논문 20-45-12-18 The Journal of Korean Institute of Communications and Information Sciences '20-12 Vol.45 No.12
https://doi.org/10.7840/kics.2020.45.12.2140

사람의 동작 인식 및 기기 제어를 위한 가속도와 초음파 융합 기반 손동작 분류 기법

천 진 원*, 최 선 웅°

Hand Gesture Classification Technique Based on Acceleration and Ultrasound Data Fusion for Human Movement Recognition and Device Control

Jinwon Cheon*, Sunwoong Choi°

요 약

IT 기술이 발전함에 따라 웨어러블 기기와 IoT(Internet of Things) 기기들이 늘어나고, 사람들에게 보편화되고 있다. 휴대성을 위해 기기들의 크기가 점점 작아지면서, 일반적인 버튼이나 터치를 이용한 기술로 제어하는 것은 점점 어려워지고 있다. 이러한 문제점을 해결하기 위해서, 사람의 동작을 인식하여 기기를 제어하는 새로운 방식의 인터페이스에 대한 연구들이 진행되고 있다. 본 논문에서는 추가적인 센서를 사용하지 않고, 스마트폰의 마이크와 스피커, 스마트워치에 내장되어있는 가속도계만으로 손동작을 분류하는 방법을 제시한다. 가속도 데이터와 초음파 데이터를 융합(fusion)하여, 손동작 분류의 정확도를 높이는 방법을 제안한다. 데이터를 융합하는 연산에 따라서 Max, Add, Concat 모델로 구분한다. 융합 모델은 한 가지의 데이터만 학습한 모델에 비해 분류 정확도가 향상되는 것을 확인하였다. 성능이 가장 높은 모델인 Concat 모델은 10가지 패턴에 대해서 90.0%의 분류 정확도를 보이는 것을 확인하였고, 이는 가속도 데이터만 사용하여 학습한 모델보다 5.8%, 초음파 데이터만 사용한 모델보다는 16.4% 향상된 분류 정확도이다.

Key Words : Gesture Recognition, Acceleration Data, Ultrasound Data, Fusion Model, Short-Time Fourier Transform

ABSTRACT

As IT technology grows, wearable devices and Internet of Things devices are increasing, and they are becoming more common for people. As devices get smaller and smaller in size for portability, it is becoming increasingly difficult to control using general buttons or touches. To solve this problem, studies are being conducted on a new interface that controls a device by recognizing human motion. In this paper, we present a method of classifying hand gestures with only the microphone and speaker in the smart phone, and accelerometer built into the smart watch without using an additional sensor. We propose a method to increase the accuracy of hand gesture classification by fusion of acceleration data and ultrasound data. According to the

※ 본 연구는 과학기술정보통신부의 재원으로 한국연구재단, 무인이동체원천기술개발사업단의 지원을 받아 수행된 기초연구사업 및 무인이동체원천기술개발사업을 통해 수행되었음.(No.2016R1A5A1012966, No. 2020M3C1C1A01084837)

* First Author : Kookmin University, Department of Security Enhanced Smart Electric Vehicle, jinwontoo@kookmin.ac.kr, 학생(석사과정), 학생회원

° Corresponding Author : Kookmin University, School of Electrical Engineering, schoi@kookmin.ac.kr, 정교수, 종신회원

논문번호 : 202008-203-C-RN, Received August 19, 2020; Revised September 10, 2020; Accepted September 11, 2020

fusion operation, it is classified into Max, Add, and Concat models. These fusion models improved classification accuracy compared to models that train only one part of the data. The Concat model, which is the best performing model showed 90.0% classification in 10 patterns, this classification accuracy is improved by 5.8% compared to the model trained using only acceleration data and 16.4% compared to the model using only ultrasound data.

Ⅰ. 서 론

IT 기술이 발전함에 따라 웨어러블 기기와 IoT(Internet of Things) 기기들이 늘어나고, 사람들에게 보편화되고 있다. 스마트폰과 더불어 웨어러블 기기, IoT 기기 등 다양한 종류의 제품들에는 일반적으로 버튼이나 터치 기술이 적용되고 있다. 하지만 이 기기들은 휴대성을 위해 크기가 점점 소형화되었고, 버튼이나 터치 기술은 이를 조작하는데 한계가 있기 때문에 새로운 인터페이스의 도입이 필요해졌다. 이러한 이유로 음성 인식이나 동작 인식 기술에 대한 연구가 활발히 이루어지고 있다.

음성 인식 기술은 마이크를 통해 얻은 신호를 분석하여 단어나 문장으로 변환하는 것을 말한다. 이 기술을 활용한 대표적인 서비스는 음성 인식 비서 서비스이다. Apple의 Siri[1]나 삼성의 Bixby[2], Google의 Google Assistant[3] 등 다양한 기업에서 음성 인식 서비스를 개발하고 있다. 하지만, 음성 인식 기술은 말을 할 수 없는 상황이나 소음이 심하게 발생하는 곳에서는 사용이 제한된다는 단점을 가지고 있다.

동작 인식 기술은 다양한 센서로부터 사람의 동작을 분석하고 판단하는 기술을 말한다. 이 기술에는 카메라와 3D 센서, 관성 측정 장치(IMU, Inertial Measurement Unit), 초음파 신호 등이 활용된다. 카메라와 3D 센서를 이용하는 기술들은 시각 정보를 받아들여 사용자의 동작을 인식한다[4-9]. 이 기술은 센서를 따로 거치하여 사용해야 한다는 단점이 존재한다. 관성 측정 장치를 사용하는 기술은 가속도 센서, 자이로스코프, 지자기 센서 등을 사용하여 사용자의 동작을 인식할 수 있다[10,11]. 또한, 초음파 신호를 활용해 사람의 손동작을 인식하는 연구들도 있다[12-16]. 이 연구들은 기기의 스피커에서 초음파 신호를 발생시킨 뒤, 손에 의해 반사되는 신호를 마이크에서 받아들이고, 그 신호를 분석하여 동작을 인식한다. 초음파 신호가 손에 맞고 반사될 때 도플러 효과로 인해 파동의 진동수에 변화가 발생하는데, 이를 분석하여 손동작을 인식할 수 있다.

본 논문에서는 음성 인식 기술과 카메라 기반 동작 인식 기술의 단점을 극복하기 위해 스마트워치에 내장된 가속도 센서와 스마트폰의 초음파 신호를 활용하여 손동작을 분류하는 시스템을 개발하였다. 10가지 패턴에 대한 가속도 데이터와 전처리된 초음파 데이터를 각각 CNN(Convolutional Neural Network) 모델로 학습하고 분류한다. 우리는 두 CNN 모델을 융합(fusion)하여 분류 정확도를 높이는 세 가지 방법을 제시한다. 세 방법은 모델을 어떻게 연산하는지에 따라 Max, Add, Concat(concatenate) 모델이라고 부른다. Max와 Add 모델은 fully connected의 결과 행렬을 연산한 후 학습을 진행하고, Concat 모델은 CNN층을 이어 붙여 학습한다. 가속도 데이터만 학습시킨 CNN 모델과 초음파 데이터만 학습시킨 CNN 모델의 분류 정확도는 84.2%, 73.6%이었지만, 융합 모델의 분류 정확도는 단일 모델보다 모두 향상되었다. 특히, Concat 모델은 90.0%의 분류 정확도를 보여 단일 모델에 비해 5.8%, 16.4% 분류 정확도가 향상된 것을 확인할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 손동작 분류에 대한 이전 연구와 논문에 대해 소개한다. 3장에서는 본 논문에서 제시하는 방법에 대해 설명하고, 4장에서는 실제 실험이 이루어진 환경과 모델에 대한 평가를 기술한다. 마지막으로, 5장에서는 결론으로 마무리 짓는다.

Ⅱ. 관련 연구

이번 장에서는 손동작 인식 및 분류가 어떻게 활용되고 있는지 간략히 소개하고, 스마트워치에 내장된 관성 측정 장치를 활용하여 손동작을 분류하는 연구들과 초음파 신호를 활용한 연구, 그리고 CNN 모델을 융합하는 연구에 대해서 설명한다.

먼저, 손동작 인식 및 분류에 대한 연구 현황을 소개한다[17]. 손동작 인식 및 분류를 위한 데이터는 주로 세 가지 유형의 센서로부터 얻게 된다. 첫째로 가속도계와 자이로스코프 등으로 손과 손가락의 움직임을 감지하는 마운트 기반 센서, 둘째로 모바일 장치에 직접 터치를 하는 멀티 터치 스크린 센서, 마지막으로

시각 기반 센서가 있다. 이러한 센서로부터 데이터를 얻은 후, 학습한 모델들은 수화 인식, 가상 조작, 게임, 인간-로봇 상호작용 등에 적용되고 있다.

스마트워치 관련 연구로는, 애플 워치의 가속도 센서만을 활용하여 손동작 패턴을 인식하는 연구가 있다[11]. CNN 모델에 x, y, z 축의 가속도 데이터를 학습시켜 손가락으로 그릴 수 있는 10가지 패턴을 분류한다. 다른 논문[10]에서는 가속도 센서와 자이로스코프를 활용하여 몇 가지 손동작과 손가락으로 알파벳을 그리는 동작을 분류한다.

초음파 신호를 활용한 연구에는, 스마트폰과 스마트워치를 활용하여 미세한 손가락의 위치를 추적하는 FingerIO[12]라는 연구가 있다. 이 연구에서는 18-20㎑ 주파수 대역의 초음파 신호를 사용하고 OFDM(Orthogonal Frequency Division Multiplexing) 방식을 활용하여 손가락의 미세한 위치를 추적한다. LLAP[14](Low-Latency Acoustic Phase)라는 다른 연구에서는 추가적인 디바이스 없이 스마트폰으로만 손동작을 추적하는 시스템을 제안한다. 이 연구에서는 스마트폰에 내장된 스피커와 마이크만을 가지고 손과 손가락을 추적한다. 핵심 아이디어는 음향 위상을 사용하여 미세한 이동 방향 및 거리를 알아내는 것이다. LLAP는 손과 손가락의 움직임으로 발생하는 음성 신호의 위상 변화를 측정하고 이를 거리로 변환함으로써 위치를 파악한다. EchoTrack[13]이라는 논문에서도 스마트폰의 스피커와 마이크를 활용해 손의 움직임을 추적한다. 이 연구에서는, 스피커에서 16-23㎑ 주파수 대역의 초음파 신호를 발생시키고, 마이크에서는 손 움직임에 의해 반사되는 신호를 받아들인다. ToF(Time of Flight) 방식을 활용하여 손과 마이크 사이의 거리를 측정하고, 지속적으로 손의 움직임을 추적한다.

마지막으로, 다수의 CNN 모델을 구성하고, 이를 융합하여 시스템의 분류 성능을 향상시키는 연구들이 있다. 그 중 한 논문[18]에서 제안하는 CNN 모델은 이미지 안의 사람이 취하고 있는 행동을 인식한다. 주요한 아이디어는 GoogLeNet, VGGNet, ResNet 세 CNN 모델을 융합하여 성능을 향상시키는 것이다. 각 모델에 특징층(100차원, 40차원의 fully connected와 softmax)을 추출하고 이를 이어 붙이는(concatenate) 것이다. 한 이미지에 대해서 개별적인 모델(GoogLeNet, VGGNet, ResNet)만 사용했을 때 보다 두 가지 모델을 융합했을 때 정확도가 더 높았으며, 세 가지 모델을 전부 융합했을 때 가장 높은 정확도를 보인다. 또 다른 논문[19]에서는 스마트폰과 스마트워치에서 수집한 데이터셋을 CNN 모델에 학습시킨다. 학습시킬 때 어느 단계에서 모델을 융합하는지에 따라 데이터 단계의 융합(data-level fusion), 특징 단계의 융합(feature-level fusion), 결정 단계의 융합(decision-level fusion) 세 가지 방법으로 분류된다. 이 논문에서도 융합 방법을 사용했을 때 모델의 정확도가 향상되었다.

Ⅲ. 가속도와 초음파 기반 손동작 분류 방법

그림 1은 우리의 시스템을 간단하게 도식화한 그림이다. 시스템은 크게 데이터 수집, 데이터 전처리, 모델 학습, 융합(fusion)으로 구성된다. 먼저, 우리는 스마트폰과 스마트워치에서 우리가 개발한 어플리케이션으로 데이터를 수집한다. 수집한 데이터 중 초음파 데이터에 STFT(Short-Time Fourier Transform)를 적용한다. 그 다음, 서버에서는 제안하는 CNN 모델에 데이터를 학습시키고, 이를 융합한다.

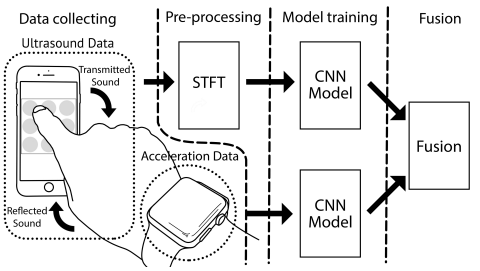


그림 1. 제안하는 모델의 시스템 개요
Fig. 1. System overview of proposed model

3.1 데이터 수집

우리는 2차원에서 그릴 수 있는 임의의 10가지 패턴에 대해 데이터를 수집했다. 그림 2는 가로와 세로가 각각 3개로 이루어진 9개의 기준점에서 만들어질 수 있는 10가지의 2차원 패턴을 도식화한 것이다. 모든 패턴은 좌측 상단 점에서 시작하여 화살표의 경로

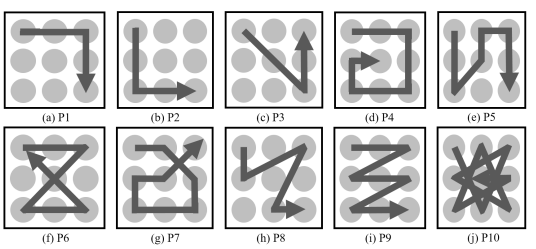


그림 2. 데이터셋으로 사용한 패턴
Fig. 2. Pattern used as dataset

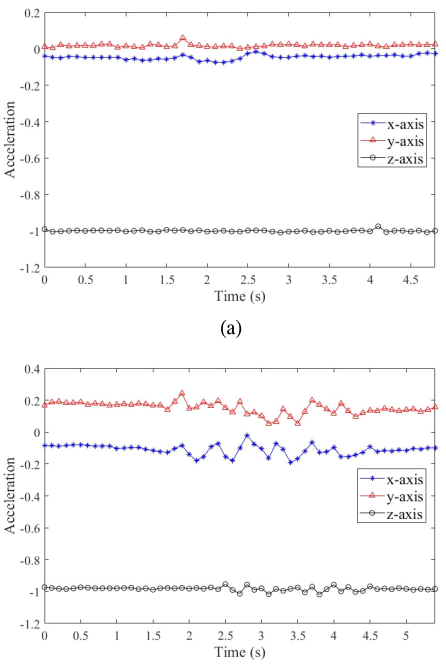
를 따라 진행되다가 화살표가 가리키는 점에서 끝난다. 스마트폰은 화면이 보이도록 평평한 책상 위에 올려두고, 스마트폰 화면에서 약 1cm 위에서 스마트워치를 착용한 손으로 동작을 취함으로써 데이터를 수집했다.

학습에 사용할 데이터는 가속도 데이터와 초음파 데이터 두 종류이다. 가속도 데이터는 스마트워치에서 수집되고, 초음파 데이터는 스마트폰의 마이크에서 단일 주파수 초음파를 녹음한 것이다. 두 기기에서 데이터를 측정하기 위해서 우리가 직접 개발한 어플리케이션을 사용했다.

3.1.1 가속도 데이터

스마트폰 어플리케이션에서 데이터 수집을 시작하면 스마트워치의 가속도계가 활성화되어 가속도 데이터를 저장한다. 하나의 패턴은 10초 안에 한 번 수행된다. 가속도계의 샘플링 속도는 10㎐ 로 x, y, z 세 축에 대해 각 100개씩 데이터가 수집되어 한 패턴에 총 300개의 데이터가 존재한다.

그림 3은 두 가지 손동작 패턴에 대해서 가속도 데이터를 그래프로 그린 것이다. 그림 3(a)는 방향 전환이 한 번 있는 P1의 손동작에서, 그림 3(b)는 방향 전환이 여섯 번 있는 P7의 손동작에서 얻은 데이터 그래프이다.



(a)

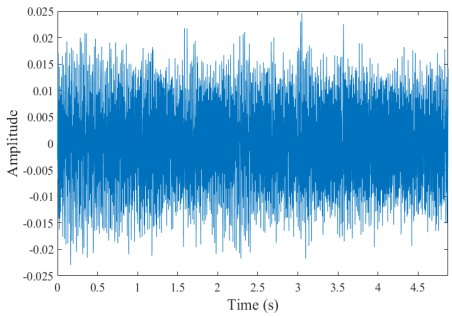
(b)

그림 3. (a) P1의 가속도 그래프; (b) P7의 가속도 그래프
Fig. 3. (a) P1 acceleration graph; (b) P7 acceleration graph

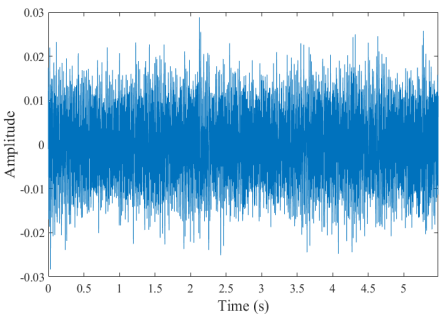
3.1.2 초음파 데이터

스마트워치에서 가속도 데이터를 수집하는 동안, 스마트폰에서는 소리 신호를 수집한다. 어플리케이션은 스마트폰의 스피커에서 20㎑의 단일 주파수 초음파를 발생시킨다. 마이크에서는 손에 맞고 반사되는 신호를 수집한다. 마이크도 가속도계와 마찬가지로 10초 동안 활성화되어 데이터를 수집한다. 이 신호의 샘플링 속도는 44.1㎑이다.

마이크에서 녹음되는 반사된 신호는 도플러 효과로 인해 손에 맞고 반사되는 동안 파동과 진동수에 변화가 생기기 때문에 패턴마다 특징이 존재한다. 그림 4는 그림 3에서 가속도 데이터 그래프를 그렸던 패턴에 대해 소리 신호의 진폭을 비교한 그림이다. 하지만 아무런 처리도 하지 않은 신호에서 사람의 눈으로 각 손동작의 특징을 알아보는 것은 쉽지 않다. 그렇기 때문에 이 신호를 전처리하여 20㎑ 대역의 초음파 데이터를 자세히 분석한다.



(a)



(b)

그림 4. (a) P1의 신호 진폭 그래프; (b) P7의 신호 진폭 그래프
Fig. 4. (a) P1 signal amplitude graph; (b) P7 signal amplitude graph

3.2 STFT(Short-Time Fourier Transform)

우리는 초음파 데이터를 시간 영역과 주파수 영역에서 분석하기 위해 STFT 과정을 거친다. STFT는 데이터를 여러 조각(window)으로 나눈 후 Fourier Transform을 적용하여, 시간 정보의 손실 없이 주파수를 분석할 수 있는 방법이다. STFT의 결과는 x 축은 시간 축, y 축은 주파수 축으로 그래프가 그려지게 되고 해당 영역의 값이 색깔로 표시되는 3차원 데이터이다. 그림 5는 그림 4의 신호를 STFT한 후 20㎑ 근처만 잘라낸 그래프이다. 그림 5(a)는 20㎑를 중심으로 4번 정도의 색깔 변화가 있는 반면, 그림 5(b)는 색깔 변화가 10번 이상 존재한다. 이처럼 STFT를 하면, 전처리하지 않은 초음파 데이터에 비해서 패턴 간의 차이를 확연하게 볼 수 있다. 모델을 학습시킬 때에는 가속도 데이터와 더불어 STFT를 거친 초음파 데이터를 학습시킨다.

초음파 데이터는 한 동작 당 44.1㎑의 샘플링 속도로 10초 간 녹음됐기 때문에, 441,000개의 데이터가 있는데, 이 데이터에 STFT를 적용하면 4,990개의 데이터를 얻을 수 있다. STFT의 자세한 사항은 다음과 같다. 윈도우 사이즈는 17,000, 오버랩은 윈도우 사이즈의 95%인 16,150, 주파수 해상도는 4,096로 설정했다. 주파수는 10.77㎐ 간격으로 나뉘는데, 관심 영역인 20㎑ 근처의 10개의 주파수 영역, 구체적으로 19.951-20.047㎑ 대역을 사용하였다. 17,000의 윈도우 사이즈와 10.77㎐로 둘러싸인 한 영역 당 499개의 데이터가 존재하므로, 10개 영역에 대해서는 4,990개의 데이터가 존재한다.

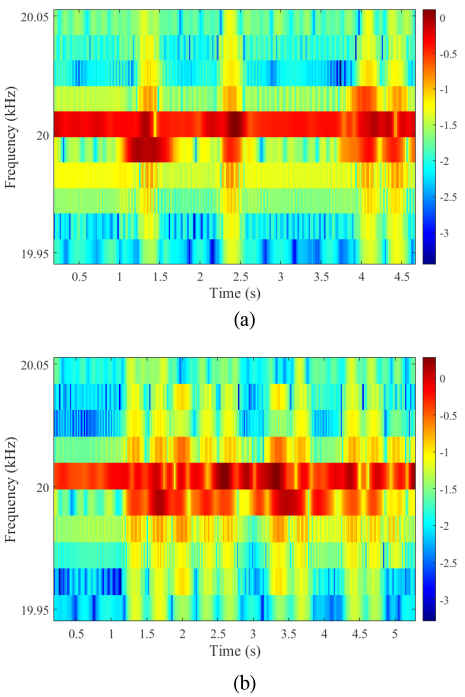


그림 5. (a) P1의 STFT 스펙트로그램; (b) P7의 STFT 스펙트로그램
Fig. 5. (a) P1 STFT spectrogram; (b) P7 STFT spectrogram

3.3 모델 학습

본 논문에서는 가속도 데이터를 학습시키는 모델을 가속도 모델, 초음파 데이터를 학습시키는 모델을 초음파 모델이라고 부른다. 가속도 모델은 가속도 데이터의 크기를 입력으로, 초음파 모델은 STFT된 초음파 데이터의 크기를 입력으로 하는 CNN 모델이다. 두 모델은 각 모델에 해당하는 데이터가 입력됐을 때, CNN 모델을 거치면서 10개의 패턴에 대한 확률을 계산하고, 그중 확률이 가장 높은 패턴을 출력함으로써 학습을 진행한다.

가속도 모델 구조는 그림 6과 같다. 모델은 크게 입력층, CNN층, fully connected, softmax로 구성되어 있다. 입력층에서는 3×100개의 데이터를 읽어온다. CNN층에서는 feature map을 두 배씩 늘리는 convolution layer와 데이터의 사이즈를 반으로 줄이는 pooling layer로 이루어져 있다. Convolution layer와 pooling layer는 총 6쌍이 존재하며, 이외에도 과적합을 방지하기 위한 drop out을 추가하였다. Fully connected에서는 CNN층의 출력을 전부 연결한다. 그 다음 softmax 함수를 거쳐 총 10가지의 동작 중 하나를 예측한다. Softmax 함수의 결과값 중 가장 큰 값을 가지는 항목을 이 모델의 예측 결과이다. 이 모델에서 사용한, 활성화 함수는 ReLU, 비용 함수는 cross entropy, 최적화기는 Adam이다.

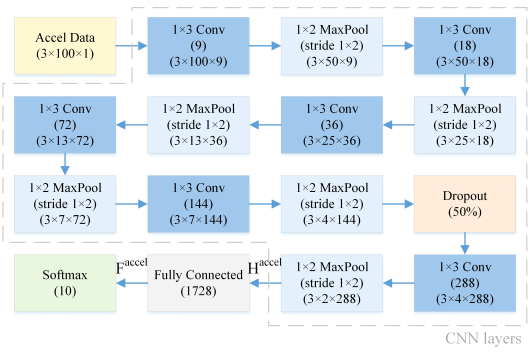


그림 6. 가속도 모델의 구조
Fig. 6. Structure of Acceleration model

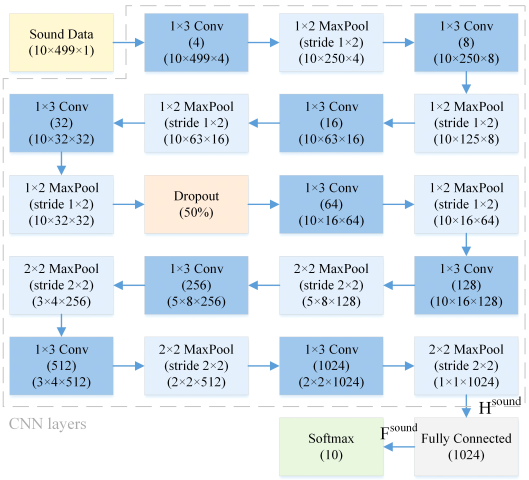


그림 7. 초음파 모델의 구조
Fig. 7. Structure of Ultrasound model

초음파 모델도 가속도 모델과 구조가 비슷하다. 모델 구조는 그림 7과 같다. 데이터가 10개의 영역에 대해서 499개씩 존재하기 때문에, 입력층에서는 10×499개의 데이터를 읽어온다. Convolution layer와 pooling layer는 가속도 모델보다 3쌍 많은 9쌍이 존재한다. 이 모델에서 사용한 활성화 함수, 비용 함수, 최적화기는 모두 가속도 모델과 같다.

3.4 모델 융합 방법

본 논문에서는 위에서 설명한 두 개의 단일 모델(가속도 모델, 초음파 모델)을 세 가지 방법으로 융합하여 손동작을 분류하는 모델을 제시한다. 우리가 제시하는 융합 방법은 가속도 모델과 초음파 모델을 연산하는 방법에 따라 Max, Add, Concat 모델로 구분한다. 세 방법은 fully connected와 CNN층을 어떻게 연산하는지에 따라 나뉜다. 두 fully connected의 결과값 중 더 큰 값으로 새로운 행렬을 만들어 학습하는 Max 모델, 두 fully connected의 결과값을 더한 값으로 학습하는 Add 모델, 마지막으로 두 모델의 CNN층의 출력을 이어붙인 후 fully connected와 softmax를 거치는 Concat(concatenate) 모델을 제안한다. 융합 모델을 학습 및 평가할 때에는 한 동작에 대해서 가속도 데이터는 가속도 모델에, STFT된 초음파 데이터는 초음파 모델에 입력한 다음, 각 융합 방법에 따라서 예측값을 계산하고 그 결과로 패턴을 예측한다.

Max 모델은 한 동작에 대한 가속도 모델의 fully connected의 결과 행렬과 초음파 모델의 fully connected의 결과 행렬 중 각 손동작에 해당하는 두 개의 결과값 중 더 큰 값으로 크기가 10인 새로운 행렬을 만든다. Max 연산에 대한 식은 식 (1)과 같다. 연산의 결과 행렬인 F^{max}에 softmax한 출력이 Max 모델의 결과가 된다.

$$F^{max} = \max\{F^{accel}, F^{sound}\} \quad (1)$$

Add 모델도 Max 모델과 유사한 방식으로 학습이 진행된다. Add 연산에 대한 식은 식 (2)와 같다. 한 동작에 대한 가속도 모델의 fully connected의 결과 행렬과 초음파 모델의 fully connected의 결과 행렬의 각 요소를 더해서 크기가 10인 새로운 행렬을 만든다. 식 (1)과 마찬가지로, 새로운 행렬인 F^{add}에서 softmax한 출력이 Add 모델의 결과이다. 그림 8은 Max와 Add 모델의 모델 융합 방법을 도식화한 그림이다.

$$F^{add} = F^{accel} + F^{sound} \quad (2)$$

Concat 모델은 Max, Add 모델과는 학습 방법에 차이가 있다. Max, Add 모델이 fully connected의 출력 단계에서 연산했던 것과 다르게, Concat 모델은 CNN층을 연산한다. 가속도, 초음파 모델의 CNN층을 이어 붙여서 학습을 진행한다. 식 (3)은 Concat 모델의 연산을 수식으로 나타낸 것이다. H는 각 모델의 CNN층을 거친 결과이다. 그림 9는 Concat 모델의 모델 융합 방법을 도식화한 그림이다.

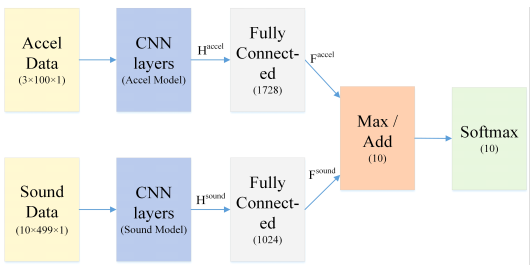


그림 8. Max와 Add 모델의 구조
Fig. 8. Structure of Max and Add model

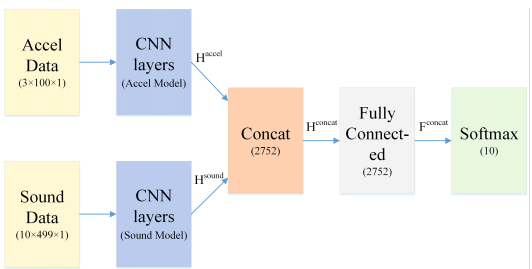


그림 9. Concat 모델의 구조
Fig. 9. Structure of Concat model

$$H^{concat} = concat[H^{accel}, H^{sound}] \quad (3)$$

Ⅳ. 성능 평가

4.1 실험 환경

데이터를 수집하는 방법으로는 우리가 제안하는 어플리케이션을 사용하였다. 데이터 수집에 활용된 기기들은 iPhone 8과 Apple Watch Series 3이다. 수집을 위한 어플리케이션 개발은 Xcode 환경에서 Swift 언어로 작성되었다.

데이터 수집은 소음이 적은 사무실에서 이루어졌으며, 그림 10처럼 스마트폰은 평평한 책상 위에 위치하였고, 스마트워치는 동작을 수행하는 손에 착용하였다. 데이터 수집을 완료한 후, 초음파 데이터는 MATLAB R2019b 버전에서 STFT 과정을 거쳤다. 그 후 두 종류의 데이터를 GPU 서버에서 두 가지 단일 모델과 세 가지 융합 모델에 학습시켰다. GPU 서버의 사양은 TITAN XP이다.

그림 10. 데이터 수집 환경
Fig. 10. Data collection environment

4.2 데이터셋

한 패턴에 대해서 50회 씩 손동작을 수행하여 총 가속도 데이터 500개, 초음파 데이터 500개를 얻었다. 학습할 때는, 동시에 수집한 한 쌍의 가속도 데이터와 초음파 데이터를 같이 사용한다. 각 패턴 당 데이터의 개수는 표 1과 같다.

표 1. 모델 학습에 사용한 데이터 개수
Table 1. Number of the used to train models

Pattern	Number of Acceleration Data	Number of Sound Data
P1	50	50
P2	50	50
P3	50	50
P4	50	50
P5	50	50
P6	50	50
P7	50	50
P8	50	50
P9	50	50
P10	50	50
Total	500	500

4.3 모델 평가

본 논문에서는 모델의 성능에 대한 신뢰도를 높이기 위하여 k-fold cross validation 기법을 사용했다. 이 기법은 데이터의 수가 적을 때 데이터셋을 k개로 나눈 후 k번 반복 학습하여 성능 평가에 대한 신뢰도를 높이는 방법이다. 우리는 5-fold cross validation을 활용하여 평가에 대한 신뢰도를 높였다.

수집한 한 쌍의 데이터 중, 가속도 모델을 평가할 땐 가속도 데이터만, 초음파 모델을 평가할 땐 STFT 된 초음파 데이터만 사용했고, 융합 모델을 평가할 땐 두 데이터를 같이 사용하여 평가를 진행했다.

본 논문에서는 precision, recall, F1-score, accuracy 등 다양한 평가지표를 활용하여 모델을 평가하였다. 그 결과는 표 2와 같다. 한 모델에 대해서 각 성능 평가지표의 값은 1-2% 정도밖에 차이가 나지 않았지만, 모델 간의 차이는 눈여겨볼만 하다. 한 종류의 데이터만 학습하는 가속도 모델과 초음파 모델보다 본 논문에서 제시하는 세 가지 융합 모델들의 성능이 더 좋은 것을 확인할 수 있다. 그 중에서도 fully connected의 결과를 연산하는 Max와 Add 모델보다 CNN층을 이어 붙여서 학습한 Concat 모델이 더 좋은 성능을 보여준다. 분류 정확도가 각각 84.2%, 73.6%인 가속도, 초음파 모델에 비해서, Concat 모델의 분류 정확도는 90.0%로 각 모델보다 5.8%, 16.4% 성능이 향상되었다.

우리는 각 클래스 별 정확도를 구체적으로 알아보

표 2. 각 모델의 성능 비교
Table 2. The performance comparison of each model

	Single models		Fusion models		
	Accel	Sound	Max	Add	Concat
Precision	84.3%	74.8%	87.5%	87.9%	90.3%
Recall	83.6%	74.9%	87.0%	87.6%	90.0%
F1-score	82.9%	72.6%	86.7%	87.1%	89.7%
Accuracy	84.2%	73.6%	87.0%	87.4%	90.0%

기 위해 confusion matrix를 작성해보았다. 그림 11과 12는 각각 가속도 모델과 초음파 모델의 confusion matrix를 작성한 것이고, 그림 13은 우리가 제안하는 모델 중 Concat 모델의 confusion matrix를 작성한 것이다. 그림 11, 12에서 볼 수 있듯이, 가속도 모델과 초음파 모델의 정확도가 높은 클래스는 차이가 있다. 가속도 모델은 P6, P7, P9에서 비교적 정확도가 떨어졌고, 초음파 모델은 P4, P5, P8에서 그러하다. 이로써 모델의 분류 정확도가 비슷하더라도 클래스 별 정확도는 큰 차이가 있음을 알 수 있다.

그림 14는 본 논문에서 사용했던 10가지 손동작 패턴에 대해서 다섯 가지 모델의 분류 정확도를 그래프로 그린 것이다. Confusion matrix에서 확인할 수 있듯이, 가속도 모델에서 분류가 잘 되는 패턴과 초음파 모델에서 분류가 잘 되는 패턴이 구분된다. 그림 14에서는 거의 모든 패턴에 대해서 융합 모델이 단일 모델에 비해서 성능이 향상되었다는 것을 확인할 수 있다. 특히, Concat 모델이 가장 큰 폭으로 성능이 향상되었다.

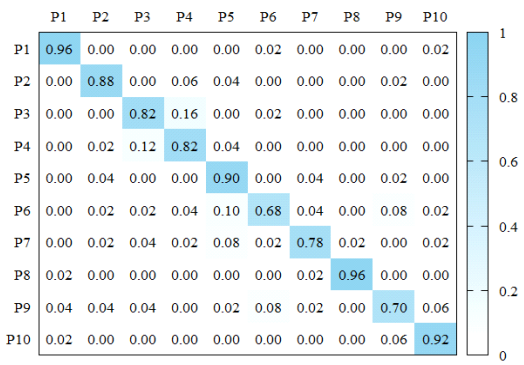


그림 11. 가속도 모델의 confusion matrix
Fig. 11. Confusion matrix of Acceleration model

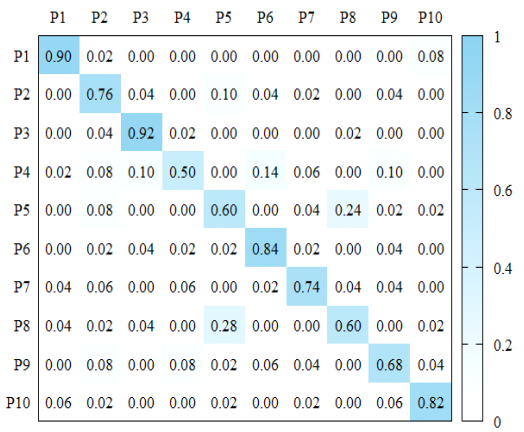


그림 12. 초음파 모델의 confusion matrix
Fig. 12. Confusion matrix of Ultrasound model

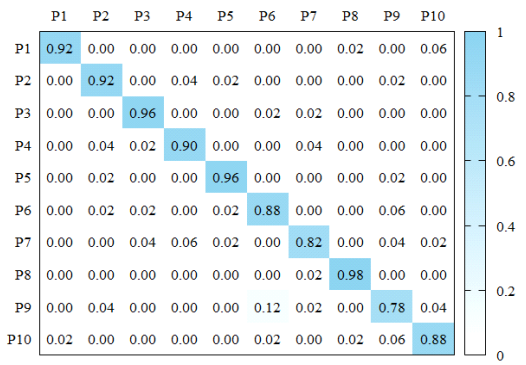


그림 13. 제안하는 Concat 모델의 confusion matrix
Fig. 13. Confusion matrix of proposed Concat model

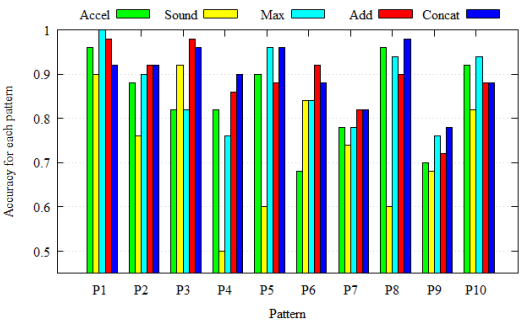


그림 14. 각 패턴에 대한 다섯 가지 모델의 분류 정확도
Fig. 14. Classification accuracy of five models for each pattern

V. 결 론

본 논문에서는 손동작을 인식할 수 있는 여러 방법 중 가속도 데이터와 초음파 데이터를 같이 사용하여 손동작을 분류하는 방법을 제안하였다. 이 방법은 스마트워치에서는 가속도계만 사용하고, 스마트폰에서는 마이크와 스피커만 사용하기 때문에 추가적인 센서 없이 내장되어 있는 센서들로 구현이 가능하다. 우리는 손동작을 수행하는 동안 스마트워치에서 가속도 데이터를, 스마트폰에서 초음파 데이터를 수집하였고, 초음파 데이터는 STFT를 거쳐 각각의 CNN 모델에 학습시켰다. 가속도 데이터만 학습한 가속도 모델은 84.2%, STFT된 초음파 데이터만 학습한 초음파 모델은 73.6%의 손동작 분류 정확도를 보였다. 우리는 두 모델을 상호보완하기 위해 모델을 융합하는 세 가지

방법을 제시하였다. 융합하는 방법에 따라 Max, Add, Concat로 분류하였고, 10가지 패턴에 대해 성능을 비교해보았을 때 하나의 데이터만 학습한 단일 모델보다 융합 모델의 성능이 더 좋았다. 그 중에서도 Concat 모델의 성능이 가장 향상되었는데, 가속도 모델보다 5.8%, 초음파 모델보다 16.4% 높은 분류 정확도를 보였다. 추가적으로, 두 단일 모델의 confusion matrix를 비교해보면 클래스별 정확도의 양상이 다른 것을 알 수 있는데, Concat 모델의 confusion matrix를 보면 두 모델이 적절하게 상호보완됐다는 것을 확인할 수 있다.

References

[1] http://www.apple.com/siri/.

[2] https://www.samsung.com/sec/apps/bixby/.

[3] https://assistant.google.com/.

[4] C. Chen, C. Ting, and Y. Huang, "3D interactive system using integral photography by embedded optical sensors for portable devices," *J. Display Technol.*, vol. 12, no. 11, pp. 1329-1334, 2016.

[5] F. Erden and A. E. Cetin, "Hand gesture based remote control system using infrared sensors and a camera," *IEEE Trans. Consumer Electron.*, vol. 60, no. 4, pp. 675-680, 2014.

[6] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni, "Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 234-245, 2019.

[7] N. M. DiFilippo and M. K. Jouaneh, "Characterization of different microsoft kinect sensor models," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4554-4564, 2015.

[8] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning actionlet ensemble for 3D human action recognition," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 36, no. 5, pp. 914-927, 2014.

[9] G. Plouffe and A. Cretu, "Static and dynamic hand gesture recognition in depth data using dynamic time warping," *IEEE Trans. Instrumentation and Measurement*, vol. 65, no. 2, pp. 305-316, 2016.

[10] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proc. 16th HotMobile '15*, pp. 9-14, New York, NY, USA, 2015.

[11] M.-C. Kwon, G. Park, and S. Choi, "Smartwatch user interface implementation using CNN-based gesture pattern recognition," *Sensors*, vol. 18, Basel, Switzerland, 2018.

[12] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "FingerIO: Using active sonar for fine-grained finger tracking," *CHI'16*, pp. 1515-1525, San Jose, CA, USA, May 2016.

[13] H. Chen, F. Li, and Y. Wang, "Echotrack: Acoustic device-free hand tracking on smart phones," in *IEEE INFOCOM 2017 - IEEE Conf. Comput. Commun.*, pp. 1-9, Atlanta, GA, USA, May 2017.

[14] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. MobiCom '16*, pp. 82-94, New York, NY, USA, Oct. 2016.

[15] J. Kim and S. Choi, "Hand gesture classification based on nonaudible sound using convolutional neural network," *J. Sensors,* vol. 2019, pp. 1-9, Nov. 2019.

[16] S.-M. Yang, W.-J. Song, I.-S. Choi, and S.-J. Yoo, "Implementation of deep learning-based motion classification system for IoT device control in ultrasonic sound environments," *J. KICS,* vol. 42, no. 9, pp. 1796-1805, Sep. 2017.

[17] H. Cheng, L. Yang, and Z. Liu, "Survey on 3D hand gesture recognition," in *IEEE Trans. Cir. Syst. for Video Technol.*, vol. 26, no. 9, pp. 1659-1673, Sep. 2016.

[18] Y. Lavinia, H. H. Vo, and A. Verma, "Fusion based deep CNN for improved large-scale image action recognition," in *2016 IEEE ISM*, pp. 609-614, 2016.

[19] F. M. Noori, M. Riegler, Md Z. Uddin, and J. Torresen, "Human activity recognition from multiple sensors data using multifusion representations and CNNs," *ACM Trans.*

Multimedia Comput. Commun. Appl., vol. 16, no. 2, May 2020.

천 진 원 (Jinwon Cheon)

2019년 2월 : 국민대학교 전자공학부 졸업
2019년 3월~현재 : 국민대학교 보안-스마트 전기자동차학과 석사과정
<관심분야> 기계학습, 사물인터넷, 임베디드 시스템
[ORCID:0000-0002-1506-6745]

최 선 웅 (Sunwoong Choi)

1998년 2월 : 서울대학교 전산과학과 졸업
2000년 2월 : 서울대학교 전산과학과 석사
2005년 8월 : 서울대학교 전기, 컴퓨터공학과 박사
2007년 3월~현재 : 국민대학교 전자공학과 정교수
<관심분야> 유무선 네트워크, 기계학습, 사물인터넷
[ORCID:0000-0002-8719-8181]