

DNN을 이용한 낮은 복잡도를 가지는 손 동작 인식에 관한 연구

송호성, 이원석

한국의국어대학교, 세종대학교

kaiger0313@naver.com, scu008@nate.com

A Study on the Hand Gesture Recognition with Low Complexity Using DNN

Ho Sung Song, Won Seok Lee

Korea Foreign Univ., Sejong Univ.

요약

손 동작을 인식하여 여러가지 사물의 기기를 올바르게 제어하기 위하여 다양한 인공지능 기술이 이용되고 있으며, 그 중에서도 CNN(Convolutional Neural Network) 기술은 성능이 우수하여 많이 이용되고 있다. 그러나 CNN 기법은 성능은 우수하지만, 높은 복잡도를 가지고 있는 단점이 있다. 따라서 본 논문에서는 손 동작 인식을 위하여 낮은 복잡도를 가지면서 CNN과 유사한 성능을 갖도록 전처리 과정을 거치는 DNN(Deep Neural Network) 기법을 제안한다. 이를 위하여 입력으로 받은 영상으로부터 손 동작을 인식하기 위하여 Canny 알고리즘을 이용하여 엣지를 검출한 후에, DNN 기법을 이용하여 낮은 복잡도를 갖는 알고리즘을 연구하였다. 이러한 알고리즘은 손 동작을 통한 다양한 사물 인터넷 기기의 제어와 응용에 활용가능할 것으로 예상된다.

I. 서론

최근에 사물 인터넷의 실현을 위하여 다양한 기기가 확산되면서 손 동작 인식을 통해 여러가지 기기를 제어하는 기술이 주목받고 있다. 특히, 손을 자유로이 사용하여 사용자의 명령을 내릴수 있다는 점에서 손 동작 인식 및 제어 기술에 대한 연구가 활발히 진행 중이다. 이러한 것을 구현하기 위해서는 손 동작의 정확하고 효율적인 검출 및 인식 기능이 요구된다[1], [2]. 최근 영상 인식 분야에서는 사람이 직접 영상에서 특징을 추출하여 영상을 인식하는 방법 대신 인공지능 기술인 딥러닝(Deep Learning)을 사용하여 사람이 미처 인지하지 못하는 특징을 이용하는 방법이 각광을 받고 있다. 딥러닝이란 인간의 뇌에 있는 뉴런의 동작을 모방한 인공 신경망을 여러 층 쌓은 심층 신경망을 이용하는 방법이다.

최근에는 손 동작 인식 분야에서 딥러닝 알고리즘을 많이 적용하고 있으며, CNN(Convolutional Neural Network)은 그 대표적인 알고리즘이다 [3], [4], [5]. 그러나 CNN 알고리즘은 성능은 우수하지만 높은 복잡도를 가지는 단점이 존재한다. 따라서 본 논문에서는 손 동작 인식 주변 환경에 따라서 CNN보다 낮은 복잡도를 가지면서 유사한 성능을 가지는 DNN(Deep Neural Network) 기법을 제안한다. DNN 기법에 적용하기 전에 전처리 과정으로 입력 받은 손의 모양을 Canny 알고리즘을 이용하여 엣지를 검출한 후에, DNN 기법을 이용하여 낮은 복잡도를 가지도록 설계하였다. 이러한 기술은 손 동작을 통해 사물 인터넷의 다양한 기기들의 제어를 포함하여 다양한 응용에 활용가능할 것으로 예상된다.

II. 본론

최근에는 영상 등의 사물을 인식하기 위한 방법으로 다양한 인공지능 기법들이 사용되고 있으며, 그 중에서도 우수한 성능을 얻을 수 있는 CNN 기법이 많이 연구되고 있다. 초창기에는 레이어의 모든 노드가 다음 레이어의 모든 노드에 연결되는 FC Layer(Fully Connected Layer)를 가지고 있었으며, 1차원적 배열 형태로 한정되어 있는 학습이었다. 그러나

대부분의 이미지는 모두 3차원 이상의 데이터들이다. 3차원 이상의 데이터를 1차원 안에서 학습시키는 것은 어렵다. 평면화 시키는 과정에서 공간 정보가 손실이 되기 때문에 이미지 공간정보 유실로 인한 정보 부족으로 인공 신경망이 특징을 추출 및 학습하는 것이 비효율적이고 정확도를 높이는 데 한계가 있다. 그래서, 이미지의 공간정보를 유지한 상태에서 학습이 가능한 모델이 바로 CNN이다. CNN은 각 레이어의 입력력 데이터 형상을 유지하고, 이미지의 공간 정보를 유지하면서 인접 이미지와의 특징을 효과적으로 인식할 수 있고, 2개 이상의 필터로 3차원 이상의 이미지를 특징을 추출해서 학습을 시킬수 있다. 장점으로 필터를 공유 파라미터로 사용하기 때문에, 이전의 인공 신경망과 비교할 때, 학습 파라미터가 매우 적은 것이 장점이다. CNN 기법은 여러 레이어가 있다는 점은 DNN과 동일하지만, 레이어의 일부에 컨벌루션 레이어를 사용한다. 컨벌루션 레이어는 이전 레이어와 현재 레이어의 노드가 모두 서로 연결되는 것이 아니라 컨벌루션 커널을 통하여 연결됨으로써 파라미터의 개수를 레이어 크기의 곱이 아닌 컨벌루션 커널의 크기로 축소한다. 활성화 함수(Activation Function)는 컨벌루션 레이어를 통해 계산된 값을 다음 레이어로 넘기기 위해 사용하는 함수로, sigmoid, ReLu(Rectified Linear Unit) 등을 사용한다. 풀링(Pooling) 레이어는 레이어의 출력을 줄여 네트워크의 사이즈를 줄이고 중요한 특징만을 얻기 위해 사용한다. CNN 기법은 우수한 성능을 가지지만 복잡도가 큰 단점을 가지고 있다. 따라서 본 논문에서는 특정한 환경에서는 CNN과 유사한 성능을 가지면서도 복잡도를 줄이기 위하여 전처리 과정을 거치는 DNN 기법을 이용한 손 동작 인식 기법을 제안한다. 이를 위하여 입력으로 받은 손 동작의 엣지를 검출한 후에 DNN 기법을 이용하여 낮은 복잡도를 가지도록 한다.

입력으로 받은 영상 신호에서 엣지(edge)는 영상의 특징을 기술하기 때문에 중요한 시각 정보를 갖는다. 영상 신호의 엣지를 검출하는 방법으로는 Sobel, Laplacian, Canny 등의 다양한 검출 기법이 존재한다. 그러나 Sobel 함수나 Laplacian과 같은 단순 이진 엣지 검출 기법은, 검출한 엣지

가 필요 이상으로 두꺼워 객체를 식별하기가 어렵게 되며, 영상의 모든 중요한 엣지를 검출하기 위한 명확한 경계값을 찾기가 어려워질 수 있다.

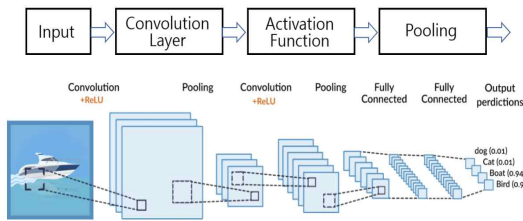


그림 1. CNN 구조도

따라서 많은 경우에는 이러한 문제를 해결하기 위하여 Canny 알고리즘을 사용하며, 다음과 같은 절차를 통하여 엣지를 검출하게 된다. 첫번째로 가우시안 필터링을 하여 영상을 부드럽게 한다. 두번째로 Sobel 연산자를 사용하여 기울기(gradient) 벡터의 크기(magnitude)를 계산한 후, 세번째로 가나단 엣지를 얻기 위해 기울기 벡터 방향에서 기울기 크기가 최대 값인 화소만 남기고 나머지는 0으로 만든 후, 마지막으로 연결된 엣지를 얻기 위해 두 개의 임계값을 사용한다. 높은 값의 임계값을 사용하여 기울기 방향에서 낮은 값의 임계값이 나올 때까지 추적하며 엣지를 연결하는 히스테리시스 임계값(hysteresis thresholding) 방식을 사용한다. 따라서 본 논문에서는 엣지 검출을 위하여 Canny 전처리 과정을 사용하였다. (그림 2 참조).

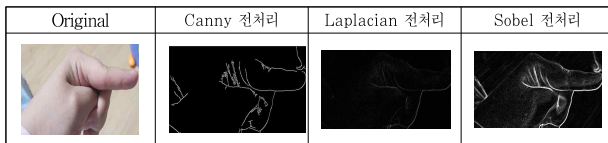


그림 2. 엣지 검출 전처리 과정

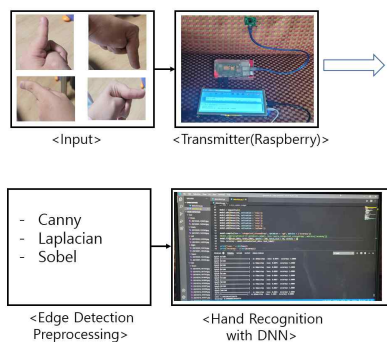


그림 3. 전처리와 DNN을 이용한 손동작 인식 구조도

그림 3은 엣지 검출의 전처리 과정을 거친 후에 DNN 기법을 적용하는 손 동작 인식의 전체 구조도를 나타내며, 그림 4는 실험 환경을 나타낸다. CNN보다 복잡도를 줄이기 위하여 입력된 손 동작의 엣지만을 검출하기 위한 전처리 과정을 사용하였다. 엣지를 검출한 후에 DNN을 기법을 적용하였다. 응용의 예로서 사물의 기기에 있는 카메라를 통하여 손 동작을 입력 받은 후에 Raspberry를 통하여 손 동작 인식을 처리하기 위한 장치로 보내 후에, 손 동작을 인식하여 기기를 제어하게 된다. 표 1과 2는 실험을 통한 손 동작 인식의 Epoch(반복횟수), Loss(손실 및 오차 계산), Accuracy(정확성)을 나타낸다. 표 1은 학습한 손 동작과 유사하며, 주변의 환경도 학습한 것과 유사할 때 실험한 결과를 나타내며, 표 2는 학습한 손 동작의 모습과 다르며 주변 환경도 학습할 때와는 다른 환경에서 실험

한 결과이다. 학습과 주변 유사 환경에서는 epoch을 증가시킴에 따라서 정확도가 100%에 이른다. 그러나 학습과 주변이 다른 환경에서의 손 동작 인식 성능은 epoch에 따라 정확도가 증가하다가 한계점에 도달하게 된다. 이러한 엣지 검출 전처리 과정을 거친 후에 DNN 기법을 적용한 손 동작 인식의 성능은 학습과 주변 환경에 따라서 CNN 보다 복잡도는 감소하지만 유사한 성능을 가지게 된다. 향후에는 좀더 CNN 기법과 유사한 성능을 가지기 위한 엣지 검출의 전처리 과정과 DNN의 최적화 파라미터에 대한 연구가 진행되어야 할 것이다. 그러나 이러한 낮은 복잡도를 가지는 인공지능 기술은 다양한 사물 인터넷 기기의 제어에 활용될 가능성이 있다.

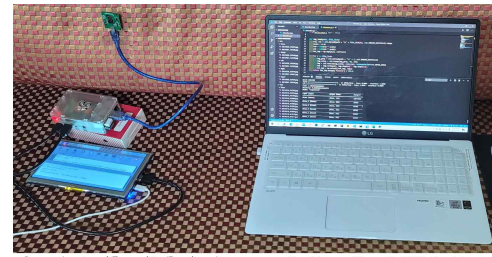


그림 4. DNN을 이용한 손동작 인식 실험 환경

표 1. 학습과 주변 유사 환경에서 손 동작 인식 성능

epochs	20	40	60	80	100
loss	1.09	0.59	0.43	0.14	0.09
accuracy(%)	40	80	80	100	100

표 2. 학습과 주변 다른 환경에서 손 동작 인식 성능

epochs	20	40	60	80	100
loss	1.36	1.46	1.17	1.58	1.42
accuracy(%)	40	40	70	70	70

III. 결론

본 논문에서는 CNN 기법보다 낮은 복잡도를 가지면서 유사한 성능을 가지는 전처리를 결합한 DNN 기법을 제안한다. 이를 위하여 입력으로 받은 영상 데이터를 Canny 알고리즘을 이용하여 엣지를 검출하는 전처리 과정을 수행한 후에, DNN 기법을 적용하는 방법을 제안하였다. 이러한 기법은 손동작을 통한 다양한 사물 인터넷 기기의 제어와 응용에 활용 가능할 것으로 예상된다.

참 고 문 헌

- [1] 지명근, 전주철, "딥 러닝을 이용한 영상 인식 기술 동향," 한국인터넷 정보학회논문지, 제18권 제2호 pp. 26-33, 2017.
- [2] 이병희, 오동환, 김태영, "딥러닝 기반 손 제스처 인식을 통한 3D 가상 현실 게임," 컴퓨터그래픽학회논문지, 제24권 제5호 pp. 41-48, 2018.
- [3] 정창영, 김태용, "손 제스처 인식을 위한 특징 강화 CNN 알고리즘과 사용자 의사 결정 판단," 전자공학회논문지, 제57권 제2호 pp. 60-70, 2020.
- [4] 김문환, 황선기, 배철수, "손동작 인식 시스템을 위한 동적 학습 알고리즘," 한국정보통신기술학회논문지, 제2권 제2호 pp. 51-56, 2009.
- [5] 홍성진 외 6인, "차세대 콘텐트를 위한 AI 기술 활용 동향 및 전망," 전자통신동향분석, 제35권 제5호 pp. 123-133, 2020.