인간-컴퓨터 상호작용을 위한 CNN 기반 객체 검출

박명숙, 김상훈*
한경대학교 전기전자제어공학과
e-mail:kimsh@hknu.ac.kr

CNN-based Object Detection for Human-Computer Interaction

Myeong-Suk Pak, Sang-Hoon Kim*
Dept of Electrical, Electronic and Control Engineering, Hankyong National University

요 약

비전 기반 제스처 인식은 비 침입적이고 저렴한 비용으로 자연스러운 인간-컴퓨터 상호 작용을 제공한다. 로봇의 사용이 증가함에 따라 인간-로봇 상호 작용은 점점 더 중요해질 것이다. 최근 효율적인 딥러닝 기술이 연구되고 있다. 본 연구는 인간 컴퓨터 상호 작용을 위해 CNN을 기반으로 한 얼굴 및 손 동작의 인식을 위해 객체 검출 기법의 적용 결과를 제시한다.

1. 서론

현재 생활에서 사람뿐만 아니라 컴퓨팅 장치와의 상호 작용이 필요해졌으며, 대부분의 컴퓨터 응용 프로그램에서 효과적으로 사용하기 위해 점점 더 많은 상호 작용이 필요하다 [1]. 비전 기반 제스처 인식은 비 침입적이고 저렴한 비용으로 자연스러운 인간-컴퓨터 상호 작용을 제공한다. 최근에 숙제 로봇, 엔터테인먼트 로봇, 보조 로봇, 가정용 감시 로봇과 같은 "인간 친화적" 시스템이라는 로봇은 식당, 병원 및 서비스 지역에서 점점 더 자주 등장했다 [2]. 따라서 인간-로봇 상호 작용이 점차 중요해질 것이다.

지난 몇 년 동안 객체 분류 및 탐지에서 우수한 성능을 보여주는 딥 러닝 기술이 발표되었다. 최근에, 모바일 장치에 적용 가능한 효율적인 기술이 연구되었고, 물체 검출 기술과 함께 좋은 결과를 보여 주었다. 본 연구는 인간 컴퓨터 상호 작용을 위한 CNN을 기반으로 한 얼굴 및 손 동작의 인식에 중점을 둔다. 효율적인 CNN 기술을 조사하고 움직임 인식의 이전 단계로서 얼굴의 검출에 대해 실험한다.

2. CNN 기반 객체검출

2.1 비전기반 인간-컴퓨터 상호작용

컴퓨터 비전을 사용하여 사용자와 상호 작용하기 위해 전신 운동, 얼굴 움직임 및 손 움직임 [3]의 세 부분으로 나눌 수 있다. 우리는 얼굴과 손 움직임의 인식에 중점을 둔다.

얼굴 움직임에는 얼굴 인식, 머리 및 얼굴 추적, 눈 추적 및 표정 인식이 포함된다. 이를 위해서는 얼굴 인식이 선행되어야 한다. 손 동작의 사용은 인간의 컴퓨터 상호 작용을 위한 이러한 성가신 인터페이스 장치에 대한 매력적이고 자연스러운 대안을 제공하고 인간의 컴퓨터 상호 작용에 필요한 용이성과 자연성을 제공한다 [1]. 비전 기반 손 제스처 인식 시스템은 감지, 추적 및 인식의 세 단계로 구성된다. 손 제스처 인식 시스템의 주요 단계는 손을 감지하는 것이다.

2.2 객체검출을 위한 CNN 기법

얼굴 및 손 동작 인식을 위해서는 먼저 얼굴 및 손의 검출을 수행해야한다. 이 절에서는 객체 검출을 위한 CNN 기술을 살펴본다.

MobileNetV1 [4]는 모바일 환경에 적합한 아키텍처에 대해 발표되었으며, 깊이 컨볼 루션과 포인트 컨볼 루션을 사용하여 계산을 줄이는 네트워크이다. MobileNetV2 [5]는 기존 bottlenect 구조와 다르게 확장 방향으로 변경하여 계산 복잡성을 개선한다. 이 방법은 높은 정확도와 높은 메모리 및 속도를 갖는다. ShuffleNetV2 [6]는 기존 모바일 환경을 위한 많은 아키텍처가 FLOP를 염두에 두고 설계되었지만 속도 및 대기 시간과 같은 직접 메트릭을 고려해야한다고 강조한다. 효율적인 네트워크 설계를 위해 채널 수를 동일하게 유지하고 그룹 컨볼루션 비용을 인식하며 단편화를 줄이며 요소 별 작업을 줄인다. 실험에서 대부분의 지수에서 최고의 성능을 보여주었다.

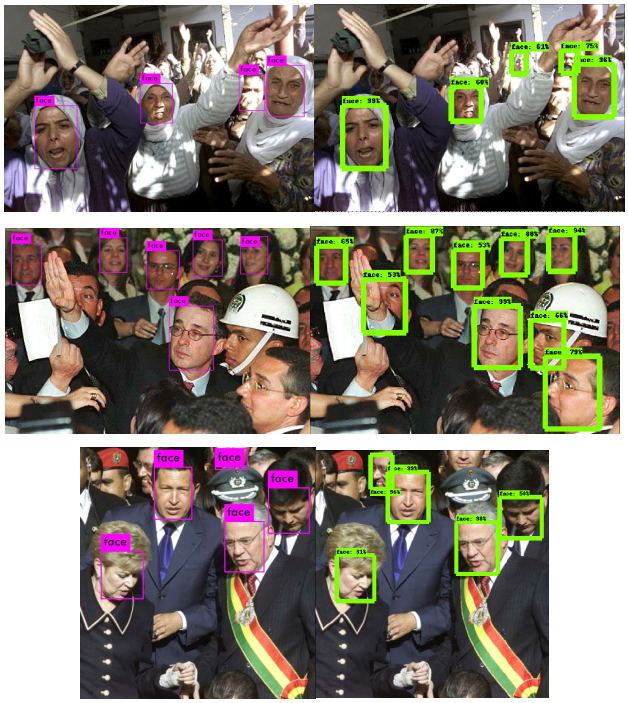
객체 검출을 위한 CNN 기술에서, 단일 스테이지 검출기는 성능과 속도에서 모두 만족스러운 결과를 보여주었다. SSD [7]는 객체의 후보 영역을 검출하고 픽셀 또는 특징을 리샘플링하는 2 단계 검출 방법과 달리, 이러한 프

로세스를 제거하고 계산 시간을 감소시켜 실시간 객체 검출 성능을 보여준다. 높은 검출 정확도를 얻기 위해 크기가 다른 특징 맵과 종횡비가 다른 기본 상자를 사용한다. 이는 속도와 정확성의 균형을 향상시킨다. MobilNetV2 [5]와 함께 사용하면 객체 검출 성능이 우수하다. YOLO [8]는 단일 컨볼루션 신경망을 통해 단일 이미지에서 경계 상자와 클래스 확률을 동시에 예측한다. 네트워크는 이미지를 그리드 셀로 나누고 객체의 중심점이 있는 그리드의 경계 상자를 추정하고 신뢰 점수를 계산한다. YOLOv2에서 Darknet-19 및 YOLOv3는 Darknet-53이라는 새로운 네트워크를 사용하여 이전 버전보다 강력하다. 320 x 320에서 YOLOv3는 SSD [7]만큼 정확하지만 3 배 더 빠르다. 라이트 버전 Tiny YOLO [8]는 각 버전에 포함되어 있으며 빠르고 모델 크기가 작다.

3. 실험

이 장에서는 Tiny YOLOv3와 SSD-MobilenetV1을 사용한 얼굴 검출 실험에 대하여 설명한다. 실험은 NVIDIA GeForce RTX 2080을 사용하였다. 데이터 셋은 WIDER FACE[11]를 이용하여 훈련하였고 FDDB[12]를 이용하여 테스트하였다.

그림 1은 얼굴 검출 결과를 보여준다. 왼쪽은 Tiny YOLOv3의 결과이고 오른쪽은 SSD-MobilenetV1의 결과이다. 일부 가려진 얼굴에 대해 Tiny YOLOv3가 잘 찾는 경우가 있지만, SSD-MobilenetV1이 더 많은 얼굴을 찾는 것으로 나타났다.



(그림 1) 얼굴 검출 결과

4. 결론

로봇의 사용이 증가함에 따라 인간-로봇 상호 작용이 점점 더 중요해지고 있으며, 최근 효율적인 딥러닝 기술이 연구되고 있다. 본 연구는 인간 컴퓨터 상호 작용을 위해 CNN을 기반으로 한 얼굴 및 손 동작의 인식의 이전 단계로서 얼굴 검출에 대해 실험하였다. 실험에 이용한 두 알고리즘 모두 방향, 스케일, 가려짐 등이 있는 경우에도 얼굴을 검출하였으며 SSD-MobilenetV1이 더 많은 얼굴을 검출하였다. 다음으로 손 검출에 대한 실험이 필요하고, 검출된 얼굴과 손의 동작인식에 대한 연구를 진행하고자 한다.

참고문헌

[1] Siddharth S. Rautaray, Anupam Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, Artificial Intelligence Review 43(1) (2015) 1-54

[2] Wei He, Zhijun LI and C.L.P. Chen, A survey of human-centered intelligent robots: Issues and challenges, IEEE/CAA J. of Autom. Sinica vol.4, no. 4 (2017) 602-609

[3] Jason J. Corso, TECHNIQUES FOR VISION-BASED HUMAN-COMPUTER INTERACTION, The Johns Hopkins University Baltimore, Maryland, August (2005)

[4] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017)

[5] Sandler M., Howard A., Zhu M. Zhmoginov A. and Chen L.C., MobileNetV2: Inverted Residuals and Linear Bottlenecks, arXiv preprint arXiv:1801.04381, (2018)

[6] Ma N., Zhang X., Zheng H.T. and Sun J., ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design, arXiv preprint arXiv:1807.11164, (2018)

[7] Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.Y., Berg A.C., SSD: Single Shot MultiBox Detector, European Conference on Computer Vision, (2016) 21-37

[8] Redmon J. and Farhadi A., YOLOv3: An incremental improvement, arXiv preprint arXiv:1804.02767, (2018)