. 갤러리 이미지의 수가 10 인 경우 Rank-1 정확도는 86.6%이며 갤러리를 구성하는 이미지의 수가 증가할수록 정확도는 감소하였다. 갤러리 이미지의 개수를 적절히 관리하면 적용 시나리오에 따라 사용 가능한 수준이다.

표 1 정확도(%)

Gallery No.	Rank-1	Rank-5
10	86.6	97.7
50	75.5	89.6

III. 결론

본 논문은 3 개의 이미지를 입력하여 이미지들 간의 차이점과 공통점을 동시에 학습할 수 있도록 하는 두 쌍의 CIND 특징 맵을 구성하는 사람 재인식 모델을

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국저작권위원회의 2020 년도 저작권보호 및 이용활성화 연구개발 지원사업으로 수행되었음

참 고 문 헌

- [1] Li, Wei, et al. "Deepreid: Deep filter pairing neural network for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
- [2] Lin, Yutian, et al. "Improving person re-identification by attribute and identity learning." Pattern Recognition, 2019.
- [3] Ahmed, Ejaz, Michael Jones, and Tim K. Marks. "An improved deep learning architecture for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.