

CNN 기반의 포즈 매크로 시스템

한재혁¹, 이상무^{1*}, 김미혜^{1*}

¹충북대학교 컴퓨터공학과

e-mail: allsog@naver.com, mhkim@cbnu.ac.kr

CNN based Pose Macro System

Jae-Hyeock Han¹, Sang-Mu Lee^{1*}, Mi-Hye Kim^{1*},

¹Dept of Computer Engineering, Chungbuk University

요 약

CNN(Convolution Neural Network)을 사용해 동작을 학습시키고, 사용자의 동작을 카메라로 촬영한다. 촬영한 동작들을 인식해서 인식한 동작들의 조합을 확인하고 조합에 해당되는 값을 출력한다. 출력된 값을 통해 간단한 동작 인식으로 다양한 매크로들을 실행시킬 수 있다.

1. 서론

유비쿼터스에 대한 연구가 이루어지면서 편의와 스마트 홈 환경 등을 조성하기 위한 여러 인터페이스 시스템에 대한 연구도 이루어지고 있다. 그 중 영상으로부터 인간의 동작을 인식하는 것은 진보적인 인터페이스 구축을 위하여 매우 중요한 핵심기술이라고 할 수 있다[1]. 본 논문에서는 CNN 을 사용하여 카메라 이미지로 손 동작을 인식하고 인식된 손 동작의 조합을 인식해서 매크로로 시스템을 제어할 수 있는 인터페이스 환경을 제안한다.

2. 관련연구

동작 인식 방법과 알고리즘들에 대한 연구는 지속적으로 진행되어 왔으며, 많은 모션 인식 방법들이 개발되었다. 주로 HMM(Hidden Markov Model), 웨어러블 장비, CNN 을 사용한 인공지능 학습 등이 대표적인 동작 인식 알고리즘, 방법이라고 할 수 있다[2,3,4].

여기서 CNN 은 인공지능망에서 딥 러닝을 위해 사용하는 심층 신경망 중 하나로 인간 뇌의 시각 피질에서 영상을 처리하는 것과 유사한 기능을 하는 신경망이다. CNN 은 특징을 직접 학습하기 때문에 학습에 필요한 특징을 수동으로 추출해야 할 필요가 없다[5]. 그리고 높은 수준의 인식 결과를 보이기 때문에 카메라로 촬영된 이미지를 바로 인공지능망의 학습에 사용할 수 있다. 즉 카메라로 촬영된 이미지를 사용해 CNN 으로 학습된 인공지능망은 카메라로 촬영된 이미지를 바로 인식하는 것이 가능하다.

이와 같이 CNN 을 사용해서 인공지능망을 학습시킨다면 다른 신경망에 비해서 학습 데이터 준비와 동작 인식에 사용할 데이터 준비의 오버헤드를 줄일 수 있다.

3. 인공지능망 학습

본 논문에서는 Tensorflow 와 CNN 구조의 Inception v3 모델을 사용해서 인공지능망을 학습시킨다. Inception v3 모델은 패턴 분류에서 한 번에 여러 크기의 필터를 동시에 사용하는 것으로 다른 CNN 기반의 Oxford VGG(Visual Geometry Group) 모델 등에 비해 이미지의 특징을 더 잘 잡아낸다[6].

이미지 인식에 대해 검증된 모델인 Inception v3 모델의 모델 구조와 parameter 들을 사용해서, 이를 기반으로 새로운 데이터셋에 retraining 을 진행하는 Transfer Learning 방식을 사용한다.

Tensorflow 를 사용한 Inception v3 모델에 retraining 에서는 각각의 데이터셋을 폴더로 나누고 학습데이터로 입력하는 것으로 이미지 데이터셋에서 학습 데이터와 실험 데이터를 무작위로 나누고, 나뉜 데이터들을 사용해서 학습을 하게 된다.

표 1. 다양한 각도의 이미지 인식률

	인식률(%)				
실험데이터	(a)	(b)	(c)	(d)	(e)
포즈 A	85%	91%	99%	99%	91%
실험데이터	(f)	(g)	(h)	(i)	(j)
포즈 B	88%	91%	73%	90%	90%

학습한 인공지능망에 학습에 사용한 데이터셋의 이미지를 제외한 새로운 이미지를 사용해서 인식 실험을 했을 때 표 1 과 같이 포즈 A 이미지에서는 85% 이상, 포즈 B 이미지에 대해서는 73% 이상의 인식률을 보여줬다.

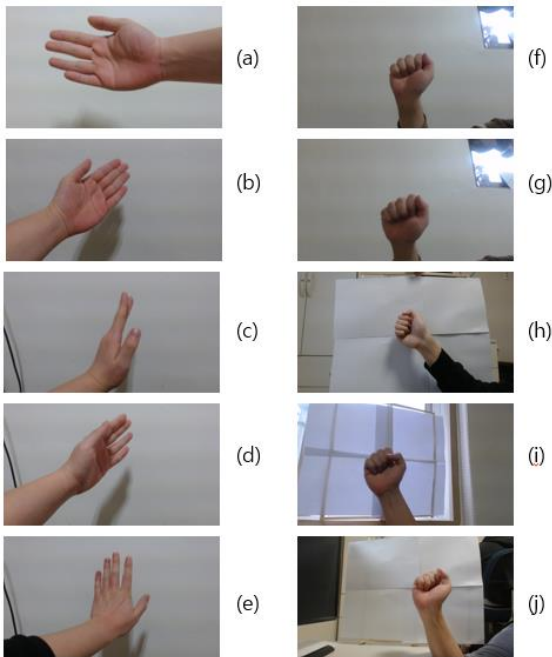


그림 2. 실험에 사용된 이미지

4. 시스템 환경

본 논문에서는 카메라와 중앙 시스템 그리고 그와 연동된 장치들로 시스템을 구성했다. 카메라로 촬영된 이미지를 학습된 인공신경망을 가진 중앙 시스템으로 전송하고, 중앙 시스템에서는 전송된 이미지들의 종류와 순서를 조합하여 그에 맞는 매크로 프로그램을 실행해서 연동된 장치들을 작동시키게 된다.

5. 구현

본 논문에서 제안한 CNN 의 Transfer learning 과 시스템 환경을 사용해서 시스템의 초기형을 구현하였다.

각각의 동작에 200 장 이상의 영상을 촬영하고 이미지 데이터셋을 3 가지로 나누고 각각을 펼쳐진 손, 주먹을 쥔 손, 그리고 오류를 검출하기 위한 배경 이미지의 데이터셋으로 구성하였다. 이 학습 데이터로 Transfer learning 을 사용해서 동작을 인식할 수 있게 한 다음, 카메라로 사용자의 손을 3 번 촬영하고, 3 번의 동작 인식을 통해 미리 설정된 매크로로 컴퓨터를 제어할 수 있도록 하였다. 이 시스템의 경우 다른 추가적인 장비 없이 카메라만으로 매크로를 통해 컴퓨터를 제어할 수 있다. 또한, 두 종류의 동작과 그 동작들의 조합으로 사용자는 많은 학습 없이 기기 제어를 할 수 있다.

6. 결론

본 논문에서는 CNN 을 사용한 딥 러닝으로 동작 인식을 제공하고, 2 가지 동작의 조합을 통한 동작 인식이 가능하도록 한 인터페이스를 구현하였다. 간단한 동작과 그 동작들의 조합으로 기기를 제어함으로써 사용자는 적은 학습으로 시스템을 사용할 수 있고, 인식해야 되는 동작이 적으므로

인식 실패나 잘못된 인식의 경우가 줄어든다. 동작 조합의 직관성과 인식 정확도에 대한 과제를 앞으로의 연구로 해결한다면, 동작 조합해서 사용하는 것을 앞으로의 동작 인식을 통한 기기 제어에 대해 하나의 방법으로 사용할 수 있을 것이다.

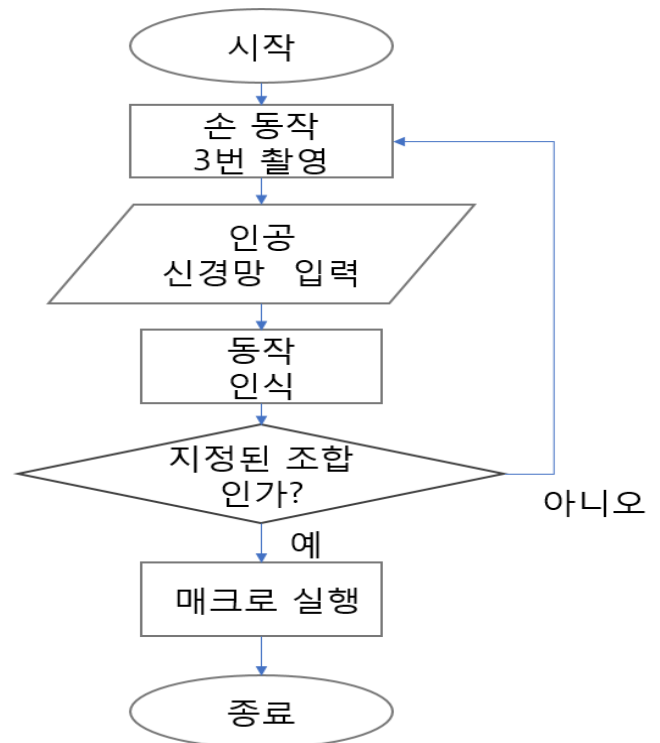


그림 3. Pose Macro System 의 순서도

참고문헌

- [1] 김호준, “CNN 모델과 FMM 신경망을 이용한 동적 수신호 인식 기법”, 지능정보연구 제 16 권 제 2 호, pp. 95-108, 2010.
- [2] 양수명, 송원재, 최익수, 유상조, “초음파 환경에서의 사물인터넷 기기 제어를 위한 딥러닝 기반 동작 인식 시스템 구현”, 한국통신학회논문지, 제 42 권, 제 9 호, pp. 1796-1805, 2017.
- [3] 김아람, 이상용, “HMM 을 이용한 자연스러운 손동작 인식”, 한국지능시스템학회 논문지, 제 22 권, 제 5 호, pp. 639-645, 2012.
- [4] 박인혜, 이상엽, 고재진, “차량 내 제스처 인식을 위한 밴드형 웨어러블 디바이스 개발 및 서비스에 관한 연구”, 한국통신학회지(정보와 통신), 제 33 권 제 6 호, pp. 39-46, 2016.
- [5] Hsien-I Lin, Ming-Hsiang Hsu and Wei-kai Chen, “Human hand gesture recognition using a convolution neural network”, 2014 IEEE Inter. Conf. on Automation Science and Engineering (CASE), 10.1109/CoASE.2014.6899454, 2014.
- [6] Christian Szegedy, et al., “Going Deeper with Convolutions”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-9, 2015.