

논문 2022-17-28

# 웨어러블 센서를 활용한 경량 인공신경망 기반 손동작 인식기술

(A Light-weight ANN-based Hand Motion Recognition Using a Wearable Sensor)

이형규\*

(Hyung Gyu Lee)

**Abstract:** Motion recognition is very useful for implementing an intuitive HMI (Human-Machine Interface). In particular, hands are the body parts that can move most precisely with relatively small portion of energy. Thus hand motion has been used as an efficient communication interface with other persons or machines. In this paper, we design and implement a light-weight ANN (Artificial Neural Network)-based hand motion recognition using a state-of-the-art flex sensor. The proposed design consists of data collection from a wearable flex sensor, preprocessing filters, and a light-weight NN (Neural Network) classifier. For verifying the performance and functionality of the proposed design, we implement it on a low-end embedded device. Finally, our experiments and prototype implementation demonstrate that the accuracy of the proposed hand motion recognition achieves up to 98.7%.

**Keywords :** Hand Motion Recognition, Artificial Neural Network, Wearable Sensor, AIoT, Flex Sensor

## I. 서 론

IT 기술에 기반을 둔 인공지능 기술의 발달과 함께 사람의 동작이나 움직임을 인식하여 활용하는 다양한 응용들이 개발되고 있다. 특히 손은 다른 신체 부위와 비교할 때 상대적으로 적은 에너지로 가장 정확하게 움직일 수 부분으로, 인간-컴퓨터 상호작용 (Human-Computer Interaction, HCI)의 효율적인 인터페이스 수단으로 사용되고 있다 [1-3].

전통적으로 신체 움직임 인식에는 이미지 (영상) 분석 방법이 주로 사용되고 있다. 이미지 분석 방식은 사용자가 전자센서를 착용하지 않고도 적용 가능하다는 장점이 있다. 그러나 현재의 이미지 기반의 신체 움직임 인식기술은 여러 가지 단점이 존재한다. 첫째, 높은 인식 정확도를 얻기 위해서는 동작의 범위를 크게 해야 할 뿐 아니라 고해상도의 이미지를 얻을 수 있는 카메라를 사용해야 한다. 둘째, 인식의 정확도는 카메라와 물체와의 거리, 주변 다른 물체의 존재 여부, 물체의 밝기와 각도 등 다양한 환경에 영향을 받게 된다 [4]. 무엇보다도 이미지 기반 인식을 위해서는 비교적 많은 컴퓨팅/메모리 자원이 요구되기 때문에 상대적으로 적은 컴퓨팅/메모리 자원을 사용하는 임베디드 디바이스에서는 손가락을 포함하는 손동작의 정확한 인식에는 효율적이지 않다. 따라서 본 연구에서는 일반적인 방법인 이미지 기반처리 대신 비교적 간단한 플렉스 센서 (Flex Sensor) [5]

를 활용한 손동작 인식기술을 설계하고자 한다.

플렉스 센서는 밴드 같은 간단한 도구만으로 손에 부착 가능한 형태로, 움직임 인식을 위해 카메라나 레이저와 같은 장비를 설치할 필요가 없으면서도, 사용자의 물리적 위치, 환경, 자세에 영향을 받지 않으므로 손의 움직임을 관찰하기 유용한 센서 중 하나이다. 또한, 이미지 기반 동작 인식보다 상대적으로 적은 양의 데이터를 생산하기 때문에 임베디드 디바이스 기반의 헬스케어 [6], 수화 인식 [7], 자세 교정 [8]과 같은 웨어러블 IoT 응용 분야에서 활용되고 있다.

플렉스 센서는 손가락 구부림을 포함하는 움직임의 정도를 각도 (Degree) 정보로 변환해주기 때문에, 생성되는 각도의 정보를 분석하면 어떠한 형태로 손이 움직였는지를 판단할 수 있다. 최근 인공신경망 기술 (ANN, Artificial Neural Network)을 활용한 동작 인식에 관한 연구들이 이루어지고 있다. 움직임 인식의 정확도를 높이기 위해 플렉스 센서와 가속도 센서로부터 수집된 데이터에 ANN을 사용하여 사람의 일상적인 활동을 인식하는 연구가 이루어졌다 [9]. 이 연구의 경우 인공신경망 기술을 활용한 움직임 인식기술이 저사양의 웨어러블 장치에서도 높은 인식 정확도로 구현될 수 있음을 보여주고 있다. 본 논문에서는 하나의 센서로 2축 움직임을 측정할 수 있는 최신의 플렉스 센서 [5]와 경량의 ANN기술을 이용하여 저사양의 임베디드 디바이스에서도 구현이 가능한 손 움직임 감지 기술을 제안하고 구현한다.

제안된 손동작 인식기술은 데이터 수집부, 전처리 필터, 인공신경망 기반 움직임 인식 분류기로 구성되어 있다. 데이터 수집부는 손가락 부위에 부착된 한 개의 플렉스 센서로 부터 주기적으로 데이터를 수집하는 기능을 한다. 전처-

\*Corresponding Author (hglee@duksung.ac.kr)

Received: Mar. 21, 2022, Revised: Apr. 22, 2022, Accepted: May 23, 2022.  
H.G. Lee: Duksung Women's University (Assoc. Prof.)

※ 본 연구는 2021년도 덕성여자대학교 교내연구비 지원에 의해 이루어졌다.

리 필터부는 수집된 데이터를 움직임 인식 분류기로 보내기 전 센싱 잡음을 줄이기 위한 잡음 제거, 하나의 손동작 인식을 위한 데이터 크기를 조절하는 세그멘테이션/다운 샘플링, 사용자 간/동작 간 편차를 줄이기 위한 정규화 필터로 구성된다. 마지막으로 동작 분류기는 인공신경망 모델을 사용하여 설계되며, 입력/출력 계층과 함께 사용되는 은닉 계층 (Hidden Layer)의 수, 사용되는 노드 (뉴런, Neuron)의 수에 대한 설계공간 탐색을 통해 최소의 컴퓨팅/메모리 자원으로 최대의 정확도를 확보할 수 있는 경량 인공신경망 모델을 설계한다.

제안된 손동작 인식기술의 기능 및 성능을 검증하기 위해 저사양의 임베디드 시스템용 마이크로 컨트롤러를 사용하여 프로토타입을 구현하였으며, 제안된 설계방법은 10가지의 손동작에 대해 최대 98.7%의 정확도로 실시간 인식이 가능함을 보인다.

본 논문의 구성은 II 장에서는 손동작 인식 시스템에 사용되는 웨어러블 플렉스 센서의 특성 및 동작, 그리고 관련 연구를 소개한다. III 장에서는 제안되는 손동작 인식기술 및 설계방법을 소개하며, IV 장에서는 손동작 인식 프로토 타입 설계 및 성능평가를 진행한다. 마지막으로 V 장에서는 결론으로 논문을 마무리한다.

## II. 웨어러블 플렉스 센서 및 관련 연구

### 1. 웨어러블 플렉스 센서 기본 및 특성 분석

특정 물체의 구부러짐의 정도를 측정할 수 있는 센서를 플렉스 센서 또는 휠 센서라고 한다. 플렉스 센서는 구성 물질 또는 방법에 따라 전도성 잉크 (Conductive Ink), 광섬유 (Fibre Optic), 전도성 섬유/폴리머 (Conductive Fabric/Polymer) 등으로 구분될 수 있다. 센서들의 동작 원리는 플렉스 센서가 휘어짐 또는 구부러짐 등을 나타낼 때 센서에 사용되는 물질의 전기적 특성이 달라지는 현상을 이용한다. 이러한 플렉스 센서는 종류에 따라 휘어지는 최대 각도, 내구성, 측정되는 값의 안정성 등이 다르게 나타나게 된다. 예를 들어 전도성 잉크 기반 플렉스 센서의 경우 저비용의 장점으로 많이 사용되고 있지만, 정확도가 다소 낮고, 신축성 없이 길이가 고정되는 제약뿐 아니라 온도 및 습도와 같은 측정 환경에 따라 값이 조금씩 달라지기 때문에 보정 또는 필터링의 과정이 필요하다. 광섬유 기반의 경우 정확한 구부림 각도 측정이 가능하지만, 빛을 사용해야 한다는 단점과 비용이 높다는 단점이 있다. 전도성 섬유/폴리머의 경우 다른 방법에 비해 센서의 신축성으로 웨어러블 응용에 활용할 수 있지만, 고비용이라는 단점이 있다. 무엇보다도 기존 대부분의 플렉스 센서는 한 방향 (1축)의 구부러진 각도만 측정할 수 있어서 복합적인 움직임을 측정하기 위해서는 여러 개의 플렉스 센서를 사용해야 하는 단점이 존재한다.

최근 전도성-비전도성 소재가 계층적으로 첨가된 실리콘 탄성중합체 (silicon elastomer) 기반의 플렉스 센서가 개발되었다 [5]. 이 센서는 구부러진 정도를 값의 혼들림 없이



a) 2-Axis Flex Sensor

(b) Attached on the Index Finger of the Right Hand

그림 1. 손동작 인식에 사용되는 2축 플렉스 센서  
Fig. 1. 2-Axis Flex Sensor Used for Recognizing Hand Motion.

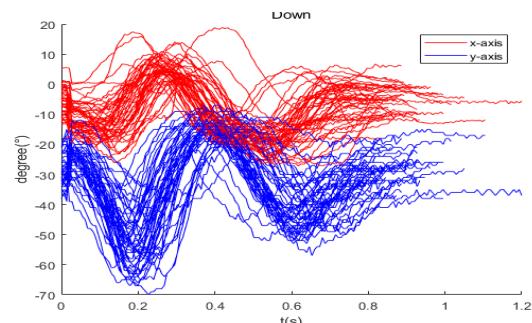


그림 2. 플렉스 센서에서 측정되는 값 예시 (센서 동작 주파수: 100Hz)

Fig. 2. Example of Measured Raw Data from a 2-Axis Flex Sensor (Sensor Operating Frequency: 100Hz)

안정적으로 측정할 수 있을 뿐 아니라, 실리콘 소재로 신축성이 있으며, 하나의 센서로 2방향 (X축, Y축)에 대한 구부러진 각도 단위로 측정이 가능한 장점이 있다. 이에 본 연구에서는 이 플렉스 센서를 사용한다 (그림 1 (a) 참조).

이 플렉스 센서는 0.18도 단위로 구부림 각도 측정이 가능하며, 전력 소모도 78uA로 매우 낮은 편에 속하며, 저전력 대기모드의 경우 1.7uA까지 낮출 수 있다. 측정된 구부러짐 정도는 아날로그값이 아닌 디지털 값으로 I<sup>2</sup>C 인터페이스를 통해 전송하기 때문에 별도의 ADC (Analog to Digital Converter)가 필요하지 않은 장점이 있다. 본 연구에서는 하나의 플렉스 센서만을 이용한 손동작 인식을 위해 그림 1 (b)와 같이 움직임의 자유도가 비교적 높은 집게손가락 상에 위 센서를 부착하여 사용하는 것을 기준으로 손동작 인식 시스템을 설계한다.

그림 2는 오른쪽 집게손가락에 2축 플렉스 센서를 부착한 상태에서 특정 움직임들을 통해 얻어지는 시간에 따른 플렉스 센서의 데이터 값 변화를 나타낸다. 기준점에서 시작하여 손가락을 위에서 아래로 내리는 동작에 대해 X축 및 Y축에서 측정되는 센서 데이터를 직접 표현한 그래프로서, 같은 동작을 반복하여 수행한 후의 결과를 정리한 그림이다. 측정되는 값은 손가락 움직임에 따른 각도 변화를 나타내며, 반복되는 동작마다 측정되는 값에 대한 편차는 존재하지만 특정한 형태의 패턴의 정보가 생성됨을 알 수 있다. 이러한 패턴은 동작의 종류에 따라 다르게 나타나며, 사용자에 따라서도 어느 정도 차이를 보이므로 이를 고려한 손동작 인식기가 설계되어야 한다.

## 2. 관련 연구

손동작 인식을 위한 다양한 연구가 수행되었다. 초기 연구는 손동작 인식보다는 다양한 센서를 이용한 손동작을 모니터링 하는 기술로, 손의 움직임을 측정할 수 있는 센서가 포함된 스마트 장갑을 착용한 손동작 인식 연구가 진행되었다 [10]. 이 연구에서는 다수의 광학 센서와 기반 플렉스 센서와 차이로 센서를 혼합하여 사용하였으며, 손가락의 움직임을 모니터링 할 뿐 동작 인식기술은 포함하고 있지 않다.

두 대의 카메라를 사용하는 스테레오 카메라 기반 손동작 인식 시스템이 제안되었다 [2]. 이 시스템은 두 대의 카메라를 사용하는 키넥트 (Kinect) 카메라를 이용하여 손 부위를 인식하고, 손의 각 부분의 거리를 측정한 후 손동작을 인식하는 방식을 사용한다. 이러한 방법은 다양한 손동작을 통해 컴퓨터 또는 기계를 제어할 수 있는 HCI 인터페이스를 구현한 것에 의의를 둘 수 있지만, 인식을 위한 카메라들이 필요하며, 수많은 신호처리 과정을 통해 손의 각 부분이 움직인 상대적 거리를 측정하는 방식으로 동작하기 때문에 계산을 위한 많은 컴퓨팅 자원뿐 아니라 인식의 정확도가 높지 않은 단점이 있다.

관성 측정 장치 (IMU, Inertial Measurement Unit)도 손동작 인식에 사용될 수 있다 [11]. IMU 센서에서 생성되는 데이터는 시계열 데이터로서 기존 방법에서는 DTW (Dynamic Time Wrapping) 기술을 사용하여 분석하였다. DTW 기술의 경우 복잡한 학습 환경으로 인해 실시간 학습/추론에 사용하지 못하는 단점을 극복하기 위해, [11]에서는 RCE (Restricted Column Energy) 신경망 기반의 인식기술을 제안하였다. 제안된 방법은 기존 방법과 비교하여 학습의 복잡도를 개선하면서도 높은 정확도를 나타냈다. 하지만 IMU 센서 자체의 샘플링 속도가 높아야 할 뿐 아니라, 잡음이 많은 센서의 특성상 복잡한 인공신경망 모델을 사용해야 한다는 단점이 여전히 존재한다.

손동작 인식을 위해 근전도 (EMG, Electromyography) 센서를 이용한 연구가 진행되었다 [12]. 이 연구에서는 EMG 센서로 수집된 생체신호를 기반으로 움직임 정보를 경량 인공신경망 모델에 적용하였으며, 중립 (Neutral) 상태를 포함한 7개의 동작을 인식에 높은 정확도를 보였지만, 사용된 네트워크 모델의 파라미터 수가 최소 수만 개로 매우 복잡하여 IMU 센서를 사용한 연구와 같이 저사양 임베디드 시스템에는 적용하기 어려운 단점이 있다.

다른 센서에 비해 비교적 제어가 간단하고, 안정적인 플렉스 센서도 손동작 움직임 인식에 많이 사용된다. 특히 전도성 잉크 기반 플렉스 센서를 사용한 다양한 손동작 움직임 인식 연구가 진행되었다 [13-16]. 전도성 잉크 기반 플렉스 센서는 특성상 1축의 휨 값만을 측정할 수 있기 때문에 복잡한 손동작 인식을 위해서는 다수의 플렉스 센서를 손가락 별 또는 손가락의 매 관절 별로 부착한 후 수집된 데이터를 기반으로 손동작을 인식하기 위해 다양한 기계학습 (Machine Learning) 방법들을 사용하였다. 그러나 대부분의 연구는 플렉스 센서로부터 수집된 손동작 움직임 자료들을

기반으로 고사양의 컴퓨터에서 학습하고 평가하여 달성한 인식 정확도만을 분석할 뿐, 제안된 동작 인식 모델이 임베디드 디바이스에서 실시간 동작 인식에 사용될 수 있음을 밝히고 있지 않다 [14], 일부 연구의 경우 아두이노 같은 저사양 디바이스는 데이터 수집 용도로만 사용될 뿐 실시간 동작 인식에는 라즈베리파이와 같은 비교적 높은 사양의 임베디드 디바이스를 사용한다 [15]. 딥러닝 기법을 사용하여 손동작을 인식할 수 있는 플렉스 센서를 사용한 스마트 글로브에 관한 연구가 진행되었다 [16]. 그러나 이 연구의 경우 장갑에 부착된 28개의 센서에서 생성되는 63개 채널의 신호를 분석하여 다양하고 정밀한 손동작 인식을 목표로 하는 연구로서 임베디드 디바이스에 적용하기에는 복잡한 모델을 사용하고 있다.

이에 본 논문에서는 카메라나, IMU, EMG, 다수의 플렉스 센서를 사용하는 기존 연구 방법들이 저사양의 임베디드 디바이스에 적용이 어려운 단점을 극복하기 위해 비교적 낮은 샘플링 속도에서 저전력으로 손동작 움직임을 관찰할 수 있는 최신의 단일 플렉스 센서 [5]를 활용하여 계산량을 최소화하면서도 높은 정확도를 갖는 인공신경망 기반 손동작 인식방법을 제안하고 설계한다.

## III. 손동작 인식 시스템 설계

### 1. 손동작 인식 시스템의 구조

그림 3은 본 논문에서 제안하는 단일 플렉스 센서를 활용한 손동작 인식 시스템의 구조를 나타낸다. 제안된 시스템은 데이터 수집 (Data Collection), 전처리 (Preprocessing) 필터, 분류 (Classification)의 3단계로 구성된다.

손동작 인식을 위한 첫 단계는 2축 플렉스 센서로부터 생성되는 움직임 데이터를 수집하는 단계이다. 집게손가락에 부착된 플렉스 센서로부터 집게손가락의 X축, Y축의 움직임 (구부림 정도)을 각각 16비트씩, 총 32비트의 각도 정보를 생성한다. 움직임의 속도에 따라 적절한 샘플링 속도를 조정할 수 있으며, 사용된 플렉스 센서의 경우 최대 200Hz의 샘플링 속도로 동작할 수 있다. 따라서 하나의 손동작 당 생성되는 데이터의 크기 (샘플의 개수)는 하나의 동작을 실행하는데 소요되는 시간 ( $t$ ), 플렉스 센서의 샘플링 속도 ( $f_s$ )에 따라 달라진다. 동작 인식의 정확도를 높이기 위해서는  $f_s$ 를 높게 설정할수록 좋지만, 이로 인한 계산량이 증가할

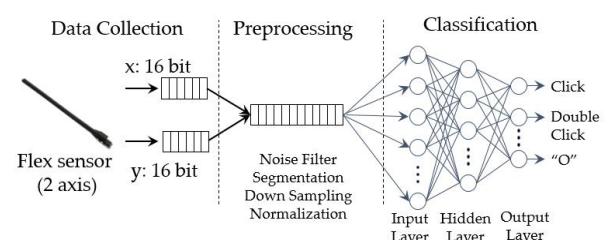


그림 3. 손동작 인식 시스템의 구조  
Fig. 3. System Architecture of the Proposed Hand Motion Recognition

수 있으므로 인식의 정확도를 고려하면서도 데이터의 크기 를 줄일 수 있는 적절한 샘플링 속도 설정이 필요하다.

플렉스 센서로부터 수집된 원본 데이터 (Raw Data)는 바로 동작 인식 분류기의 입력으로 사용할 수 있다. 하지만 수집된 원본 데이터의 경우 다양한 측정 잡음을 포함하고 있을 뿐 아니라, 인식되는 동작에 따라, 같은 동작이라 하더라도 사용자에 따라, 같은 동작/같은 사용자라 하더라도 시간에 따라 적지 않은 편차가 존재하게 된다. 또한, 동작 인식을 위해 분류기로 보내지는 데이터의 경우 다른 동작과 관련된 데이터가 혼합되지 않은 하나의 동작 인식과 관련된 데이터만으로 이루어져야 한다. 이를 위해 제안된 시스템에서 수집된 데이터의 전처리 과정을 포함하며, 전처리 필터 설계에 관한 내용은 다음 장에서 더 자세히 설명한다.

마지막으로 전처리 필터를 통해 정리된 데이터들은 동작 인식 분류기로 보내져 정해진 동작 중 하나로 분류된다. 본 연구의 주된 목적은 제한된 컴퓨팅 환경 (저사양의 임베디드 시스템)에서도 구현 가능한 높은 정확도를 갖는 인식 시스템의 설계 및 구현이다. 이를 위해 본 논문에서는 초경량 인공신경망 기반 분류기를 설계한다. 인공신경망 모델 설계 시 계층의 수, 계층별 노드의 수를 통한 최적화 과정이 필요하며, 이에 대한 설명 역시 다음 장에서 자세히 설명한다.

## 2. 전처리 필터 설계

본 논문에서 제안하는 전처리 단계는 그림 4에서 표현하는 바와 같이 잡음 제거 필터 (Noise Filter), 세그멘테이션 (Segmentation)과 다운 샘플링 (Down Sampling) 필터, 그리고 데이터 정규화 (Normalization) 필터로 구분된다.

### 2.1 잡음 제거 필터

플렉스 센서에서 수집되는 데이터는 인식하고자 하는 각 동작 별로 고유한 패턴을 가진다. 그러나 데이터 수집 과정에서 센서 자체, 센서-제어기 연결 회로, 센싱된 데이터를 디지털로 변환하는 과정 등 다양한 경로에서 잡음을 포함하게 된다. 이렇게 발생한 잡음들은 무작위로, 일정하지 않은 패턴으로 발생하게 되어 동작 별 고유한 데이터의 패턴을 교란하는 역할을 한다. 이러한 잡음은 곧 인식 정확도의 저하 또는 정확한 인식을 위해 사용해야 하는 컴퓨팅 자원의 수를 증가시키게 된다. 이러한 현상을 최소화하기 위한 전처리의 첫 단계로 잡음 제거 필터를 사용한다. 본 설계에서는 상대적으로 적은 계산량을 요구하는 잡음 제거 필터인 IIR (Infinite Impulse Response) 필터를 사용한다.

### 2.2 데이터 세그멘테이션과 다운 샘플링

하나의 동작에서 수집되는 데이터의 크기 (수)는 플렉스 센서의 동작 속도뿐 아니라 그림 2와 같이 동작의 종류에 따라, 사용자에 따라, 같은 사용자라 하더라도 동작의 시기에 따라 가변적이다. 하나의 동작 인식을 위해 처리하는 데이터 수가 가변적일 경우 인공신경망 모델을 설계하기 어렵게 되고, 처리하는 데이터의 수가 필요 이상 많아질 경우 복잡하고 많은 계산을 요구하는 신경망 모델을 설계해야만 한다. 이러한 점을 고려하여 동작 단위로 데이터를 수집할

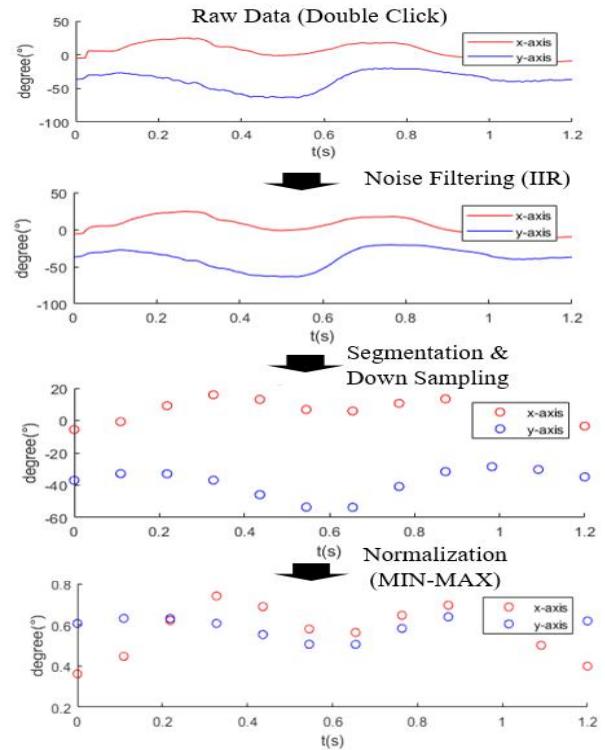


그림 4. 플렉스 센서로부터 수집된 데이터의 전처리 과정  
Fig. 4. Preprocessing of the Data Collected from the Flex Sensor

수 있는 세그멘테이션 필터를 설계해야 한다. 본 연구에서 손가락은 의도된 패턴의 동작을 수행할 때만 움직이고 다른 경우에는 멈추어 있게 되므로 수집되는 데이터 값의 평균 변화율과 확인하면 움직임의 시작 및 끝을 알 수 있다.

실험을 통해 손가락을 움직이지 않을 경우 플렉스 센서에서 측정되는 x축, y축의 평균 변화율은 대부분  $1^\circ$  (20ms 측정시간 기준) 이하로 관측되었으며, 잡음 및 편차를 고려하여 x축과 y축의 평균 변화율의 합이  $3^\circ/20\text{ms}$  미만이면 움직임이 없는 것으로 판단될 수 있다. 따라서 평균 변화율 합이  $3^\circ/20\text{ms}$  이상 관측되는 경우 움직임의 시작을 이하일 경우 움직임의 끝으로 판단된다. 다만 손의 움직임에 따른 측정 각도가 움직임 도중 반대 방향으로 변하는 움직임의 경우 순간적으로는 정지상태로 인정될 수 있으므로 정지상태의 기준을 평균 변화율과 지속시간 (본 연구에서는 100ms 이상 평균 변화율이  $3^\circ$ 이하로 유지될 경우)을 기준으로 정지상태를 판단한다. 평균 변화율의 경우 잡음 제거 필터와 같이 계산하면 이로 인한 추가적인 지연은 없게 된다.

세그멘테이션 필터에 의해 하나의 동작과 관련된 데이터들만으로 그룹을 생성한 후, 그룹으로 분류된 데이터들은 정확도의 손실이 크게 일어나지 않는 범위 내에서 다운 샘플링, 평탄화 (Smoothing) 알고리즘을 통해 고정된 개수의 입력 데이터를 생성한다. 손가락 움직임의 특성상 세그멘테이션과 다운 샘플링은 같은 단계에 수행하여 계산을 위한 실행 시간을 줄인다. 어느 수준까지 다운 샘플링 하는지에 대한 선택은 인공신경망 모델의 입력 계층 설계와 연관된

문제로써, 이에 대해서는 논의는 입/출력 계층 설계 부분에서 설명한다.

### 2.3 데이터값의 정규화

앞 장에서 설명한 것과 같이 하나의 동작에서 수집되는 데이터는 데이터의 수뿐 아니라 움직임의 각도 정보에서도 많은 편차가 존재한다. 특히 정의된 동작을 수행하는 시점에서 사용자별 손가락의 길이에 따라 다르게 나타난다. 또한, 같은 사용자, 같은 동작이라 하더라도 대기 상태에서의 손가락 각도, 움직임의 크기 등이 달라지고, 이는 곧 수집되는 데이터에 많은 편자가 발생하게 되는 원인이 된다. 위와 같은 이유로 발생할 수 있는 편차들을 줄이기 위해 수집된 데이터의 최대-최소값 (MIN-MAX)을 기준으로 정규화 과정을 진행하여 동작 인식에 사용되는 분류기의 복잡도를 최소화하면서도 동작 인식의 정확도를 향상한다.

### 3. 손동작 인식에 적합한 경량 인공신경망 모델 설계

인공신경망 기반 움직임 인식 모델 설계에 있어서 중요한 것은 적절한 입력 계층 (Input Layer), 은닉 계층 (Hidden Layer), 출력 계층 (Output Layer)에 대한 설계이다. 입력 계층의 경우 하나의 동작 인식에 사용되는 입력 데이터 수를, 출력 계층의 경우 인식되는 동작의 수를 고려할 경우, 설계를 위한 선택의 폭이 그리 넓지는 않다. 반면 은닉 계층의 경우 사용되는 모델의 종류, 학습을 위해 수집된 데이터의 양에 따라 다양한 선택을 할 수 있다. 본 장에서는 움직임 인식의 특성과 플렉스 센서로부터 생성되는 데이터의 특성에 따른 입출력계층 설계와 인식의 정확도를 높이면서도 저사양의 임베디드 디바이스 특성을 고려하여 자원사용을 최소화하기 위한 은닉 계층을 위한 설계공간 탐색에 대해 논의한다.

#### 3.1 입력/출력 계층 설계

입력 계층에 사용되는 노드의 수는 하나의 동작 인식에 사용되는 입력 데이터 수에 의존적이다. 앞장에서 설명한 내용과 같이, 동작 인식의 특성상 하나의 동작 인식을 위해 수집되는 데이터의 수는 동작의 종류, 사용자에 따라 다르게 나타난다. 이 경우 시계열 데이터 분석에 적합한 RNN (Recursive Neural Network) 모델을 사용하지만, 본 연구의 목표는 저사양 임베디드 디바이스에 구현 가능한 손동작 인식기술로, 초경량 인공신경망 모델 구현을 위해 본 논문에서는 비교적 단순하면서도 높은 정확도를 나타낼 수 있는 FNN (Feedforward Neural Network) 모델을 사용한다. FNN의 경우 특성상 입력 데이터의 수가 고정되어야 한다. 이를 위해 전처리 단계의 데이터 세그멘테이션 필터는 정보 손실을 최소화하는 범위 내에서 하나의 동작 당 입력 데이터의 수를 고정하는 역할을 한다. 인식을 위한 최적의 입력 데이터의 개수 선정을 위해 플렉스 센서의 동작 속도는 100Hz로 고정한 후 필터링 과정을 거친 출력 데이터 개수를 6에서 200까지 변경하여 실험한 결과 손동작의 경우 움직임이 빠르지 않은 특성상 그림 5와 같이 하나의 동작 당 데이터 수가 24개부터 성능 향상이 정체되는 결과를 얻었기

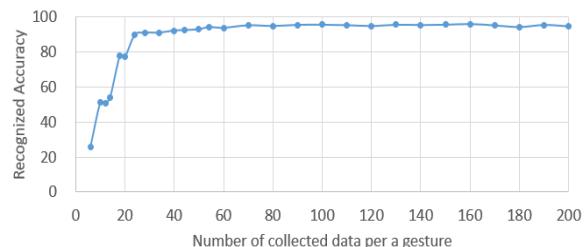


그림 5. 하나의 동작 당 수집되는 데이터 수에 따른 인식 성능  
Fig. 5. Accuracy Changes by the Number of Collected Data per a Gesture

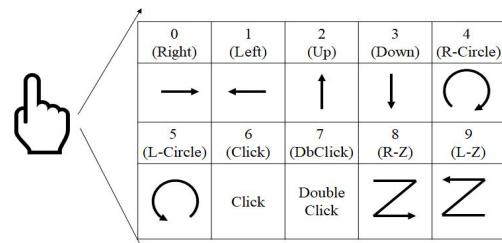


그림 6. 인식되는 10가지의 손동작  
Fig. 6. Ten Types of Recognized Hand Motions

에 입력되는 데이터의 수는 24개로 설정하여 설계한다.

출력 계층의 경우 사용되는 노드의 수는 인식되는 동작의 수로 결정되기 때문에 본 연구에서 최종인식 대상 손동작 수는 그림 6에 표시된 것처럼 10가지 동작으로 이를 반영하여 출력 노드의 수를 10개로 설정한다. 출력 계층의 각 노드에서는 인식되는 동작에 대한 확률값을 생성하기 위해 소프트 맥스 (Softmax) 활성화 함수를 사용하며, 가장 높은 확률을 갖는 손동작이 최종 결과로서 출력되는 방식으로 설계한다.

#### 3.2 은닉 계층 설계

입력과 출력 계층의 노드 수는 설계 대상에 따라 결정되는 반면, 은닉 계층 설계의 주요 요소인 은닉 계층 수 및 계층별 노드의 수는 컴퓨팅/메모리 자원 요구량과 정확도에 대한 분석이 필요하므로 은닉 계층의 경우 다양한 설계공간 탐색을 통해 인식 정확도를 높이면서도 메모리와 계산량을 최소화할 수 있는 설정을 찾아내야 한다.

본 연구의 경우 플렉스 센서에서 생성되는 데이터의 양은 카메라로부터 생성되는 이미지/영상 처리보다 상당히 작은 수의 입력 데이터를 사용하며, 인식되는 동작의 수도 10가

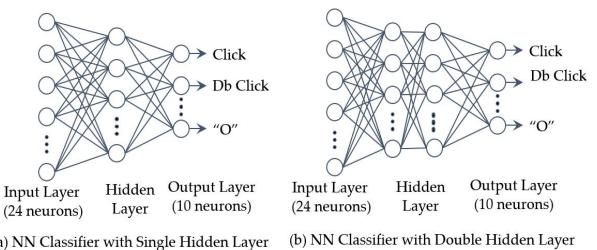


그림 7 설계공간 탐색을 위한 인공신경망 모델 구조  
Fig. 7 Neural Network Models Used in the Design Space Exploration

지로 비교적 복잡하지 않은 응용이기 때문에 사용되는 은닉 계층의 수는 그림 7과 같이 단일 은닉 계층 (Single Hidden Layer)과 이중 은닉 계층 (Double Hidden Layer)으로 한정하고, 이에 대한 설계공간 탐색을 수행한다.

이렇게 설정된 단일 은닉 계층 모델과 이중 은닉 계층 모델에 대해 각 계층 당 사용되는 노드의 수를 변경하며 설계공간 탐색을 진행한다. 설계공간 탐색을 위해 설정된 다양한 모델들에 대한 학습을 각각 진행하며, 학습 시 사용한 하이퍼 파라미터의 경우 학습률 (Learning Rate), 반복 수 (Epoch), 배치 크기 (Batch size)는 각각 0.0075, 500, 1로 설정되었으며, 활성화 함수로는 ReLu 함수를 적용한다. 설계공간 탐색의 결과는 IV장의 성능평가에서 자세히 설명한다.

#### IV. 프로토타입 구현 및 성능평가

##### 1. 프로토타입 구현 및 실험 환경

그림 8은 본 연구를 위해 구현된 프로토타입이다. 프로토타입은 동작 인식 제어 하드웨어와 집계손가락에 부착되는 단일 플렉스 센서로 구성되어 있다. 제어용 하드웨어의 경우 블루투스 통신 기능이 내장된 Cortex-M4 기반의 마이크로 컨트롤러를 사용하여 직접 제작하였으며, 블루투스 통신을 위한 안테나까지 PCB (Printed Circuit Board) 상에 바로 구현하여 손동작 인식 결과를 블루투스를 통해 컴퓨터나 스마트폰으로 전송할 수 있도록 구현하였다. 플렉스 센서로는 Bendlabs사의 2축 플렉스 센서 하나를 사용하여 구현하였다. 플렉스 센서는 제어 하드웨어에 구현된 전용 커넥터로 연결되며 I2C 채널을 통하여 통신한다. 제어 하드웨어의 경우 블루투스를 통해 PC 또는 스마트폰과 연결된다. 프로토타입 구현에 사용한 마이크로 컨트롤러와 플렉스 센서의 상세한 특성은 표 1의 내용과 같다.

제작된 프로토타입은 데이터수집 모드와 실시간 동작 인식을 통한 성능검증 모드의 두 가지의 모드로 사용된다. 데

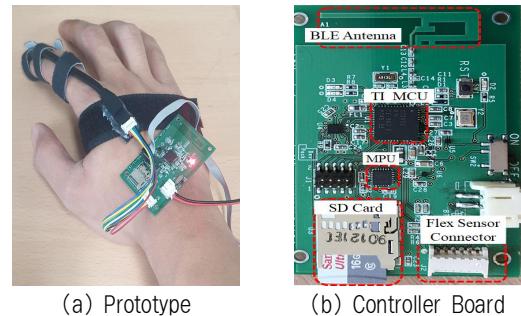


그림 8. 구현된 프로토타입  
Fig. 8. Prototype Implementation

표 1. 프로토타입 구현에 사용된 마이크로 컨트롤러, 플렉스 센서의 사양  
Table 1. Specifications of Micro controller and Flex Sensor Used in the Prototype

Micro-controller [17]	TI CC2652R Simplink Cortex-M4F 32-bit RISC (Max. 48MHz) 88KB SRAM, 352KB Flash Bluetooth 5.2, Zigbee 3.0 Power consumption: 71uA/MHz
Flex Sensor [5]	Bendlabs 2-Axis Sensor Max Sampling Frequency: 200Hz Power Consumption: 223uA@100Hz Measurement Resolution: 0.18° Dimensions: 100mm X 4mm X 4mm I <sup>2</sup> C: 16-bit data for each of X and Y axis

이터수집 모드의 경우 제어 하드웨어는 플렉스 센서로부터 수집된 원본 데이터들을 바로 PC로 전송하게 되고, 이렇게 수집된 원본 데이터들은 제안된 인공신경망 모델의 학습에 사용되게 된다. 그림 9은 인식되는 10가지의 손동작에 따른 수집된 손가락 움직임 데이터의 특징을 나타낸다. 본 연구에서는 손동작 인식을 위해 5명의 사용자 (남성 4명, 여성 1

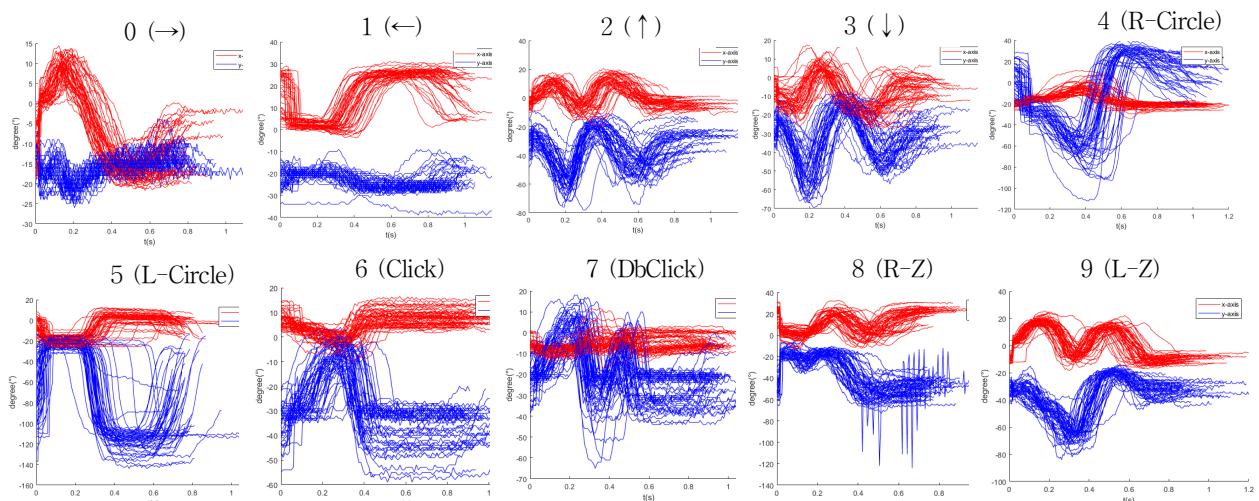


그림 9. 인식 동작에 따른 손가락 움직임 데이터의 특징

Fig. 9. Characteristics of Hand Movement Data

명)로부터 1,500개 (동작 당 150개)의 데이터 세트를 수집하였다. 수집된 데이터는 개인별 차이에 따른 과적합을 방지하고 실험결과의 범용성을 높이기 위해, 같은 동작 내에서 전체 데이터셋을 무작위로 섞었으며, 무작위로 섞인 데이터의 80%는 학습/검증 용도로, 나머지 20%의 데이터는 평가를 위해 사용하였으며, 교차 검증방법을 통해 학습에 사용되는 파라미터의 최적화를 진행하였다.

## 2. 구현된 프로토타입의 성능 (처리시간) 평가 결과

표 2는 구현된 프로토타입 상에서 측정한 제안된 손동작 인식 시스템의 처리시간 (성능)을 나타낸다. 데이터 수집단계는 구현된 시스템의 성능이 아닌 인식되는 동작의 종류와 사용자에 따라 다르게 나타나며 실험을 통해 0.8초~1.2초 정도의 시간이 소요되는 것으로 나타났다. 반면 전처리 필터와 분류단계에서는 사용되는 데이터의 수에 따라 다르게 나타나며, 특히 분류 동작의 경우 345us에서 최대 3.9ms 정도의 변화를 보였으며, 이는 사용되는 네트워크 파라미터의 크기에 따라 실행 시간이 크게 증가 되지만, 전처리와 분류의 모든 계산과정은 수ms 미만으로 저사양 MCU 상에서도 실시간 손동작 인식을 하기에 충분히 짧은 처리시간으로 분석된다.

## 3. 설계공간 탐색을 통한 인식 정확도 평가 결과

앞장의 설명한 것과 같이 은닉 계층의 경우 단일 계층과 이중 계층을 대상으로 계층 당 노드의 수를 변경하며 설계 공간 탐색을 진행하였다. 계층 수와 노드의 수 대신 컴퓨팅/메모리 사용량과 직접적 연관이 있는 네트워크 모델 파라미터의 수를 사용하여 표현하였으며, 네트워크 모델 파라미터의 수는 아래의 수식을 사용하여 계산하였다.

$$i \cdot h_1 + \sum_{k=1}^{n-1} (h_k \cdot h_{k+1}) + h_n \cdot o + \sum_{k=1}^n h_k + o. \quad (1)$$

수식에서  $i$ 와  $o$ 는 입력과 출력 계층에 사용된 노드의 수를,  $h_k$ 는  $k$ 번째 은닉 계층에 사용된 노드의 수를, 마지막으로  $n$ 은 사용된 은닉 계층의 수를 의미한다. 참고로 제어기로 사용하고 있는 TI사 마이크로 컨트롤러의 경우 최대 18,100개의 모델 파라미터를 사용할 수 있으며, 이보다 더 경량 마이크로 컨트롤러인 ATMEGA128과 MSP430의 경우 구현에 따라 약간의 차이가 있지만 대략 900개 모델 파라미터를 사용할 수 있다.

그림 10은 제안된 인공신경망 모델 기반 분류기에 사용된 모델 파라미터의 수를 변경해가면서 측정한 인식 정확도를 나타낸다. 파라미터 수 외에도 제안된 전처리 필터의 효과를 평가할 수 있도록 전처리 필터를 사용한 경우 ("W/ Preprocessing")와 사용하지 않는 경우 ("W/O Preprocessing")를 비교하였다. 단 세그멘테이션 필터와 다른 샘플링의 경우 두 가지 방법 모두에 적용하여 평가하였다. 설계공간 탐색의 범위는 입력 계층은 24, 출력 계층은 10으로 고정한 상태에서 총 115개의 모델 파라미터를 갖는

표 2. 각 단계별 처리 시간(성능)

Table 2. Processing Time for Each Step

	Data Collection	Pre-processing	Classification
Processing Time	0.8 ~ 1.2sec	103 ~ 395us	345 ~ 3,887us

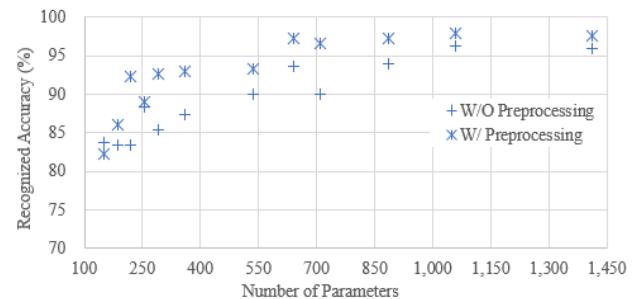


그림 10. 설계공간 탐색을 통한 최적의 네트워크 모델 탐색 결과  
Fig. 10. Result of the Design Space Exploration for Finding the Best Network Model

표 3. 제안된 설계의 인식 정확도 비교

Table 3. Accuracy Comparisons of the Proposed Designs.

# of parms.	NN Structure	Recognition Accuracy (%)		
		W/O	W/	Diff. (%)
220	24×6×10	83.3	92.3	11
360	24×10×10	87.3	93.0	6
640	24×18×10	93.7	97.3	4
710	24×20×10	90.0	96.7	7
885	24×25×10	94.0	97.3	4
1,060	24×30×10	96.3	98.0	2
2,460	24×70×10	95.3	98.7	4

24×3×10구조부터 2,460개의 모델 파라미터를 갖는 24×70×10구조 (이중 은닉 계층의 경우에도 24×35×35×10)까지 수행하였다. 다만 그레프는 일정 이상의 유의미한 정확도를 갖는 모델 구조들만 표현하였으며, 중요한 의미를 갖는 모델 및 성능에 대해서는 표 3에 상세한 구조를 정확도와 함께 표현하였다. 그레프에 표현된 바와 같이 손동작 인식의 경우 플렉스 센서에서 수집되는 데이터들의 특성이 비교적 명확하게 구분이 되기 때문에 복잡하지 않은 네트워크 모델만으로도 비교적 높은 정확도를 보였다. 전처리 필터를 사용하지 않는 경우 150개의 비교적 적은 모델 파라미터를 사용하는 구조의 정확도가 83.7%를 보였다. 이후 모델 파라미터의 수가 증가 될수록 인식의 정확도가 증가하였고, 640개의 모델 파라미터를 사용하는 구조 (24×18×10)에서 93.7%의 정확도를 보이고, 이후부터는 모델 파라미터 수의 증가에 따른 인식 정확도의 증가속도가 낮아지면서 1,060개의 모델 파라미터를 사용하는 구조 (24×50×10)에서 96.3%의 정확도를 보인 후 더는 인식의 정확도가 향상되지 않았다.

반면 제안된 전처리 필터를 사용하는 경우 150개의 모델 파라미터를 사용하는 경우 83.0%의 정확도를 보였다. 이후

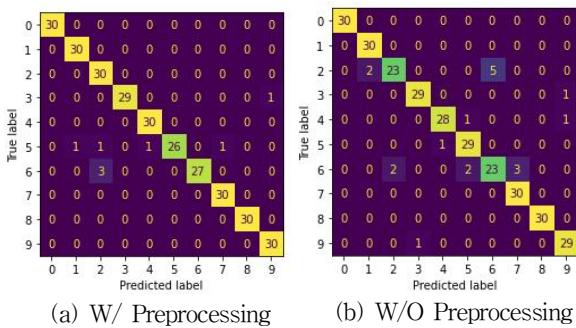


그림 11. 전처리 필터를 사용한 경우와 사용하지 않는 경우로 설계된 손동작 인식기의 혼동 행렬 ( $24 \times 18 \times 10$ )

Fig. 11. Confusion Matrix of the Designed Hand Motion Recognition With and Without Preprocessing ( $24 \times 18 \times 10$ )

모델 파라미터의 수가 증가되면서 전처리 필터를 사용하지 않는 경우보다 인식의 정확도가 빠르게 향상되었으며, 모델 파라미터의 수가 640개를 사용하는 구조 ( $24 \times 18 \times 10$ )에서 인식의 정확도가 97.3%까지 증가한 후 파라미터 수가 2,460인 구조 ( $24 \times 70 \times 10$ )에서 최대 98.7%의 정확도를 보이고 더 이상의 향상은 관찰되지 않았다. 640개의 모델 파라미터의 경우 손동작 인식 프로토 타입 구현에 사용한 TI MCU뿐 아니라 초경량 마이크로 컨트롤러인 ATMEGA128과 MSP430에서도 구현 가능한 크기로서, 제안된 손동작 인식 시스템이 저사양 임베디드 디바이스에서도 높은 97.3%의 인식 정확도를 갖는 손동작 인식기술 구현 가능하다는 것을 알 수 있다.

마지막으로 그림 11은 구현된 손동작 인식기의 혼동 행렬 (Confusion Matrix)을 나타낸다. 전처리 필터를 사용하면서 모델 파라미터의 수 대비 높은 97.3%의 정확도를 보인  $24 \times 18 \times 10$  구조의 경우 (그림 10 (a))와 동일한 모델 구조이지만 전처리 필터를 사용하지 않은 경우 (그림 10 (b))를 비교하였다. 전처리 필터를 사용한 경우 5번 손동작 ("L-Circle")과 6번 손동작 ("Click")을 제외하고는 대부분 100%의 동작 정확도를 보였다. 6번 손동작의 경우 모두 2번 손동작 ("↑")으로 잘 못 인식하는 경우가 발생하였다. 반면 전처리 필터를 사용하지 않은 손동작 인식의 경우 인식 정확도는 93.7%로, 2번 손동작을 1번 또는 6번 손동작으로 잘못 인식하는 경우와 6번 손동작을 2번, 5번, 7번으로 잘못 인식하는 경우가 다수 발생하였다. 이는 전처리 과정을 거치지 않은 입력 데이터의 경우  $24 \times 18 \times 10$  구조에서는 정확한 손동작 인식이 어려움을 의미하며, 전처리 필터를 사용하지 않는 경우 사용되는 모델 파라미터의 수를 2,460개까지 증가시킨  $24 \times 70 \times 10$  구조에서도 전처리 필터를 사용한 인식기보다 인식 성능이 낮았다. 이는 초경량 인공신경망 기반 손동작 인식기술 구현에 있어서 인공신경망 모델 최적화에는 전처리 필터의 설계도 중요한 고려가 되어야 함을 증명한다.

## V. 결 론

웨어러블 센서를 사용한 동작 인식은 인간-컴퓨터 상호

작용 (HCI, Human-Computer Interaction)의 매우 효율적인 인터페이스로 사용될 수 있다. 특히 손가락의 움직임을 포함하는 손동작의 경우 다른 신체 부위와 비교할 때 상대적으로 적은 에너지로 가장 정확하게 움직일 수 있는 신체 부위로써, 효율적인 인터페이스 수단으로 사용될 수 있는 장점이 있다. 이에 본 논문에서는 손동작 인식을 위해 최신의 웨어러블 플렉스 센서를 사용한 인공신경망 기반 인식방법을 제안하고 설계하였다. 제안된 손동작 인식기술은 데이터 수집, 수집된 데이터의 효율적인 처리를 위한 잡음 제거, 데이터 세그멘테이션과 다운 샘플링, 정규화로 구성된 전처리 필터, 그리고 경량 인공신경망 모델 기반의 손동작 인식 분류기로 구성되었다. 제안된 설계는 저사양의 마이크로 컨트롤러를 사용하는 임베디드 디바이스에서도 최소의 자원을 사용하여 구현할 수 있도록 플렉스 센서의 특성과 인식되는 동작의 특성을 고려한 전처리 필터를 설계하였으며, 인공신경망 모델의 설계 공간 탐색을 통해 최대 98.7%의 인식 정확도를 보였다. 마지막으로 제안된 손동작 인식기술은 저사양의 마이크로 컨트롤러와 단일 플렉스 센서를 이용한 프로토 타입을 구현을 통해 그 기능 및 성능을 검증하였다.

본 연구는 웨어러블 센서를 사용한 경량 인공신경망 기반 움직임 인식기술로 움직임 인식에 초점을 두고 연구를 진행하였다. 하지만 웨어러블 형태로 구현되는 특성상 에너지 소모도 고려해야 할 중요한 설계 요소 중 하나이다. 이에 향후 본 연구를 기반으로 에너지 소모까지 고려할 수 있는 에너지-인식성능 최적화 연구가 필요하다.

## References

- [1] K. Oka, Y. Sato, H. Koike, "Real-time Fingertip Tracking and Gesture Recognition," IEEE Computer Graphics and Applications, Vol. 22, No. 6, pp. 64-71, 2002.
- [2] Z. Ren, J. Meng, J. Yuan, Z. Zhang, "Robust Hand Gesture Recognition with Kinect Sensor," Proceedings of ACM international conference on Multimedia (MM '11), pp. 759 - 760, 2011.
- [3] S. Y. Jeon, E. S. Kim, B. Y. Park, "CNN-based Hand Gesture Recognition Method for Teleoperation Control of Industrial Robot," IEMEK Journal of Embedded Systems and Applications, Vol. 16, No. 2, pp. 65-72, 2021.
- [4] L. M. Danga, K. Min, H. Wang, M. J. Piran, C. H. Lee, H. Moon, "Sensor-based and Vision-based Human Activity Recognition: A Comprehensive Survey," Pattern Recognition, 108, 2020.
- [5] Bendlabs, 2-axis Soft Flex Sensor, <https://www.bendlabs.com/products/2-axis-soft-flex-sensor/>
- [6] B. S. Lin, P. C. Hsiao, S. Y. Yang, C. S. Su, I. J. Lee, "Data Glove System Embedded With Inertial Measurement Units for Hand Function Evaluation in Stroke Patients," IEEE Transaction on Neural Systems

- and Rehabilitation Engineering, Vol. 25, No. 11, pp. 2204–2213, 2017.
- [7] C. E. A. Quiapo, K. N. M. Ramos, "Development of a Sign Language Translator Using Simplified Tilt, Flex and Contact Sensor Modules," IEEE Region 10 Conference (TENCON), (2016).
- [8] H. J. Kim, H. T. Park, W. H. Lee, J. W. Kim, Y. L. Park, "Design of Wearable Orthopedic Devices for Treating Forward Head Postures using Pneumatic Artificial Muscles and Flex Sensors," URAI 2017.
- [9] G. Bhat, R. Deb, V. V. Chaurasia, H. Shill, U. Y. Ogras, "Online Human Activity Recognition using Low-Power Wearable Devices," IEEE International Conference on Computer-Aided Design (ICCAD), 2018.
- [10] L. Majeau, J. Borduas, S. Loranger, Y. El-Iraki, J. Lavoie, D. Banville, V. Latendresse, V. Béland, J. Daniel-Rivest, A. Thiaw, H. D. Bambara, T. P. Beausoleil, W. Trottier-Lapointe, J. Lapointe, "Dataglove for Consumer Applications," 2011 7th International Workshop on Fibre and Optical Passive Components, 2011.
- [11] M. Kim, J. Cho, S. Lee, Y. Jung, "IMU Sensor-Based Hand Gesture Recognition for Human-Machine Interfaces," Sensors, Vol. 19, No. 18, 2019.
- [12] L. Chen, J. Fu, Y. Wu, H. Li, B. Zheng, "Hand Gesture Recognition Using Compact CNN via Surface Electromyography Signals," Sensors, Vol. 20, No. 3, 2020.
- [13] W. Jung, H. G. Lee, "Design and Performance Analysis of ML Techniques for Finger Motion Recognition," Journal of the Korea Industrial Information Systems Research, Vol. 25, No. 2, pp. 129–136, 2020.
- [14] A. K. Panda, R. Chakravarty, S. Moulik, "Hand Gesture Recognition using Flex Sensor and Machine Learning Algorithms," EMBS Conference on Biomedical Engineering and Science, 2020.
- [15] W. C. Chuang, W. J. Hwan, T. M. Tai, D. R. Huang, Y. J. Jhang, "Continuous Finger Gesture Recognition Based on Flex Sensors," Sensors, Vol. 19, No. 18, August, 2019.
- [16] J. Chan, E. Veas, J. Simon, "Designing a Sensor Glove Using Deep Learning," International Conference on Intelligent User Interface, April, 2021.
- [17] Texas Instruments, Datasheet of CC2652R  
<https://www.ti.com/product/CC2652R>

### Hyung Gyu Lee (이형규)



2001 Computer Science and Engineering from Seoul National University (M.S.)  
 2007 Computer Science and Engineering from Seoul National University (Ph.D)  
 2021~Duksung Women's University (Assoc. Prof.)

#### Career:

2007~2010 Senior Engineer, Samsung Electronics  
 2010~2012 Research Faculty, Georgia Institute of Technology  
 2012~2021 Assistant/Associate Professor, Daegu University  
 Field of Interests: Embedded System Design, Low Power System, Memory System, Energy Harvesting, Wearable IoTs  
 Email: hglee@duksung.ac.kr