얼굴과 신체의 영상 정보를 조합한 비제약적 환경에서의 사용자 인식

이 상호[1], 허 준희[2], 곽 노준[3]
서울대학교 융합과학기술대학원[1,3]
SK텔레콤 미래기술원[2]
dcmber12@snu.ac.kr[1], junhee.heu@sk.com[2], nojunk@snu.ac.kr[3]

요약

영상에서의 사용자 인식(Person identification)은 사람이 촬영된 영상 내에서 해당 인물이 누구인지 예측하는 것이다. 기존의 사용자 인식은 영상 내에서 사람의 신체 중 주로 얼굴 정보를 위주로 수행되었으나, 좀 더 비제약적 환경 하에서는 사람의 얼굴 정보만으로는 한계가 있었다. 이 연구에서는 한 개의 합성곱 신경망(Convolutional neural network)의 일반성을 이용하여 사람의 얼굴 정보뿐만 아니라 신체 정보도 함께 분석하여 사용자 인식을 수행하였으며, 얼굴 정보만을 가지고 수행한 사용자 인식에 비해 향상된 정확도를 얻을 수 있었다.

1. 서론

영상 사용자 인식은 ID 카드나 비밀번호 등 별도의 번거로운 정보 없이 영상만으로 인간을 구분할 수 있도록 하는 기술이며, 감시시스템, 생체 인식, 사물인터넷, 로봇 등등 분야로 활용도가 많다. 최근의 사용자 인식은 합성곱 신경망을 사용하여 특징맵(feature map)을 추출하는 방식으로 정확도가 매우 상승하였으며 이에 따라 그 활용도가 매우 높아졌다.

2. 관련 연구

[1, 2]는 합성곱 신경망을 활용하여 각각 97.35%, 98.95% 의 사용자 인식 정확도를 발표하였다. 그러나 합성곱 신경망을 활용한 사용자 인식은 비제약적 환경에서 영상이 정면으로 잘 정렬(align)되지 않거나, 조명, 얼굴 가림 등의 영향으로 정확도가 감소할 수 있다. [3]은 [2]를 포함한 최근의 얼굴 분류 연구를 비제약적 환경에서 수행함으로써 그 성능이 현저히 감소함을 보였다.

[4, 5, 6]등의 연구는 신체정보, 관계 및 그룹 정보 등등 얼굴 이외의 정보를 활용하여 비제약적 환경에서 정확도가 감소하는 현상을 보완하였다. 그러나 [4, 5, 6]의 연구는 한 종류의 정보에 대해서 한 개 이상의 신경망을 사용함으로써 신경망의 개수가 매우 많다는 한계가 있다. [4]의 경우는 총 107개의 신경망을 조합하여 사용자 인식을 수행한다. 이 경우 임베디드 환경 등 메모리에 한계가 있는 환경에서 적용하기가 어렵다.

본 연구에서는 한 개의 합성곱 신경망의 일반성을 활용하여, 비제약적 환경에서 한 개의 합성곱 신경망이 얼굴과 얼굴이외에 신체의 정보를 모두 분석함으로써 사용자 인식을 수행하였으며, 얼굴정보만을 이용한 방식보다 향상된 성능을 얻었다.

3. 제안하는 방법

본 연구에서는 여러 개의 신경망이 아닌 한 개의 합성곱 신경망을 사용하여, 영상 내에 등장하는 사람의 얼굴과 얼굴 이외의 신체영역에 대한 정보를 모두 한 개의 신경망에 통과시킴으로 추출한 특징맵(feature map)들을 기반으로 사용자 인식을 수행한다. 그림 1 은 본 연구에서 제안한 방식의 전체적인 과정을 요약하여 보여준다.

3.1 ROI 영역

그림 2 는 실험에 사용된 신체의 영역을 보여준다. 사용된 영역은 총 세 가지로, 머리영역(h)과 전신영역(b), 상체영역(u) 으로 이루어져 있다. 머리영역(h)은 본 실험에서 사용한 데이터셋에 주석으로 제공된 정보를 사용하였다. 전신영역(b)은 머리영역의 가로(w), 세로(h) 길이 중 더 작은 값($l = \min(w, h)$)을 기준으로, 머리영역을 상단 가운데에 밀착시킨 가로로 길이 2l, 세로로 길이 6l의 영역으로 정의하였다. 상체영역(u)은 전신역역의 상단 절반($2l \times 3l$)을 사용하였다[5, 6].

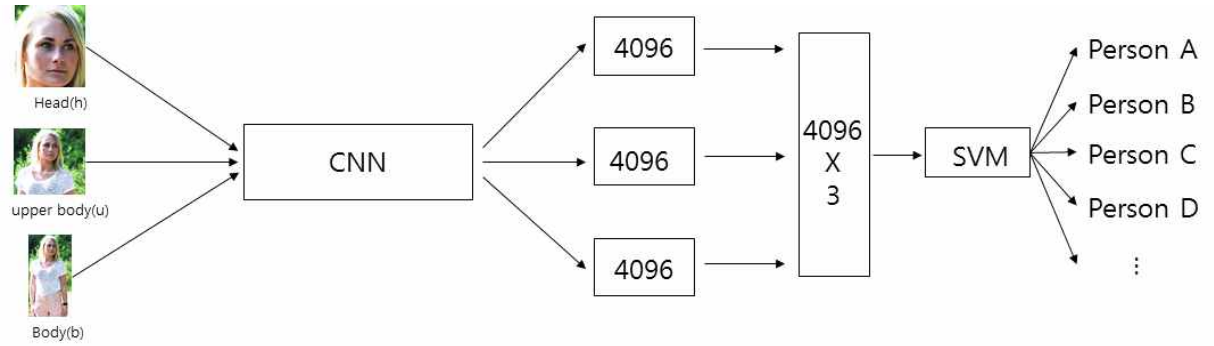

그림 1 제안하는 방법의 총체적인 모식도. PIPA 데이터셋의 인물사진은 얼굴(h), 상체(u), 전신(b) 영역으로 나뉘어져 하나의 합성곱 신경망(CNN)에 통과된 뒤, 신경망의 fc6 계층에서 4096 차원 특징맵을 생성한다. 해당 특징맵은 4096 × 3 차원으로 결합(concatenating)된 후에 SVM의 입력값으로 전달되어 최종 결과값(인물 정보)을 반환한다.

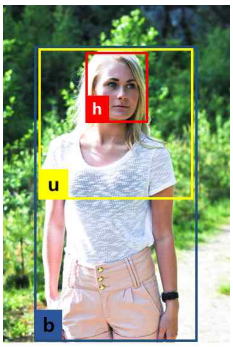

그림 2 사용자 인식에 사용한 신체의 부위. 빨간색의 머리영역(h)과 노란색의 상체영역(u), 파란색의 전신영역(b) 으로 이루어져 있다.

3.2 합성곱 신경망 및 SVM

실험에 사용된 합성곱 신경망은 [2]에서 제안된 VGG network를 사용하였다. 해당 합성곱 신경망에 얼굴영역, 신체영역의 정보를 통과시킨 후 각각 fc6 계층에서 4096 차원의 특징맵을 추출하였다.

해당 특징맵을 학습 데이터로 하여 linear SVM학습기를 학습시켰으며, SVM의 특수 매개변수(hyper parameter)는 C=1 으로 학습하였다[6].

4. 실험 및 분석

본 실험은 PIPA(People In Photo Album)[4] 데이터셋에서 수행되었다. PIPA 데이터셋은 train, test, validation, leftover로 나뉘어져 있다. 합성곱 신경망은 train 데이터에서 사람의 얼굴영역 정보를 이용해 미세조정(fine-tune) 되었다.

실험은 PIPA 데이터셋의 test 데이터를 사용하였다. 충분한 데이터를 학습시키기 위해, 한 인물 당 데이터가 25장 이상인 인물들만 추출한 뒤에 실험하였다. test 데이터를 두 하위 데이터셋으로 나눈 뒤, 각자 합성곱 신경망을 통과하여 특징맵을 얻었다. 이 후, 한 개의 하위 데이터셋에서 얻은 특징맵으로 SVM 분류기를 학습한 뒤, 나머지 한 개의 하위 데이터셋에서 얻은 특징맵으로 SVM 분류기를 테스트 하였다.

표 1 은 실험결과를 보여준다. 표 1 의 구분 A~B는 [2]의 신경망을 PIPA 데이터셋에 바로 적용한 결과이며, 구분 C~H는 PIPA 데이터셋의 train 데이터에서 얼굴영역을 이용하여 미세조정 한 실험 결과이다. 구분E의 정확도가 높았으며, 구분 D, E, F의 실험 결과를 비교해 보면 전신 정보는 얼굴과 함께 쓰이면 정확도를 약간 증가시킬 수 있지만 상체 정보보다는 정확도에 기여하는 바가 낮은 것을 볼 수 있다. 이는 얼굴 정보에 익숙한 합성곱 신경망이 전신보다는 상체에 대해 좀 더 정확한 표현력을 갖는 것으로 해석할 수 있다.

구분	사용한 정보	정확도
A	얼굴	85.81%
B	전신	52.9%
C	얼굴	86.32%
D	상체	64.96%
E	전신	53.97%
F	얼굴 + 전신	86.94%
G	얼굴 + 상체	**87.76%**
H	얼굴 + 상체 + 전신	87.64%

표 1 인물의 얼굴, 신체 정보 조합에 따른 합성곱 신경망과 SVM의 정확도

5. 결론

본 연구는 비제약적 환경에서 얼굴 분류의 정확도를 향상시키기 위해 사람의 얼굴 정보와 얼굴 이외 신체의 정보를 하나의 합성곱 신경망에 적용하는 연구를 수행하였다.

합성곱 신경망은 얼굴정보만 학습하였지만 그 특징맵의 구분능력은 인물의 상체, 전신 등 얼굴이 아닌 정보로도 일반성을 확장시킬 수 있었다.

참고문헌

[1] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[2] Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." British Machine Vision Conference. Vol. 1. No. 3. 2015.

[3] Ghazi, Mostafa Mehdipour, and Hazim Kemal Ekenel. "A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition." arXiv preprint arXiv:1606.02894 (2016).

[4] Zhang, Ning, et al. "Beyond frontal faces: Improving person recognition using multiple cues." 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.

[5] Li, Haoxiang, et al. "A Multi-Level Contextual Model For Person Recognition in Photo Albums."

[6] Joon Oh, Seong, et al. "Person recognition in personal photo collections." Proceedings of the IEEE International Conference on Computer Vision. 2015.