

한국차세대컴퓨팅학회 논문지
Vol.14 No.6

ISSN : 1975-681X(Print)

이중흐름 3차원 합성곱 신경망 구조를 이용한 효율적인 손 제스처 인식 방법

최현중, 노대철, 김태영

To cite this article : 최현중, 노대철, 김태영 (2018) 이중흐름 3차원 합성곱 신경망 구조를 이용한 효율적인 손 제스처 인식 방법 , 한국차세대컴퓨팅학회 논문지, 14:6, 66-74

① earticle에서 제공하는 모든 저작물의 저작권은 원저작자에게 있으며, 학술교육원은 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

이중흐름 3차원 합성곱 신경망 구조를 이용한 효율적인 손 제스처 인식 방법

An Efficient Hand Gesture Recognition Method using Two-Stream 3D
Convolutional Neural Network Structure

최현중, 노대철, 김태영¹⁾

Hyeon-Jong Choi, Dae-Cheol Noh, Tae-Young Kim

(02713) 서울특별시 성북구 서경로 124 서경대학교 컴퓨터공학과
{hip4652, sheocjf1025, tykim}@skuniv.ac.kr

요 약

최근 가상환경에서 몰입감을 높이고 자유로운 상호작용을 제공하기 위한 손 제스처 인식에 대한 연구가 활발히 진행되고 있다. 그러나 기존의 연구는 특화된 센서나 장비를 요구하거나, 낮은 인식률을 보이고 있다. 본 논문은 정적 손 제스처와 동적 손 제스처 인식을 위해 카메라 이외의 별도의 센서나 장비 없이 딥러닝 기술을 사용한 손 제스처 인식 방법을 제안한다. 일련의 손 제스처 영상을 고주파 영상으로 변환한 후 손 제스처 RGB 영상들과 이에 대한 고주파 영상들 각각에 대해 텐스넷 3차원 합성곱 신경망을 통해 학습한다. 6개의 정적 손 제스처와 9개의 동적 손 제스처 인터페이스에 대해 실험한 결과 기존 텐스넷에 비해 4.6%의 성능이 향상된 평균 92.6%의 인식률을 보였다. 본 연구결과를 검증하기 위하여 3D 디펜스 게임을 구현한 결과 평균 34ms로 제스처 인식이 가능하여 가상현실 응용의 실시간 사용자 인터페이스로 사용가능함을 알 수 있었다.

Abstract

Recently, there has been active studies on hand gesture recognition to increase immersion and provide user-friendly interaction in a virtual reality environment. However, most studies require specialized sensors or equipment, or show low recognition rates. This paper proposes a hand gesture recognition method using Deep Learning technology without separate sensors or equipment other than camera to recognize static and dynamic hand gestures. First, a series of hand gesture input images are converted into high-frequency images, then each of the hand gestures RGB images and their high-frequency images is learned through the DenseNet three-dimensional Convolutional Neural Network. Experimental results on 6 static hand gestures and 9 dynamic hand gestures showed an average of 92.6% recognition rate and increased 4.6% compared to previous DenseNet. The 3D defense game was implemented to verify the results of our study, and an average speed of 30 ms of gesture recognition was found to be available as a real-time user interface for virtual reality applications.

1) 교신저자

키워드: 손 제스처 인식, 딥러닝, 합성곱 신경망, 텐스넷, 이중흐름 신경망

Keyword: Hand Gesture Recognition, Deep Learning, Convolutional Neural Network, DenseNet, Two-Stream Network

1. 서론

최근 가상현실 기술이 발전함에 따라 가상현실 상에서 몰입감을 증가시키기 위한 사용자 친화적 인터페이스(NUI, Natural User Interface)에 대한 연구가 활발히 진행되고 있다[1]. 가상현실 속의 인터페이스는 키보드나 마우스와 같은 단순한 입력 장치를 넘어 사용자에게 자연스럽고 직관적인 느낌을 줄 수 있어야 한다.

사용자에게 가장 자연스러운 인터페이스는 사람이 일상생활에서 사용하는 손 제스처를 가상공간 속에서도 인식할 수 있도록 하는 것이다. 이와 같은 손 제스처 인식에 대한 연구로 키넥트나 립모션과 같은 장비를 이용한 손 제스처 인터페이스에 관한 연구[2-7]가 있었지만 특화된 장비 혹은 센서가 필요하고 조명이나 거리와 같은 주변 환경에 제약을 받는다는 단점이 있었다.

인공지능과 고성능 GPU의 발달로 등장한 딥러닝(Deep Learning) 기술을 손 제스처 인식에 적용하는 연구가 최근 진행되고 있다. 이와 관련한 기존 연구로 스테레오 비디오로부터 구한 깊이 정보와 색 정보를 이용하여 검출한 손 윤곽선 정보를 학습시켜 인식하는 연구[8], 깊이 카메라로 얻은 관절 정보를 학습하여 인식하는 연구[9] 그리고 칼라 영상과 깊이 영상을 결합하여 3차원 합성곱 신경망(Convolutional Neural Network)으로 학습하여 인식하는 연구[10] 등이 있다. 하지만 위의 연구들 역시 특정 장비나 센서를 필요로 하거나, 낮은 인식률을 보이는 문제점을 지니고 있다.

최근에는 객체 인식을 위한 다양한 신경망 구조들이 등장했는데, 그 중 텐스넷(DenseNet) [11]과 이중 흐름(Two-Stream) [12] 구조를 예로 들 수 있다. 텐스넷은 압축 계층을 지닌 고밀도 연결 구조를 사용하여 신경망이 깊어질수록 정보를 손실하는

문제를 보완하였고 적은 파라미터로도 높은 속도와 인식률을 제공하는 장점을 가진다. 이중 흐름 구조는 입력 영상의 공간적 흐름에 따른 정적인 정보 외에 시간적 흐름에 따른 동적인 정보를 동시에 학습함으로써 성능을 높였다. 즉 입력 영상에 광학 흐름(Optical Flow) 처리를 하여 동적인 정보를 추출하고, 입력 영상과 광학흐름 처리 영상 각각에 대한 합성곱 신경망을 동시에 수행하고 그 결과를 융합함으로써 인식의 정확도를 높였다.

본 연구에서는 이중흐름 구조를 사용하되 손 제스처의 특성 상 시간 흐름 정보보다 각 타임구간의 손 포스처 영상이 인식에 많은 영향을 끼치는 점을 고려하여 손 영상의 고주파(High Frequency) 정보를 별도로 학습하여 기존 결과와 융합한다. 또한 합성곱 신경망 구조로 적은 파라미터와 높은 인식률을 제공하는 텐스넷을 사용한다. 본 논문에서 제안하는 손 제스처 인식 방법은 다음과 같다. 일반적인 USB 카메라로 최근 20 프레임의 손 제스처 영상을 입력 받아 이중 흐름 구조의 첫 번째 흐름의 입력으로 가공하지 않은 RGB 영상을, 두 번째 흐름의 입력으로 원본 영상에서 가우시안 블러(Gaussian Blur) 처리를 거친 저주파 영상을 뺀 고주파 영상을 사용한다. 각각의 흐름의 마지막 합성곱 연산을 수행한 후 RGB 흐름과 고주파 흐름에서 나온 특징맵(Feature Map)에 대하여 1차 융합(Fusion) [13]을 실시한다. 이후 RGB 흐름에서 출력된 특징맵과 융합된 특징맵에 대해 각각 전역 평균 풀링(Global Average Pooling) 연산을 수행한다. 그 후 최종적으로 출력된 결과에 대해 2차 융합을 실시하고, 소프트맥스(Softmax) 함수를 거쳐 인식 결과를 출력하게 된다. 본 방법의 검증을 위하여 조명, 배경, 영상의 손 위치와 거리 등 다양한 상황을 고려하여 제작한 15가지 정적 및 동적 손 제스

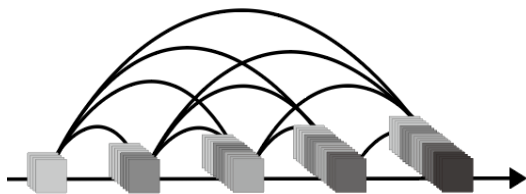
처 데이터 세트로 실험한 결과 기존 텐스넷에 비해 4.6%의 성능이 향상된 평균 92.6%의 인식률을 보였다.

본 논문의 구성은 다음과 같다. 2장에서 관련 연구로 텐스넷과 이중 흐름 신경망에 대해 소개한다. 3장에서 본 논문에서 제안하는 손 제스처 인식을 위한 텐스넷 기반 이중 흐름 신경망 구조에 대해 설명한다. 4장에서 본 방법에 대한 성능 분석 및 가상 현실 게임에 적용한 결과를 기술한 후 5장에서 결론을 맺는다.

2. 관련 연구

2.1 텐스넷

텐스넷은 CVPR 2017에서 최우수 논문으로 선정된 고밀도 연결 구조(Dense Connectivity) 네트워크이다. 고밀도 연결 구조는 (그림 1)에서 보는 바와 같이 각 계층의 출력을 이후의 모든 계층에 덧붙여 연결함으로써 신경망이 깊어질수록 초기 계층의 정보를 잃어버리는 단점을 보완하기 위한 구조이다. 또한 단순한 행렬 덧셈으로 계층 사이를 연결했던 기존의 합산(summation) 방식 대신 이전 계층의 특징맵을 결합해 기존 정보를 유지하는 결합(Concatenation) 방식을 제안하였다(그림 1).



(그림 1) 고밀도 연결 구조

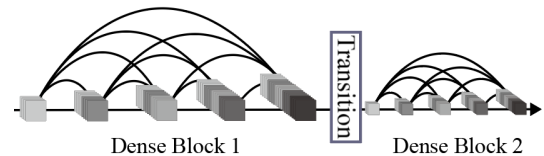
식 1에서 보는 바와 같이 고밀도 연결 구조의 각 계층의 입력 x_l 은 이전 계층의 출력 특징맵 $H_l(x_{l-1})$ 에 이전 계층이 입력 받은 특징맵을 결합하여 구성된다. 이전 계층의 입력 특징맵 x_{l-1} 은 $H_l(x_{l-1}) + H_{l-1}(x_{l-2}) + x_{l-2}$ 이기 때문에, 각

계층의 입력에는 그 이전 모든 계층의 출력들이 포함되어 있다(식 2). 이와 같이 마지막 계층은 첫 번째 계층의 출력 특징맵 $H_1(x_0)$ 부터 바로 전 단계의 출력 특징맵 $H_l(x_{l-1})$ 까지를 결합하여 입력으로 활용한다.

$$x_l = H_l(x_{l-1}) + x_{l-1} \quad (1)$$

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (2)$$

고밀도 연결구조 방식을 수행하기 위해서는 (그림 1)에서 보는 바와 같이 각 계층에 대한 특징맵의 해상도가 같아야 한다. 합성곱을 수행하기 위해서 다운 샘플링(Down Sampling)이 필수적이므로 (그림 2)와 같이 고밀도 연결이 수행되는 범위를 고밀도 구역(Dense Block)으로 구분하고 각 구역 사이의 이행 계층(Transition Layer)에서 3D 평균 풀링을 수행하여 해상도를 줄인다.

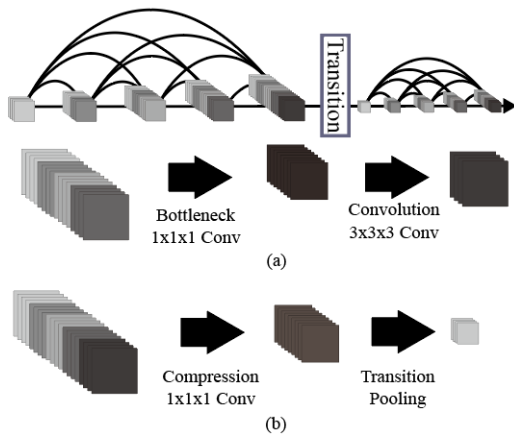


(그림 2) 고밀도 구역으로 나눈 모델 구조

고밀도 연결 구조는 신경망의 깊이가 깊어질수록 특징맵의 깊이가 늘어나기 때문에 이와 비례하여 파라미터 수가 커지게 된다. 이와 같은 문제를 해결하기 위해 각각의 합성곱 계층에서 사용하는 커널의 깊이를 기존 모델에 비해 작은 값을 사용함으로써 출력되는 특징맵의 깊이를 축소시켜 특징맵의 크기가 지나치게 커지지 않도록 한다. 또한 특징맵의 밀도(Compactness)를 높여 메모리를 절약하기 위한 병목 계층(Bottleneck Layer)과 압축 계층(Compression Layer)을 추가하여 특징맵의 크기를 조절한다.

병목 계층과 압축 계층은 NIN[14]이 제안한 1

$\times 1$ 합성곱 연산을 통해 크기를 축소시키고 특징맵이 과도하게 커지지 않게 조절한다. 병목 계층은 고밀도 구역 내 합성곱 계층의 이전 단계에서 입력 특징맵의 크기를 조정하고, 압축 계층은 고밀도 구역의 끝에서 이행 계층으로 넘어가는 특징맵의 크기를 조정하여 3D 합성곱 계층의 파라미터 크기를 줄인다 (그림 3).



(그림 3) 덴스넷의 병목 계층(a)과 압축 계층(b)

2.2 이중 흐름 합성곱 신경망

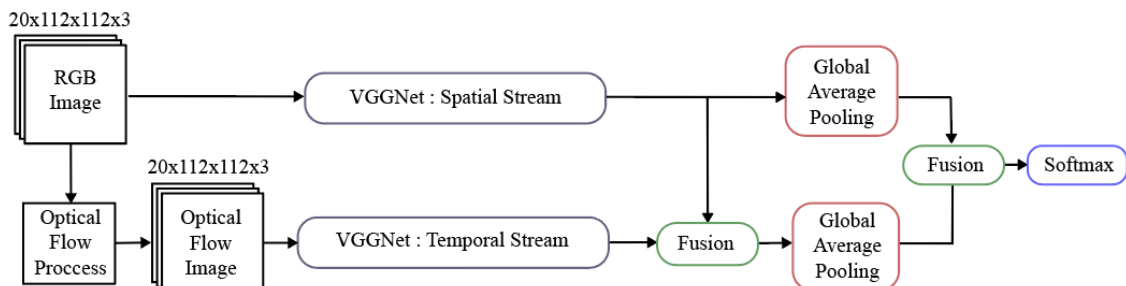
이중 흐름 구조는 서로 다른 두 개의 합성곱 신경망에 각각 공간 및 시간 정보를 입력 영상으로 사용함으로써 동적 영상 인식에 높은 성능을 제공하는 구조이다 (그림 4). 이중 흐름 구조에서 적용한 합성곱 신경망은 VGGNet [15]으로, 이 네트워크는 합성곱 층(Convolutional Layer) - 최대 풀링 계층(Max Pooling Layer) - 완전 연결 계층(Fully

Connected Layer) 순서로 이루어진 간단하고 사용하기 쉬운 특징을 가지고 있는 신경망이다.

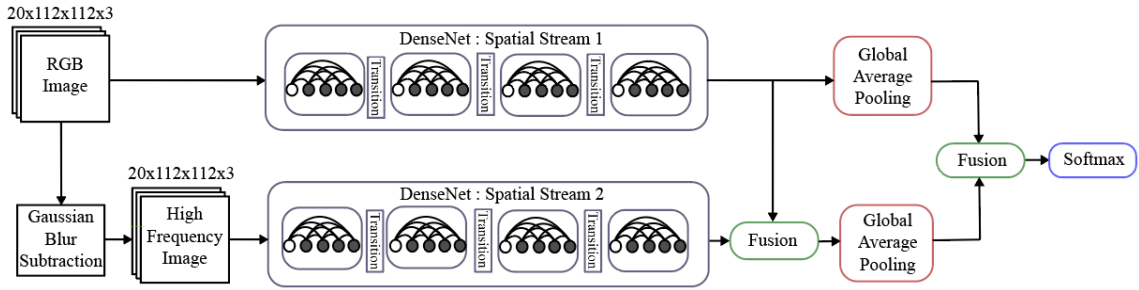
이중 흐름 구조의 공간 흐름 구조에는 RGB 영상을 입력 데이터로 사용하고, 시간 흐름 구조에는 RGB 영상에 광학 흐름(Optical Flow) 처리를 거친 영상을 입력 데이터로 사용한다. 광학 흐름 처리는 이전 프레임에서 특징 점을 추출한 후 현재 프레임에서 동일한 특징 점을 찾아 그 점 사이의 움직임의 변화를 시각적으로 볼 수 있는 처리 방식이다. 위와 같은 특징 때문에 광학 흐름 처리는 입력 데이터의 시간 흐름에 따라서 변화하는 특징을 가져온다. 학습 이후 공간 흐름 구조와 시간 흐름 구조 각각에서 출력된 특징맵에 대한 융합 과정이 이루어지며, 이후 첫 번째 전역 평균 풀링 계층에는 공간 흐름의 특징맵이 입력되고, 두 번째 전역 평균 풀링 계층에는 융합된 특징맵이 입력된다. 전역 평균 풀링을 거친 각각의 특징맵은 2차 융합을 거친 후 소프트맥스 함수를 통해 인식 결과를 출력한다.

3. 이중 흐름 신경망을 이용한 손 제스처 인식 방법

본 논문은 실시간 손 제스처 인식을 위한 덴스넷 기반 이중 흐름 신경망 구조(그림 5)를 제안한다. 기존의 광학 흐름 데이터와 VGGNet은 느린 처리 속도로 3D 합성곱 신경망에 적합하지 않고 실시간 인터페이스 구현에도 제약을 가진다. 이러한 단점을 개선하기 위해 광학 흐름 처리 대신 손의 상세 정보를 부가적으로 학습하여 성능을 높이기 위하여 상대적으로 처리 시간이 짧은 가우시안 블러를 활



(그림 4) 이중 흐름 신경망의 구조



(그림 5) 덴스넷 기반의 3차원 이중 흐름 신경망 구조그림 8

용한 고주파 손 영상을 입력 데이터로 사용하고, 신경망으로 VGGNet 대신 속도와 성능 양쪽에서 우위를 보이는 덴스넷을 사용하였다.

손 제스처 인식을 위해 일반적인 USB 카메라에서 초당 30 프레임의 속도로 사용자의 제스처를 촬영한 후, 최근 20 프레임의 영상으로 입력 데이터를 제작한다. 입력 데이터는 메모리 절약을 위해 112x112 크기로 재조정(Resize)을 거친 후 원본 RGB 영상과 고주파 영상으로 나누어지며, 고주파 영상은 원본 영상에서 가우시안 블러 처리를 거친 저주파 영상을 뺀으로서 제작된다(그림 6).



(그림 6) 원본 영상(좌)과 저주파 제거 영상(우)

신경망에 입력된 손 제스처 영상은 3x7x7 커널을 사용하여 합성곱을 한번 수행한 후 4개의 고밀도 구역을 거친다. 고밀도 블록 내부의 합성곱 계층은 모두 배치 정규화(Batch Normalization) [16]와 ReLU 활성화 함수[17]를 거친 후 수행한다. 각 고밀도 블록은 1x1x1 커널을 사용한 병목 계층과 3x3x3 커널을 사용한 합성곱 계층을 4번 반복한다. 각 프레임의 정보를 보존하기 위해, 이행계층에서 영상의 크기를 다운 샘플링 할 때에도 프레임의

크기는 유지한다.

4 개의 고밀도 구역을 거친 후, 1차 융합을 통해 두 개의 신경망에서 출력된 특징맵을 하나로 융합한다. 융합 기법은 두 개의 특징맵을 서로 합친 후 나누어주는 평균 융합 기법을 사용한다. 융합이 끝난 후 첫 번째 전역 평균 풀링 계층에는 RGB 흐름에서 출력된 특징맵이 그대로 입력되고, 두 번째 전역 평균 풀링 계층에는 융합된 특징맵이 입력된다. 이 때 과적합(Overfitting)에 대한 대책으로 드롭아웃(Dropout)을 수행한다. 위와 같은 단계를 거쳐 출력된 각각의 특징맵은 2차 평균 융합을 거친 후 소프트맥스 함수를 통해 인식 결과를 출력한다.






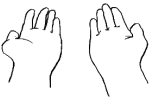


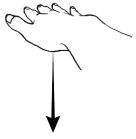

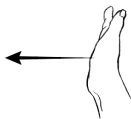
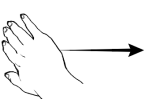
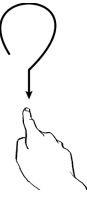
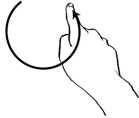

4. 실험

본 실험은 프로세서 Intel Core i5 8400, 그래픽 카드 GeForce GTX 1080Ti, RAM 16GB 등으로 구성된 장비와 Python 3.5 기반의 Tensorflow 1.5 GPU 버전을 개발도구로 사용한 환경에서 진행되었다.

4.1 손 제스처 정의

손 제스처는 시작부터 끝까지 손의 모양과 위치가 변하지 않는 정적 제스처와 시간에 따라 위치가 변하는 동적 제스처로 분류된다. 본 논문에서 제안한 방법을 실험하기 위하여 <표 1>에서 보는 바와 같이 6개의 정적 제스처와 9개의 동적 제스처를 정의하였다.

〈표 1〉 손 제스처 정의

정적 제스처		
Point	Grab	Spread Palm
		
Ok	No	Receive
		
동적 제스처		
Push	Pull	Swipe Down
		
Swipe UP	Swipe Left	Swipe Right
		
Question Mark	Circle	Triangle
		

4.2 데이터 세트

본 실험에 사용된 데이터 세트는 640x480 해상도의 카메라로 촬영되었고, 4명의 학습자가 참가하

여 <표 1>에서 정의한 손 제스처를 다양한 위치, 각도, 거리에서 시행하여 제작하였다. 초당 20 프레임으로 구성된 손 제스처 영상 1500 세트인 총 30,000장을 제작하였고, 과적합 문제를 개선하기 위해 제작된 손 제스처 영상의 손의 위치와 크기, 조명을 다양하게 후처리하여 5배 증가시켜 총 150,000장을 제작하였다. 총 150,000장 중 90%인 6,750 세트는 학습 데이터로, 나머지 10%인 750세트는 평가 데이터로 사용하였다.

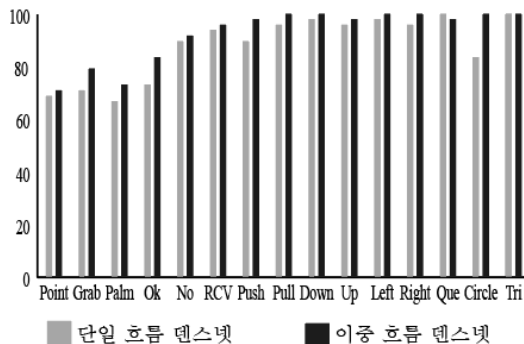
4.3 실험결과

본 논문에서 실험한 제스처의 종류는 정적 제스처와 동적 제스처이다. 정적 제스처는 한 장의 입력 영상을 가진 2D 합성곱 신경망으로 학습이 가능한 반면 동적 제스처는 일련의 입력 영상들로 이뤄진 3D 합성곱 신경망으로 학습이 가능하다. 제스처 인터페이스로 정적 제스처와 동적 제스처 두가지 모두 필요하므로 이 두 가지 제스처를 지원하고자 한다면 서로 다른 신경망을 번갈아 사용하여야하는 부담이 따른다. 본 논문은 정적 제스처와 동적 제스처를 시스템에 부담을 갖지 않도록 하면서 실시간으로 인식시키는 것을 목적으로 한다. 따라서 정적 제스처와 동적 제스처를 동시에 학습하는 통합모델을 제안하였고 실험은 배치 크기 8, Epoch 20, 드롭아웃 비율 0.5의 조건에서 진행되었다.

실험 결과는 <표 2>와 그림 7과 같다. 표에서 보는 바와 같이 상대적으로 유사한 정적 제스처인 Point, Grab, Spread Palm, OK 등의 경우 기타 제스처의 성능에 비해 인식률이 다소 떨어짐을 알 수 있다. 이점은 정적 제스처와 동적 제스처를 동일한 신경망에서 학습함으로써 인식률의 저하를 가져온 것으로 판단된다. 그렇지만 본 논문에서 제안하는 텐스넷 기반 이중 흐름 신경망은 단일 흐름 신경망과 비교하여 4.58%의 성능이 향상되어 전반적으로 92.5%의 인식률을 보였다.

〈표 2〉 단일 흐름 텐스넷과 이중 흐름 텐스넷의 제스처별 인식률

제스처	단일 흐름 텐스넷	이중 흐름 텐스넷
Point	68.75%	70.83%
Grab	70.83%	79.17%
Spread Palm	66.67%	72.92%
Ok	72.92%	83.33%
No	89.58%	91.67%
Receive	93.75%	95.83%
Push	89.58%	97.92%
Pull	95.83%	100.00%
Swipe Down	97.92%	100.00%
Swipe Up	95.83%	97.92%
Swipe Left	97.92%	100.00%
Swipe Right	95.83%	100.00%
Question Mark	100.00%	97.92%
Circle	83.33%	100.00%
Triangle	100.00%	100.00%
Total	87.92%	92.50%



(그림 7) 단일 흐름 텐스넷과 이중 흐름 텐스넷의 제스처별 인식률 비교

〈표 3〉은 단일 텐스넷을 사용했을 때의 파라미터수를 나타낸 것이다. 이중 흐름 텐스넷의 경우, 파라미터 수에 영향을 주는 합성곱 계층이 단일 텐스넷의 두 배만큼 사용되기 때문에 파라미터 수는 〈표 3〉의 두 배에 해당된다. 결과적으로 이중 흐름 텐스넷의 파라미터 수는 4.32M로 기존 연구[12]에서 사용한 VGGNet의 29.78M와 비교해 확연히 적은 수치를 보임을 알 수 있다.

〈표 3〉 단일 텐스넷의 파라미터 수

Type	Params
Convolution 1	28.22K
Convolution 2	118.78K
Convolution 3	122.88K
Convolution 4	126.98K
Convolution 5	131.07K
Transition 1	18.43K
Convolution 6	122.88K
Convolution 7	126.98K
Convolution 8	131.07K
Convolution 9	135.17K
Transition 2	25.09K
Convolution 10	124.93K
Convolution 11	129.02K
Convolution 12	133.12K
Convolution 13	137.22K
Transition 3	28.80K
Convolution 14	125.95K
Convolution 15	130.05K
Convolution 16	134.14K
Convolution 17	138.24K
Global Average Pooling	
Total	2,169.02K

〈표 4〉는 본 논문에서 제안한 방법과 기존 이중 흐름 신경망[12]의 손 제스처 인식 처리시간을 비교한 것이다. 본 방법은 속도의 제약이 컸던 기존 이중 흐름 신경망의 데이터 처리와 신경망 구조를 개선하여 약 8배의 성능 향상을 보여 실시간으로 제스처 인터페이스를 제공하는 것이 가능함을 알 수 있었다.

〈표 4〉 손 제스처 인식 처리시간

기존 연구[12]	본 방법
290 ms	34ms

4.4 응용 사례

본 논문의 텐스넷 기반 이중 흐름 신경망 학습 모델의 성능을 확인하기 위하여 가상현실 기반의 응용 프로그램을 제작하였다 (그림 8). 응용 프로그램은 다리 건너편에서 다가오는 몬스터들을 막는

디펜스 게임 프로그램으로, 주인공 캐릭터는 손 제스처 15가지를 인터페이스로 사용하여 마법 기술을 사용할 수 있다. 몬스터의 종류와 상황에 따른 다양한 마법 기술을 사용하기 위해 각 기술의 컨셉에 맞는 다양한 인터페이스를 구현하였다.

USB 카메라를 통해 사용자의 손 제스처를 초당 30 프레임 촬영하여 신경망 모델에 적용하여 실험한 결과 손 제스처를 평균 34 ms로 실시간 인식하여 손 제스처 기반 인터페이스로 가상현실 게임 실행이 가능함을 알 수 있었다.



(그림 8) 손 제스처를 활용한 가상현실 게임

5. 결론

본 논문은 USB 카메라 이외의 별도의 센서나 장비 없이 실시간으로 정적 손 제스처와 동적 손 제스처를 인식하기 위한 텐스넷 기반 이중 흐름 신경망 구조를 제안한 후 가상현실 게임 상에서 실시간으로 손 제스처 인터페이스로 적용 가능함을 보였다.

여전히 딥러닝을 이용한 응용은 고사양의 컴퓨팅 환경을 요구한다. 향후연구로 저사양 환경에서도 구동 가능한 딥러닝 기반 손 제스처 인식 인터페이스를 개발하고자 한다.

■ 감사의 글

본 연구는 2018학년도 서경대학교 교내연구비 지원에 의하여 이루어졌음.

■ 참고문헌

- [1] 박경범, 이재열, “가상현실 환경에서 3D 가상객체 조작을 위한 인터페이스와 인터랙션 비교 연구,” 한국CDE학회 논문집, 21(1), pp. 20-30, 2016. 3.
- [2] 윤종원, 민준기, 조상배, “몰입형 가상현실의 착용식 사용자 인터페이스를 위한 Mixture-of-Experts 기반 제스처 인식,” 한국HCI학회 논문지, 6(1), pp. 1-8, 2011. 5.
- [3] 나민영, 유휘중, 김태영, “스마트 디바이스 제어를 위한 비전 기반 실시간 손 포즈 및 제스처 인식방법,” 한국차세대컴퓨팅학회 논문지, 8(4), pp.27-34, 2012.8.
- [4] 이세봄, 정일홍, “키넥트를 사용한 NUI 설계 및 구현,” 한국디지털콘텐츠학회 논문지, 15(4), pp. 473-480, 2014. 8.
- [5] 고택균, 윤민호, 김태영, “HMM과 MCSVM 기반 손 제스처 인터페이스 연구,” 한국차세대컴퓨팅학회 논문지, 14(1), pp. 57-64, 2018. 2.
- [6] 김민재, 허정만, 김진형, 박소영, 장준호, “직관적인 손 동작을 고려한 림프선 기반 게임 인터페이스의 개발 및 평가,” 한국컴퓨터게임학회 논문지, 27(4), pp. 69-75, 2014. 12.
- [7] 김설호, 김경섭, 김계영, “ToF 깊이영상과 벡터내적을 이용한 손 모양 인식,” 한국차세대컴퓨팅학회 논문지 12(4), pp. 89-101, 2016.8.
- [8] 문현철, 양안나, 김재곤, “웨어러블 응용을 위한 CNN 기반 손 제스처 인식,” 방송공학회 논문지, 23(2), pp. 246-252, 2018. 3.
- [9] A. Sinha, C. Choi, and K. Ramani, “DeepHand: Robust hand pose estimation by completing a matrix imputed with deep features,” In IEEE Conference on Computer Vision and Pattern

Recognition, pp. 4150-4158, 2016.

- [10] P. Molchanov, S. Gupta, K. Kim, and J. Kauts
"Hand Gesture Recognition with 3D Convolutional
Neural Networks," In IEEE Conference on
Computer Vision and Pattern Recognition,
pp.1-7, 2015.
- [11] G. Huang, Z. Liu, K. Q. Weinberger, and L.
van der Maaten. "Densely connected convolutional
networks," In IEEE Conference on Computer
Vision and Pattern Recognition, pp 3-11,
2017.
- [12] K. Simonyan, A. Zisserman. "Two- stream
convolutional networks for action recognition
in videos," In NIPS, 2014.
- [13] C. Feichtenhofer, A. Pinz, A. Zisserman,
"Convolutional two-stream network fusion
for video action recognition," The IEEE
Conference on Computer Vision and Pattern
Recognition, pp. 1933-1941, 2016.
- [14] Min Lin, Qiang Chen, Shuicheng Yan, "Network
In Network," arXiv preprint arXiv:1312.4400v1,
2013.
- [15] K. Simonyan, A. Zisserman, "Very Deep
Convolutional Networks For Large-Scale Image
Recognition," In International Conference on
Machine Learning, pp. 1-14, 2014.
- [16] Sergey Ioffe, Christian Szegedy, "Batch
Normalization: Accelerating Deep Network
Training by Reducing Internal Covariate
Shift," arXiv preprint arXiv:1502.03167v3,
2015.
- [17] Vinod Nair, Geoffrey E. Hinton, "Rectified
Linear Units Improve Restricted Boltzmann
Machines," In International Conference on
Machine Learning, pp. 807-814, 2010.

■ 저자소개

◆ 최현중



- 2013년 3월~현재 서경대학교 컴퓨터
공학과 학사 재학
- 관심 분야: 가상현실, 게임프로그래밍,
컴퓨터 비전, 머신 러닝

◆ 노대철



- 2014년 3월~현재 서경대학교 컴퓨터
공학과 학사 재학
- 관심 분야: 가상현실, 게임프로그래밍,
컴퓨터 비전, 머신 러닝

◆ 김태영



- 1991년 2월 이화여자대학교 전자계산
학과 학사
- 1993년 2월 이화여자대학교 전자계산
학과 석사
- 1993년 3월~2002년 2월 한국통신
멀티미디어연구소 선임연구원
- 2001년 8월 서울대학교 전기컴퓨터 공
학부 박사
- 2002년 3월~현재 서경대학교 컴퓨터
공학과 부교수
- 관심 분야: 실시간 렌더링, 증강현실,
딥러닝, 영상처리, 모바일 3D