

Visualization of Signals with Variable Length for Character Recognition

문자 인식을 위한 가변 길이 신호의 시각화

Sang-Yoon Lee¹, Jung-Ha Hwang², Doo-Hyun Choi³

이상윤¹, 황정하², 최두현³

¹ Student, School of Electronic and Electrical Engineering, Kyungpook National University, South Korea, alexlcy0315@naver.com

² Student, School of Electronic and Electrical Engineering, Kyungpook National University, South Korea, comand4812@gmail.com

³ Professor, School of Electronic and Electrical Engineering, Kyungpook National University, South Korea, dhc@ee.knu.ac.kr

Corresponding author: Doo-Hyun Choi

Abstract: Research on human motion recognition using various sensors has been steadily conducted. However, despite the increasing spread of smartphones and the activation of research using smartphone sensors, research on character recognition methods using signals with variable length is still at a standstill. The reason is that even with the same motion, the length of the signal is always different, making it difficult to process. To solve the problem, the study proposes a method of recognizing characters using an object detection neural network as a classifier. In this study, five different subjects collect data by drawing three alphabets, I, S, and Z letters 100 times each in the air and the collected data is used to train an object detection neural network and to evaluate the performance of the network. The data changes into images through denoising and normalization. Third-order spline interpolation and Fourier transform are used to remove noise from the raw signal. The values of the acceleration sensors x, y, and z axes correspond to the values of the R, G, and B channels of an image, respectively. As a result of evaluating character recognition performance by applying image data to YOLOv5, it showed an average accuracy of 99% for three characters.

Keywords: Variable Length, Visualization, Character Recognition, Deep learning, YOLOv5

요약: 다양한 센서를 이용한 사람의 동작인식 연구는 꾸준히 진행되고 있다. 하지만 스마트폰 보급률 증가와 스마트폰 센서 활용 연구의 활성화에도 불구하고 가변 길이 신호를 활용해 문자를 인식하는 방법에 대한 연구는 진전이 없다. 그 이유는 같은 동작도 신호의 길이가 항상 달라 이것을 다루는 것이 어렵기 때문이다. 이 문제를 해결하기 위해 객체 탐지 신경망을 분류기로 사용하여 문자를 인식하는 방법을 제안한다. 본 연구에서는 다섯명의 피실험자가 I, S, Z 세가지 알파벳을 허공에 백 번씩 그려서 데이터를 수집하였으며, 수집된 데이터는 객체 탐지 신경망의 학습 및 평가에 활용되었다. 데이터는 잡음제거와 정규화를 통해 이미지 데이터로 재탄생되었다. 신호의 잡음을 제거하기 위해 3차 스플라인 보간법과 푸리에 변환을 사용하였다. 가속도 센서 x, y, z축의 값은 각각 이미지의 R, G, B 채널에

Received: February 11, 2023; 1st Review Result: March 27, 2023; 2nd Review Result: April 25, 2023
Accepted: May 31, 2023

대응된다. 이미지 데이터를 YOLOv5에 적용하여 문자 인식 성능을 평가한 결과 세가지 문자에 대해서 평균 99%의 정확도를 보인다.

핵심어: 가변 길이, 시각화, 문자 인식, 심층 학습, YOLOv5

1. 서론

개인용 컴퓨터가 세상에 처음 나타난 것은 1970년대로 모든 가정에 개인용 컴퓨터가 한 대씩 들어서기까지는 수십 년이 걸렸다. 하지만 스마트폰은 2000년대에 처음 소개되어 많은 사람이 사용하기까지 채 10년이 걸리지 않았다. 이제 스마트폰은 사람들의 일상생활에서 떼어놓을 수 없는 전자기기가 되었으며, 마치 신체의 일부인 것처럼 항상 우리와 함께 움직인다.

항상 사용자와 함께한다는 점에 착안하여 다양한 생활편의 기능을 제공하고자 기업들은 이전에 사용하지 않았던 다양한 센서들을 스마트폰에 활용하기 시작하였다. 가속도 센서, 자이로 센서, 방향 센서 등이 이에 포함된다. 전 세계 스마트폰 시장에서 독점적인 위치를 확보한 A사는 사용자의 움직임 감지하여 사용자가 낙상 사고 등을 겪게 되면, 기기 스스로 위급상황을 알리는 기능을 통해 실제 어려움에 부딪친 사람을 구한 사례가 있다. 또한, 디지털 헬스케어에 대한 관심이 증가하면서 스마트폰을 이용한 일일 운동량 측정, 생활패턴 분석 등의 생활편의 기능도 제공하기 시작하였다.

이러한 생활편의 기능을 완성하는 데는 다양한 센서를 활용한 연구들이 기여한 것으로 보인다. 예로, Akram Bayat 등은 스마트폰에 내장된 가속도 센서를 이용하여 춤추기, 느리게 걷기, 빠르게 걷기, 달리기, 계단 오르기, 계단 내려가기 6가지 동작을 구분하는 방법을 제안하였다[1]. Songbin Xu 등은 웨어러블 디바이스를 팔에 부착하고 허공에 가상의 칠판이 있다고 생각하며 팔을 움직여 허공에 문자와 숫자를 쓸 때 발생한 신호를 이용하여 알파벳 대문자와 0부터 9까지 아라비아 숫자를 인식하는 방법을 제안하였다[2]. Zhiyuan Lu 등은 가속도 센서와 표면 근전도 센서를 사용하는 자체 제작한 웨어러블 디바이스를 팔에 부착하고 주먹을 편 상태에서 팔을 상, 하, 좌, 우 직선으로 움직이거나 원을 그리는 5가지 손동작과 주먹을 쥔 상태에서 한 획 그리기로 허공에 0부터 9까지 숫자를 쓸 때 발생하는 신호를 이용하여 손동작과 숫자를 구분하는 방법을 제안하였다[3].

이와 같은 연구들은 센서를 통해 입력된 신호를 바탕으로 사용자의 움직임 또는 표현하고자 하는 문자와 숫자를 구분하기 위해 신호에서 특징을 추출하는 과정을 거친다. 이 과정에서 많은 연구는 서로 다른 길이의 신호를 특정한 길이로 가공하거나 신호로부터 추출하는 특징의 수를 연구자가 지정한 수로 제한한다[3-15]. 이에 따라서 원래 신호가 가지고 있는 정보가 소실되거나 왜곡될 수 있는 문제가 있다. 그러므로 신호가 가진 정보의 손실을 줄이기 위해서 가변 길이 신호를 있는 그대로 활용하여 특징을 추출하고 이를 통해 사용자의 움직임이나 표현하고자 하는 문자, 숫자를 구분하는 방법이 필요하다.

본 연구는 가변 길이 신호의 길이를 임의로 특정하거나 신호로부터 특정한 개수의 특징을 추출하는 방법을 적용하지 않고 신호를 있는 그대로 활용하여 이미지로 바꾸고 객체 탐지 신경망의 입력으로 활용하여 학습된 신경망의 분류 결과를 통해 신호가 나타내는 문자가 무엇인지 인식하는 방법을 다룬다. 최근 유행하고 있는 신경망 모델 중

객체 탐지를 목적으로 하는 신경망을 활용하였으며, 객체 탐지 신경망에 가변 길이 신호를 입력으로 사용하기 위해 센서 입력 신호를 이미지로 변환하는 방법을 제안한다. 실험 결과 YOLOv5 모델을 사용하여 ‘I’, ‘S’, ‘Z’ 3가지 알파벳을 99%의 정확도로 판별하였다.

본 연구 논문의 구성은 다음과 같다. 2장에서 본 연구와 관련있는 다른 연구를 살펴봄으로써 각 연구에 사용된 다양한 기법과 특징을 파악하고 장점과 단점을 확인한다. 3장에서는 본 연구에 사용한 데이터의 수집방법, 사용된 데이터 전처리 기법 및 실험 환경을 설명한다. 4장에서는 실험을 통해 얻은 결과를 제시하고 결과에 대한 분석을 통해 연구에서 제시한 가설이 맞는지 확인하고, 5장에서 연구에 대한 결론을 내린다.

2. 관련 연구

센서를 이용하여 동작을 인식하는 연구는 꾸준히 관심을 받았다. 다양한 동작 인식 연구를 구체적인 연구 내용으로 분류하자면 사람의 활동 인식, 손짓 인식, 문자 인식 3가지 형태로 구분이 가능하다. 이 중에서 가장 연구가 활발한 분야는 활동 인식 분야이지만, 문자 인식에 대한 연구도 활발하게 이루어지고 있다. 본 연구는 ‘I’, ‘S’, ‘Z’ 3가지 문자에 대한 연구라는 제한이 있지만, 최종 목표는 26개 알파벳 전체를 인식하는 것이다. 이러한 관점에서 손의 움직임과 관련성이 높은 문자 인식과 손짓 인식에 대한 연구를 집중적으로 살펴본다.

손짓 인식에 대한 연구를 요약하면 다음과 같다. Zhiyuan Lu 등은 자체 제작한 기기를 이용하여 데이터를 수집하고 베이스 선형 분류기와 히든 마르코프 모델을 이용해 여러 손짓과 숫자를 인식하는 방법을 제안하였다[3]. Fuad Al Abir 등은 보간법을 이용해 신호의 특성을 추출하는 방법을 이용하였으며 손짓과 문자에 대한 공개된 데이터를 사용함으로써 제안한 방법의 성능을 객관적으로 비교할 수 있도록 하였다[4]. Kazuya Murao 등은 27가지 손짓을 규정하고 9개의 센서를 하나의 보드에 결합하여 데이터를 수집함으로써 신호 분류의 정확성을 높이고자 하였다[5]. Xian Wang 등은 스마트폰이 움직이는 방법에 따라 6가지 손짓을 정하고 스마트폰에 내장된 센서를 통해 수집된 데이터를 패턴분석으로 분류하였다[6]. Hyunju Cho 등은 스마트폰을 가상의 테니스 채로 가정하고 스마트폰 내장 센서에서 수집한 데이터에서 패턴을 분석하여 테니스 공을 치는 네 가지 동작을 구분하는 방법을 제안하였다[7]. Tsung-Ming Tai 등은 하나의 손짓을 인식하는 기존의 연구들과 달리 연속된 손짓을 분류하는 방법을 제시하였다[16].

문자 인식에 대한 연구들은 다음과 같다. Yafeng Yin 등은 랜덤 포레스트 기법을 적용하여 26개의 소문자 알파벳을 구분하는 방법을 제안하였다[8]. Tomer Yanay 등은 K-최근접 이웃 방법을 이용해 26개 대문자 알파벳을 구분하는 방법을 제안하였다[9]. Yunzhe Li 등은 합성곱 신경망(Convolutional Neural Network)을 이용하여 0~9까지 숫자를 구분하는 방법을 소개하였다[10]. Chunyu Xie 등은 A~D, 1~4를 서포트 벡터 머신, 히든 마르코프 모델, K-최근접 이웃 방법을 적용하여 분류하였으며 3가지 분류 방법의 성능을 비교하였다[11]. Pedro Lopez-Rodriguez 등은 합성곱 신경망과 장단기 메모리(Long Short-Term Memory)를 활용하여 터치스크린에 26개의 소문자 알파벳과 0~9의 숫자를 썼을 때 이것을 분별하는 방법을 제안하였다[17]. Amitangshu Pal은 입력 신호의 패턴을 분석하고 획 구분 방법을 통해 입력된 문자를 구분하는 방법을 제안하였다[12]. Sandip Agrawal 등은 [12]의 연구와 비슷하게 패턴을 분석하고 획 구분을 통해 문자를 인식하는 방법을

제안하였으나 연속된 문자를 인식하기 위한 노력이 있었다는 점에서 차이가 있다[13]. Qian Zhang 등이 제안한 방법은 엄밀하게 따졌을 때 유일하게 연속된 문자를 인식하는 방법으로 보인다. 스마트워치를 손목에 찬 상태에서 빈 종이 등에 글을 쓰면 연속으로 입력되는 문자를 실시간으로 인식하여 보여주었다[14]. Zhangjie Fu 등은 사람의 손짓에 의해 발생하는 와이파이 간섭 신호를 분석하여 문자를 인식하는 방법을 제안하였는데, 사람이 어떤 장치를 몸에 지니지 않아도 된다는 점이 가장 큰 특징이다[15].

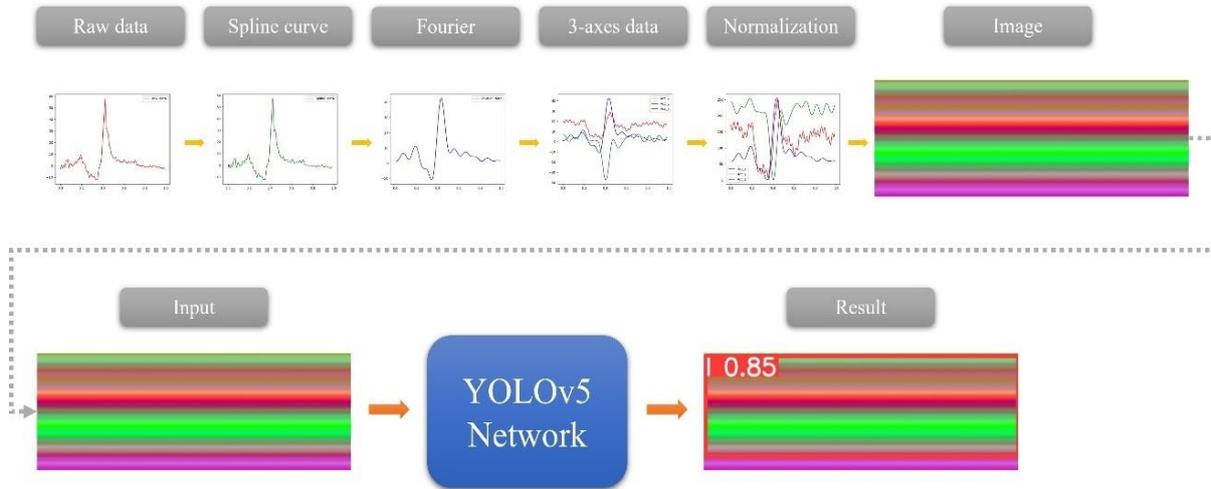
이처럼 다양한 연구가 진행되었으나 여전히 해결되지 못한 문제가 남아있다. 그것은 바로 신호 길이의 가변성에 대한 대응 방법이다. 일부 연구들은 실험 데이터를 기반으로 가장 효과적인 것으로 판단되는 길이로 전체 데이터의 길이를 특정하는 방법을 사용하였으며[3-5][8-11][14], 다른 연구들은 서로 다른 길이의 신호를 특정 개수의 구간으로 분류하거나[2] 특정 패턴을 찾아냄으로써 신호의 길이를 정규화하는 방법을 적용하였고[6][7][12][13][15], 나머지는 신호 길이에 대한 대응 방법이 명확하게 나타나지 않는다[16][17]. 실증적 방법으로 데이터 길이를 특정하거나 신호에서 특정 패턴을 추출함으로써 처리해야 하는 데이터의 길이를 정규화하는 방법은 데이터의 왜곡과 손실 등이 발생할 가능성이 높으며, 데이터를 생성하는 피실험자의 개인차에 따라 데이터의 특성 변화가 크다는 점을 감안하면 좋은 방법이라고 판단하기 어렵다.

또한, 신호를 시각화하는 기존 연구 중에서 본 연구와 유사한 방법으로는 음성신호 분석에서 주로 활용되는 스펙트로그램 기법이 있다[18][19]. 스펙트로그램 기법은 푸리에 변환을 통해 신호의 주파수와 크기를 분석한 후 시간과 주파수를 좌표로 사용하는 2차원 평면에서 각 좌표상의 점을 신호 크기에 따라 빛의 강도를 다르게 하여 이미지로 표현하는 방법이다. 이 방법은 신호를 분석하는데 매우 효과적인 방법으로 오래전부터 사용되어 왔으며, 신호의 시각화라는 점에서는 본 연구에서 제안하는 방식과 유사한 부분이 있다. 스펙트로그램 기법은 신호의 주파수와 시간은 좌표가 되고, 신호의 크기가 영상의 값이 된다. 또한 스펙트로그램 기법은 앞의 다른 연구들과 마찬가지로 사용자 임의에 따라 특정된 길이의 신호가 분석 데이터로 사용된다.

이에 따라, 본 연구에서는 가변 길이 신호를 특정 길이의 신호로 변환하거나 패턴을 분석하여 정규화하는 방법과는 다르게 신호를 있는 그대로 사용하는 방법을 제안한다. 또한, 기존 연구들에서 주로 활용되는 신호를 시각화하는 기법인 스펙트로그램을 이용하는 방식과는 다른 신호의 시각화 방법을 제안한다. 이를 위해 본 연구에서는 가속도 센서의 x, y, z축의 값을 3채널 영상의 R, G, B값으로 대응시켜 신호를 영상으로 변환하고 신호의 길이를 제한하지 않음으로써 신호가 가지고 있는 정보의 손실이나 왜곡을 방지한다. 이미지화된 신호를 객체 탐지 신경망의 학습에 활용하고 학습된 신경망을 분류기로 이용해 가변 길이 신호를 정확하게 분류하는 방법을 제안한다. 실험을 위해 사용된 객체 탐지 신경망은 YOLOv5[20] 모델이다.

3. 실험 방법

실험 방법을 명확하게 제시하기 위해서 데이터 정보, 신호 전처리 기법, 신호의 이미지화 방법, YOLOv5 신경망의 환경설정에 대해서 아래에서 명확하게 기술한다. 아래 [그림 1]은 본 연구에서 문자를 인식하는 절차를 도식화한 것이다.



[그림 1] 데이터 전처리 및 문자 인식 절차

[Fig. 1] Data Preprocessing and Character Recognition Process

3.1 데이터 정보

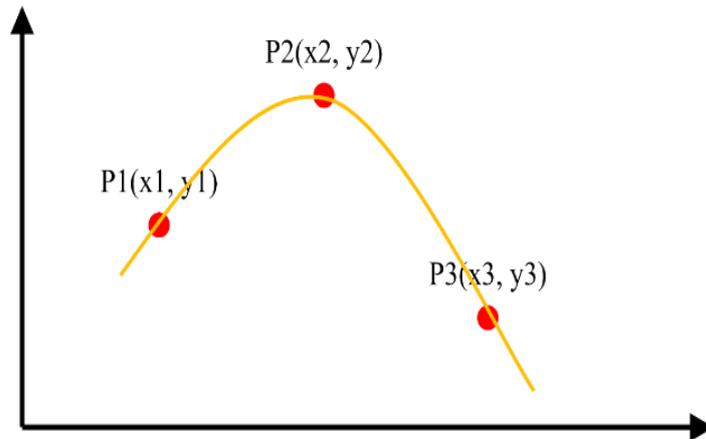
3가지 문자 ‘I’, ‘S’, ‘Z’ 데이터를 수집하기 위해 서로 다른 5명이 각 문자에 대해 100개씩 데이터를 수집하여 전체 1,500개의 데이터를 수집하였다. 수집된 데이터는 전처리를 통해 이미지로 바꾸어 사용되었으며, 전체 1,500개의 데이터 중 300개는 학습된 신경망의 최종 분류 성능을 검증하기 위해 사용되고, 나머지 1,200개의 데이터 중 960개는 신경망 학습을 위해서 사용되고, 240개는 학습된 신경망의 사전평가를 위해 사용되었다. 전체 데이터는 임의의 선택에 따라 용도별로 구분되었으며, 데이터를 수집할 때 피실험자 앞에 가상의 칠판이 있다고 생각하고 분필로 알파벳을 그리는 것처럼 문자를 허공에 그리도록 함으로써 문자 쓰기 제한사항을 최소화하여 개인차를 최대한 반영하였다. 데이터 수집에 사용된 기기는 삼성전자에서 제조한 안드로이드 버전 12의 Galaxy A31 모델이며, 본 논문에서는 스마트폰 내장 센서 중 가속도 센서의 데이터만을 이용한다. 내장된 가속도 센서를 통해 측정할 수 있는 신호는 최소 간격이 10ms로 1초 동안 최대 100개의 신호를 측정할 수 있다.

3.2 신호 전처리

가속도 센서로 측정된 데이터는 x, y, z 축 3개의 좌표로 구성된다. 원본 데이터는 잡음이 섞여 있으므로 이를 제거하기 위한 절차가 필요하다. 이 연구에서는 3차 스플라인 보간법 및 푸리에 변환을 통해 잡음을 제거하였다. 푸리에 변환을 통해 원래의 신호를 복원하기 위해서는 최소한 주파수의 2배 크기로 샘플링을 해야 한다. 가속도 센서를 통해 측정된 신호는 주파수가 100Hz이므로 푸리에 변환을 사용하기 위해서는 적어도 200Hz의 속도로 샘플링해야 한다. 따라서, 푸리에 변환을 적용하기 위해서는 드러나지 않는 센서 데이터를 추정해야 한다. 그러므로 많이 활용되는 기법의 하나인 3차 스플라인 보간법을 통해 데이터를 보정하였다.

3차 스플라인 보간법은 떨어진 서로 다른 두 점 사이가 3차 함수의 곡선으로 이어져 있다고 추정하는 방법이다. 다음의 그림 2와 같이 서로 다른 점 3개가 그래프상에

분포하고 있으며, 점의 정확한 위치를 우리가 안다면, 다음의 조건에 따라 우리는 서로 다른 두 점 사이의 3차 함수를 추정할 수 있다.



[그림 2] 3차 스플라인 보간법 이해를 위한 그림

[Fig. 2] A Figure for Understanding of Third Order Spline Interpolation

[그림 2]에서 점 P1과 P2 사이를 지나는 3차 함수 $f_1(x)$ 과 점 P2와 P3 사이를 지나는 3차 함수 $f_2(x)$ 는 다음의 식(1), (2)와 같이 표현할 수 있다.

$$f_1(x) = a(x - x_1)^3 + b(x - x_1)^2 + c(x - x_1) + d \quad (1)$$

$$f_2(x) = e(x - x_2)^3 + f(x - x_2)^2 + g(x - x_2) + h \quad (2)$$

또한, 3차 스플라인 보간법에서 각 함수는 모든 점에서 연속이며, 도함수도 모든 점에서 연속이고, 도함수의 도함수도 모든 점에서 연속이라는 전제를 가진다. f_1 과 f_2 의 도함수는 식(3), (4)와 같고 이 도함수들의 도함수는 식(5), (6)과 같다.

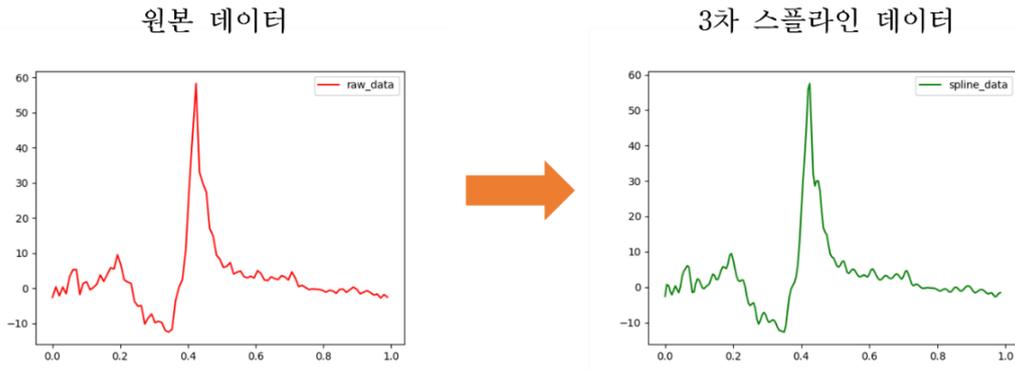
$$f_1(x)' = 3a(x - x_1)^2 + 2b(x - x_1) + c \quad (3)$$

$$f_2(x)' = 3e(x - x_2)^2 + 2f(x - x_2) + g \quad (4)$$

$$f_1(x)'' = 6a(x - x_1) + 2b \quad (5)$$

$$f_2(x)'' = 6e(x - x_2) + 2f \quad (6)$$

따라서 $f_1(x_1)=y_1$, $f_1(x_2)=y_2$, $f_2(x_2)=y_2$, $f_2(x_3)=y_3$, $f_1(x_2)'=f_2(x_2)'$, $f_1(x_2)''=f_2(x_2)''$ 가 성립한다. 위의 6개 식만으로는 8개의 미지수를 구할 수 없다. 그러나 3차 스플라인 보간법에서는 하나의 전제가 더 존재하는데 스플라인 함수의 2차 도함수는 양 끝점에서 값이 0이라는 것이다. 이 전제를 따르면 $f_1(x_1)''=f_2(x_3)''=0$ 이 되어 8개의 식이 존재하므로 8개 미지수를 전부 구할 수 있다. 원본 데이터와 3차 스플라인을 통해 추정된 데이터는 [그림 3]과 같다.



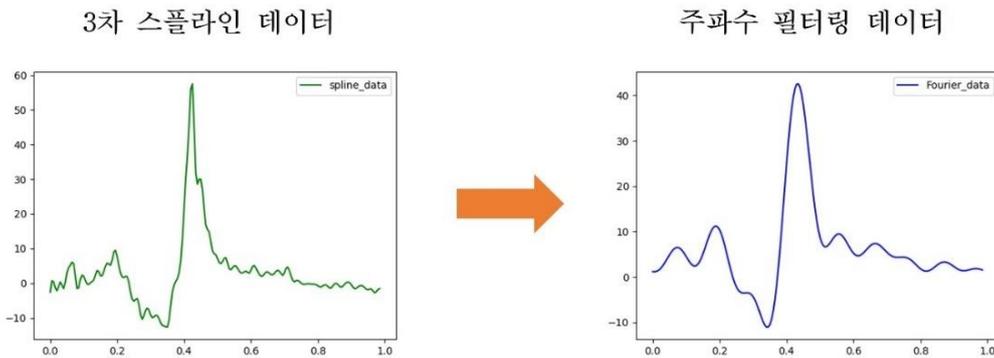
[그림 3] 3차 스플라인 보간법 적용 전후 데이터 비교

[Fig. 3] Data Comparison Before and After Third Order Spline Interpolation

푸리에 변환은 시간 대역의 신호를 주파수대역의 신호로 변환하는 방법이다. 이 방법을 사용하면 우리는 시간 대역의 신호에 큰 영향을 주는 주파수와 작은 영향을 주는 주파수를 구분할 수 있다. 그뿐만 아니라 특정 주파수를 제거하거나 강화할 수도 있다. 이 연구에서는 푸리에 변환을 통해 신호에 가장 큰 영향을 주는 상위 10개 주파수만을 복원하여 잡음을 제거하는 방식을 선택하였다. 시간 대역에서 주파수대역으로 바꾸는 푸리에 변환은 식(7)을 따르고, 역변환은 식(8)을 따른다.

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \tag{7}$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{i\omega t} d\omega \tag{8}$$



[그림 4] 주파수 필터링 전후 데이터 비교

[Fig. 4] Data Comparison Before and After Frequency Filtering

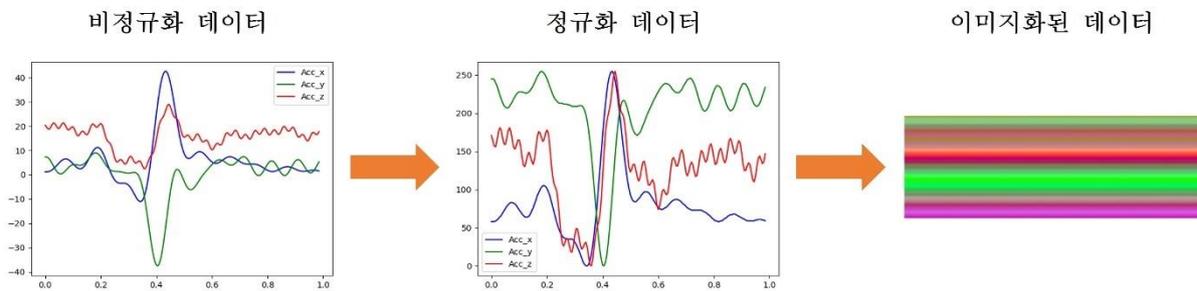
푸리에 변환을 통해 잡음을 제거한 신호가 훨씬 매끄러워진 것을 [그림 4]를 통해 확인할 수 있다. 가속도 센서 신호는 사람의 움직임에 따라 측정된 신호이므로 푸리에 변환을 통해 신호의 크기가 큰 상위 10개 주파수를 살펴보면 대부분 10Hz 이하의 저주파에 속하는 것을 확인할 수 있다.

3.3 가속도 센서 신호의 이미지화

신호 전처리를 통해 잡음을 제거한 가속도 센서 신호를 객체 탐지 신경망에 적용하기 위해서는 시간 순서 데이터를 이미지로 바꾸어야 한다. 이미지는 R, G, B 3개 채널을 가지고 채널마다 0~255까지의 픽셀 값을 가진다. 따라서 가속도 센서 신호에서 특징을 추출하기 위해서는 센서 데이터를 0~255까지 값으로 정규화해야 한다. 가속도 센서는 x, y, z 3개 축에서 값이 측정되므로 x, y, z 축을 각각 R, G, B 채널에 대응시키는 방법을 적용하였다. 정규화 방법은 다음과 같다.

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} * 255 \tag{9}$$

가속도 센서 x축 신호의 집합 $X = [x_1, x_2, \dots, x_n]$ 라고 할 때, X에서 최댓값 $\max(X)$ 와 최솟값 $\min(X)$ 를 찾은 후 식 (9)의 방식으로 정규화하여 정규화된 집합 X'을 구한다. 동일한 방식으로 y축과 z축의 신호도 정규화한다. 세 가지 축 신호를 정규화하고 이를 이미지로 나타내면 데이터는 [그림 5]와 같이 바뀐다. 정규화된 X'의 각 원소는 이미지에서 행의 픽셀값을 나타내고, X'의 원소가 n개인 경우 이미지는 가로 416, 세로 n 픽셀의 이미지가 된다. 가로 픽셀 크기를 416으로 특정한 이유는 YOLOv5 모델이 이미지를 학습할 때 사용하는 이미지 크기가 416*416이기 때문에 학습에 용이하도록 이미지의 가로 길이를 동일하게 맞추었다.



[그림 5] 데이터 정규화 및 이미지화

[Fig. 5] Data Normalization and Visualization

3.4 실험 환경

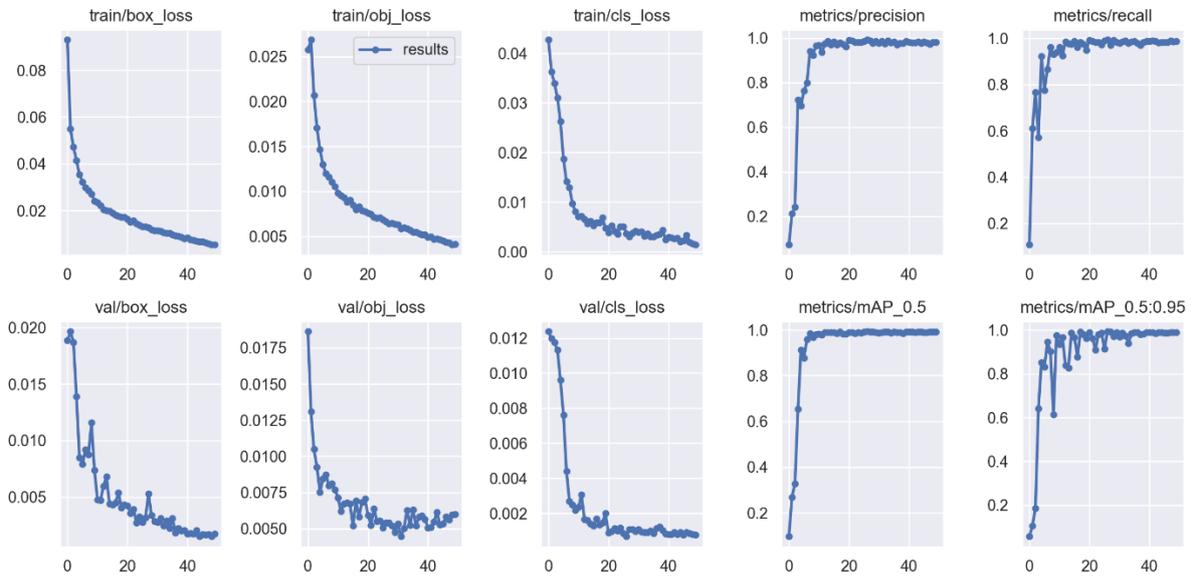
신호 전처리를 통해 이미지로 바꾼 가속도 센서 데이터 중 일부는 YOLOv5의 학습에 활용되고 나머지는 성능 평가에 활용된다. 실험 환경에 따라 연구 결과에 차이가 있을 수 있으므로 본 연구를 재현하거나 후속 연구를 하고자 하는 경우에 참고할 수 있도록 정확한 실험 환경을 명시하고자 한다. 컴퓨터의 운영체제는 윈도우 10이며, 중앙처리장치는 Intel Xeon E5-2630 v4, 그래픽카드는 NVIDIA Titan XP이다. YOLOv5를 구동하기 위해 필요한 가상환경의 주요 패키지 버전은 cuda 11.7, cudnn8.0, python 3.8, torch 1.13.1이다. 학습된 YOLOv5를 활용할 때 한 가지 주의해야 할 사항이 있다. 이 신경망의 기본 설정을 따르면 이미지를 학습할 때 416*416의 크기로 학습하지만, 학습이 끝난

신경망으로 객체를 탐지할 때는 640*640의 크기로 입력 크기를 재배열한다. 본 연구에서는 객체를 탐지할 때도 416*416의 크기로 입력 크기를 재배열하도록 YOLOv5를 수정하여 사용하였다. 입력 이미지의 가로세로 길이가 다른 경우 비율을 계산하여 긴 쪽이 416픽셀이 되며 다른 쪽은 비율에 맞추어 조정된다. 또한, 이 신경망에는 데이터 증강 기법이 포함되어 있으며, 사용자가 데이터 증강 기법의 적용을 선택할 수 있다. 본 연구에서는 가변 길이 신호를 활용하므로 가변 길이 신호에서도 문자를 잘 인식할 수 있도록 데이터 증강 기법을 적용하였다.

4. 실험 결과

YOLOv5는 신경망의 크기에 따라 n, s, m, l, x의 5가지 버전이 존재하는데, 이번 연구에서는 그중 가장 작은 n 모델을 사용하였다. 학습은 배치 크기 16, 50 에포크로 진행되었으며, 학습 결과를 나타내는 지표는 [그림 6]에 잘 나타나 있다.

YOLOv5 신경망은 학습이 잘 되었는지 판단하기 위한 3가지 손실함수를 사용한다. 손실함수는 객체 크기와 위치가 잘 맞는지 나타내는 box_loss, 객체의 존재 유무를 나타내는 obj_loss, 객체의 클래스를 잘 구분하는지 나타내는 cls_loss가 있다. 지표 중에서 정밀도(precision)는 신경망이 참인 것으로 판단한 모든 것 중 실제 참인 것의 비율을 의미하며, 재현율(recall)은 실제로 참인 것 중에서 신경망이 참인 것으로 판단한 것의 비율을 의미한다. 학습이 진행됨에 따라 손실함수는 값이 줄어들고, 정밀도와 재현율은 높아지는 것을 통해 학습이 잘 진행되었음을 확인할 수 있다.



[그림 6] 신경망 학습에 따른 지표 변화

[Fig. 6] Indicator Changes with Neural Network Training

960개의 학습 데이터와 240개의 사전평가 데이터로 학습한 신경망에 최종 평가 데이터 300개를 입력으로 사용한 문자 인식 결과는 다음의 [표 1]과 같다.

[표 1] 문자 인식 평가 결과

[Table 1] Result of Character Recognition Assessment

문자	샘플수	정밀도	재현율	평균정밀도 (mAP50)	평균정밀도 (mAP50-95)
I	96	1	0.979	0.992	0.986
S	107	0.98	0.991	0.994	0.994
Z	97	0.989	1	0.995	0.995

3가지 문자에 대한 인식 정밀도와 재현율의 평균값은 각각 0.99로 매우 정확도가 높은 것으로 나타났다. 또한, 평균정밀도(mAP50-95)의 수치 또한 0.99의 평균 값을 보이며 신호의 시작과 끝부분을 잘 찾아내는 것으로 확인된다.

또한, 본 연구의 가장 큰 목적인 가변 길이 신호를 얼마나 정확하게 인지하는지에 대해서는 신호의 길이가 얼마나 차이나는지를 통해서 알 수 있다. 데이터로 수집된 전체 1,500개의 신호에서 I, S, Z 각 문자를 표현한 신호의 최대, 최소 길이는 [표 2]와 같다.

[표 2] 각 문자별 신호의 최대 및 최소 길이

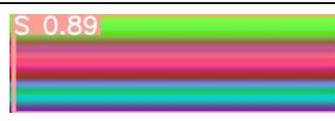
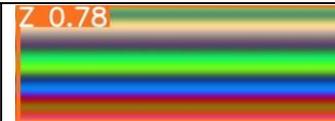
[Table 2] Maximum and Minimum Lengths of Each Character Signal

문자	최대길이	최소길이
I	198	76
S	176	80
Z	272	86

신호의 길이가 약 2배~3배까지 차이 남에도 불구하고 분류 정확도는 99%에 달한다. 이를 통해 본 연구에서 제안한 방법으로 가변 길이 신호를 높은 정확도로 구별할 수 있음이 확인된다. 그러나 여전히 100%의 정확도는 아니라는 점에서 개선의 여지가 있다. [표 3]을 통해서 전체 300개의 평가 데이터 중에서 잘못 인식한 결과를 확인할 수 있다.

[표 3] 문자 오인식 사례

[Table 3] Examples of the Wrong Character Recognition

이미지			
실제문자	I 신호 1개	I 신호 1개	S 신호 1개
인식결과	I 신호 2개	S 신호 1개	Z 신호 1개

문자를 오인식한 경우는 크게 두 가지로 나타났다. 신호를 1개로 인식해야 하지만 2개로 인식한 경우, 신호를 다른 문자로 인식한 경우가 해당한다. 이러한 오인식 문제는 신경망 학습 과정에서 사용되는 데이터 증강 기법에 의해 서로 다른 문자임에도 비슷한 특징을 찾아낼 수 있다는 점에서 발생하는 것으로 추측된다. 향후 연구에서는 이와 같은 문제를 개선하기 위한 노력이 필요하다.

5. 결론

본 연구에서는 사람의 동작을 인식하는 연구 중에서 손동작을 이용한 연구에 초점을 두었으며, 손동작을 이용한 문자 인식, 손짓 인식 등의 연구에서 가변 길이 신호를 처리하는 방법에 대한 연구가 미흡함을 인지하고, 이를 해결할 방안을 찾는 연구를 하였다. 객체 탐지 신경망은 합성곱 연산이 가지는 특성인 참조의 지역성으로 인해 영상에서 객체의 크기와 위치가 바뀌더라도 객체를 찾아낼 수 있다. 이 부분에 근거하여 신호를 영상으로 변환 후 신경망의 입력으로 사용하면 가변 길이 신호를 높은 정확도로 분류할 수 있을 것이라는 가설을 세웠으며, 실험을 통해 객체 탐지 신경망은 시각화된 문자를 표현하는 1차원 신호를 길이 변환 없이 99%의 정확도로 구분할 수 있음을 확인하였다. 실험 결과는 스플라인 보간법을 적용하고 푸리에 변환을 통해 전처리한 신호를 이미지로 바꾸는 방법이 서로 다른 신호를 높은 정확도로 구분할 수 있도록 신호에서 특징을 잘 찾아낸다는 것을 의미하며, 객체 탐지 신경망이 가변 길이 신호를 높은 정확도로 구분할 수 있음을 보여준다. 1%의 확률로 문자를 오인식하는 문제가 남아 있으나 이는 향후 연구를 통해 개선할 수 있을 것으로 보인다. 또한, 가변 길이 신호를 인식할 수 있다는 것은 연속된 신호에서 특정 신호를 찾아낼 수 있는 가능성을 포함하고 있으므로 향후 연구에서는 가변 길이 신호 인식만큼이나 어려운 문제인 연속된 신호를 정확하게 인식하는 방법을 연구할 계획이다.

References

- [1] Akram Bayat, Marc Pomplun, Duc A. Tran, A study on human activity recognition using accelerometer data from smartphones, *Procedia Computer Science*, (2014), Vol.34 pp.450-457.
DOI: <https://doi.org/10.1016/j.procs.2014.07.009>
- [2] Songbin Xu, Yang Xue, Air-Writing Characters Modelling and Recognition on Modified CHMM, 2016 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, (2016)
DOI: <https://doi.org/10.1109/SMC.2016.7844452>
- [3] Zhiyuan Lu, Xiang Chen, Qiang Li, Xu Zhang, Ping Zhou, A Hand Gesture Recognition Framework and Wearable Gesture-Based Interaction Prototype for Mobile Devices, *IEEE Transactions on Human-Machine Systems*, (2014), Vol.44, No.2, pp.293-299.
DOI: <https://doi.org/10.1109/THMS.2014.2302794>
- [4] Fuad Al Abir, Md. Al Siam, Abu Sayeed, Md. Al Mehedi Hasan, Jungpil Shin, Deep Learning Based Air-Writing Recognition with the Choice of Proper Interpolation Technique, *Sensors*, (2021), Vol.21, No.24, p.8407.
DOI: <https://doi.org/10.3390/s21248407>
- [5] Kazuya Murao, Tsutomu Terada, Ai yano, Ryuichi Matsukura, Evaluating Gesture Recognition by Multiple-Sensor-Containing Mobile Devices, 2011 15th Annual International Symposium on Wearable Computers, IEEE, (2011)
DOI: <https://doi.org/10.1109/ISWC.2011.20>
- [6] Xian Wang, Paula Tarrío, Eduardo Metola, Ana M. Bernardos, José R Casar, Gesture recognition using mobile phone's inertial sensors, *Distributed Computing and Artificial Intelligence: 9th International Conference*, Springer Berlin Heidelberg, (2012)
DOI: https://doi.org/10.1007/978-3-642-28765-7_21
- [7] Hyunju Cho, Sangchul Kim, Jinsuk Baek, Paul S. Fisher, Motion Recognition with Smart Phone Embedded 3-Axis Accelerometer Sensor, 2012 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, (2012)
DOI: <https://doi.org/10.1109/ICSMC.2012.6377845>

- [8] Tsung-Ming Tai, Yun-Jie Jhang, Zhen-Wei Liao, Kai-Chung Teng, and Wen-Jyi Hwang, Sensor-Based Continuous Hand Gesture Recognition by Long Short-Term Memory, *IEEE sensors letters*, (2018), Vol.2, No.3, pp.1-4.
DOI: <https://doi.org/10.1109/LESENS.2018.2864963>
- [9] Yafeng Yin, Lei Xie, Tao Gu, Yijia Lu, Sanglu Lu, AirContour: Building Contour-based Model for In-Air Writing Gesture Recognition, *ACM Transactions on Sensor Networks*, (2019), Vol.15, No.4, pp.1-25.
DOI: <https://doi.org/10.1145/3343855>
- [10] Tomer Yanay, Erez Shmueli, Air-writing recognition using smart-bands, *Pervasive and Mobile Computing*, (2020), Vol.66, p.101183.
DOI: <https://doi.org/10.1016/j.pmcj.2020.101183>
- [11] Yunzhe Li, Hui Zheng, He Zhu, Haojun Ai, Xiaowei Dong, Cross-People Mobile-Phone Based Airwriting Character Recognition, 2020 25th International Conference on Pattern Recognition, IEEE, (2021)
DOI: <https://doi.org/10.1109/ICPR48806.2021.9411989>
- [12] Chunyu Xie, Shangzhen Luan, Hainan Wang, Baochang Zhang, Gesture Recognition Benchmark Based on Mobile Phone, *Biometric Recognition: 11th Chinese Conference, CCBR 2016*, Springer International Publishing, (2016)
DOI: https://doi.org/10.1007/978-3-319-46654-5_48
- [13] Pedro Lopez-Rodriguez, Juan Gabriel Avina-Cervantes, Jose Luis Contreras-Hernandez, Rodrigo Correa, Jose Ruiz-Pinales, Handwriting Recognition Based on 3D Accelerometer Data by Deep Learning, *Applied Sciences*, (2022), Vol.12, No.13, p.6707.
DOI: <https://doi.org/10.3390/app12136707>
- [14] Amitangshu Pal, MicaPen: A Pen to Write in Air Using Mica Motes, 2020 16th International Conference on Distributed Computing in Sensor Systems, IEEE, (2020)
DOI: <https://doi.org/10.1109/DCOSS49796.2020.00035>
- [15] Sandip Agrawal, Ionut Constandache, Shravan Gaonkar, Romit Roy Choudhury, Kevin Caves, Frank DeRuyter, Using Mobile Phones to Write in Air, *Proceedings of the 9th international conference on Mobile systems, applications, and services*, Association for Computing Machinery, (2011)
DOI: <https://doi.org/10.1145/1999995.1999998>
- [16] Qian Zhang, Dong Wang, Run Zhao, Yinggang Yu, Jiazhen Jing, Write, Attend and Spell: Streaming End-to-end Free-style Handwriting Recognition Using Smartwatches, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Association for Computing Machinery, (2021),
DOI: <https://doi.org/10.1145/3478100>
- [17] Zhangjie Fu, Jiashuang Xu, Zhuangdi Zhu, Alex X. Liu, Xingming Sun, Writing in the air with WiFi signals for virtual reality devices, *IEEE Transactions on Mobile Computing*, (2018), Vol.18, No.2, pp.473-484.
DOI: <https://doi.org/10.1109/TMC.2018.2831709>
- [18] Szu-Wei Fu, Ting-yao Hu, Yu Tsao, Xugang Lu, Complex Spectrogram Enhancement by Convolutional Neural Network with Multi-metrics Learning, 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing, IEEE, (2017)
DOI: <https://doi.org/10.1109/MLSP.2017.8168119>
- [19] Jonathan Dennis, Huy Dat Tran, Haizhou Li, Spectrogram Image Feature for Sound Event Classification in Mismatched Conditions, *IEEE signal processing letters*, (2010), Vol.18, No.2, pp.130-133.
DOI: <https://doi.org/10.1109/LSP.2010.2100380>
- [20] <https://github.com/ultralytics/yolov5>, Feb 11 (2023)