

FSK RADAR와 Multi-Headed Convolutional Neural Network를 이용한 손동작인식 시스템

양기문, *이성주

반도체시스템공학과 및 지능형드론융합전공, *전자정보통신공학과 및 지능형드론융합전공, 세종대학교
ykm3563@itsoc.sejong.ac.kr, *seongjoo@sejong.ac.kr

Hand Gesture Recognition System Using FSK RADAR and Multi-Headed Convolutional Neural Network

Kimoon Yang, *Seongjoo Lee

Dept. of Semiconductor Systems Engineering and Dept. of Convergence Engineering for Intelligent Drone, Sejong Univ.

*Dept. of Electrical Engineering and Dept. of Convergence Engineering for Intelligent Drone, Sejong Univ.

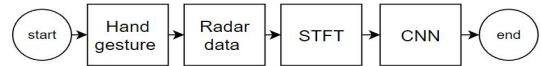
요약

본 논문은 손동작 인식 시스템에 FSK radar와 Multi-headed Convolution Neural Network를 도입해서 인식 정확도를 향상하는 방법을 제안한다. 두 개의 반송파를 교대로 전송하는 FSK radar의 각 샘플포지션에 해당하는 data stream들은 이론적으로 모두 같은 형태를 보여야 한다. 하지만 실제 측정된 결과를 보면 이상적이지 않은 PLL에 의해서 다른 형태를 띤다. 동작 별 data stream들간의 형태 차이의 규칙을 발견하였고, 이런 규칙을 반영하기 위해 Multi-headed CNN 모델을 사용하였다. 결론적으로 98.75%의 인식확률을 확인했고, 기존 방법에 비해 1.88%의 정확도 향상을 얻었다.

I. 서론

본 논문에서는 Frequency Shift Keying (FSK) radar와 multi-headed convolutional neural network (CNN)를 도입한 Hand gesture recognition system (HGR)을 제안한다. 손동작 인식은 Human-Computer Interaction (HCI)에 해당되는 분야로 IoT의 발전과 함께 빠른 속도로 연구가 진행되고 있다.[1] 본 논문에서는 손동작을 통해 디바이스를 효과적으로 컨트롤 하는 것을 목적으로 한다. 손동작은 키보드같은 전통적인 방식에 비해 사용자가 직관적으로 사용할 수 있다는 장점이 있다. 또한 동작인식에 레이더센서를 사용하면, 카메라센서의 사생활 침해 문제와, 자이로센서의 항시 착용문제를 해결할 수 있으므로 효과적이다.[2] 본 논문에서는 두 개의 carrier frequency를 교대로 전송하는 FSK radar를 채택했다. 두 개의 반송파를 교대로 전송할 때 Phase-Locked Loop (PLL)의 한계로 인해 데이터가 일정하지 않게 발생하는 현상이 관찰되었는데, 이 특성과 multi-headed CNN을 이용해서 손동작 인식의 정확도를 향상시켰다. 2장 본문에서 더 자세히 설명한다.

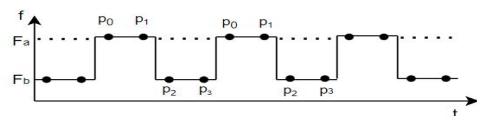
magnitude로 구성된 3차원 그래프인 spectrogram을 얻을 수 있고, spectrogram 이미지는 CNN의 input으로 사용되어 학습 및 추론된다. 최종적으로 CNN을 통해 동작이 구분되고, target device를 컨트롤한다.



<그림 1> 레이더를 이용한 손동작 인식 시스템의 flowchart

2 FSK radar에서의 data stream 분류

FSK radar sensor는 F_a 과 F_b 의 carrier frequency를 갖는 두 반송파를 스위칭하면서 전송한다. 그림2는 FSK radar로부터 얻게되는 데이터의 수신신호 분류 예시를 보여준다. p_0, p_1, p_2, p_3 의 4개의 샘플포지션이 있다고 가정했을 때, 수신된 레이더 신호는 네 개의 data streams로 분류된다. 첫 번째 data stream인 $DS_0[n]$ 은 모든 p_0 포인트에 샘플된다. 마찬가지로 $DS_1[n], DS_2[n], DS_3[n]$ 도 각각 해당하는 포인트들로 구성된다.



<그림 2> FSK radar의 수신신호 data stream 분류

II. 본론

1. 레이더를 이용한 Hand Gesture Recognition system

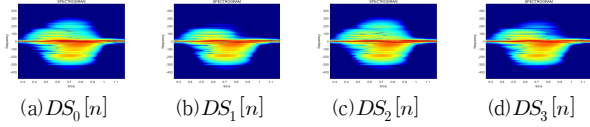
그림1은 레이더를 사용한 손동작 인식 시스템의 flowchart를 나타낸다. 손동작이 발생하면 레이더는 그에 대한 micro doppler signature를 얻는다. micro doppler signature는 움직임으로 인한 주파수 변화에 대해서 구분되는 특성을 보이므로 손동작을 인식하고 구분하는데 이용된다.[3] 무빙윈도우를 통해 짧은 시간동안 FFT를 반복적으로 수행하는 Short Time Fourier Transform (STFT)을 통해 분석된다. 이를 통해 시간의 변화에 따른 주파수 변화패턴을 찾을 수 있다. STFT를 통해 time, frequency,

3. FSK radar data streams간의 spectrogram 차이

본론 2절에서 설명했듯이 F_a 주파수에 해당하는 데이터 스트림인 $DS_0[n], DS_1[n]$ 와 F_b 주파수에 해당하는 $DS_2[n], DS_3[n]$ 을 얻을 수 있다. 각 데이터 스트림에 대한 spectrogram을 구했을 때, 같은 carrier frequency에 해당하는 data stream간의 spectrogram은 같은 형태를 띠어야 한다. 즉 $DS_0[n]$ 과 $DS_1[n]$ 의 spectrogram의 형태가 같아야 하고,

* 교신저자 : 이성주

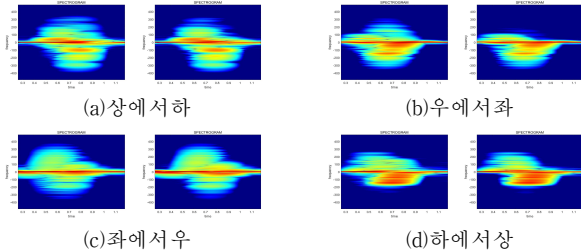
$DS_2[n]$ 과 $DS_3[n]$ 의 spectrogram이 같은 형태여야 한다. 또한 손동작 인식이 사용하는 FSK radar의 두 반송파 주파수의 차이는 매우 작으므로, 네 개의 spectrogram의 형태는 이론적으로 모두 같은 모양이다. 하지만 실제 데이터에서 spectrogram을 도출해보면 그렇지 않다는 것을 관찰할 수 있었다. 그림3은 각 data stream에 해당하는 spectrogram을 보인다.



<그림 3> 실제 데이터의 데이터스트림별 spectrogram

이 현상이 발생하는 원인은 PLL이 이상적이지 않기 때문이다. 이상적으로 두 개의 주파수를 순식간에 스위칭할 수 없으므로 주파수 이동시간이 발생한다. 주파수가 완벽히 안정적인 상태에 샘플포지션이 위치하지 않는다면 이런 현상이 발생할 수 있다. 손을 상, 하, 좌, 우로 움직이는 동작에 대해서 측정을 진행했는데, 모든 동작에서 이런 현상이 발생하였다.

동작 별로 spectrogram을 분석하였다. 그 결과, 동작 별로 해당 현상이 약간씩 다르게 일어나고 있음을 확인하였고, 각 동작 내에서는 일정한 규칙을 가지고 있음을 확인했다. 그림4에 해당 내용이 정리되었다. ‘우에서 좌’와 ‘하에서상’ 동작이 ‘상에서하’,와 ‘좌에서우’ 동작에 비해 상대적으로 이 현상이 크게 일어났다. 그리고 ‘좌에서우’는 좌측하단의 형태 차이가 더 큰 반면에 나머지 동작들을 우측상단의 차이가 더 크다. 그러므로 본 논문은 기존의 방법보다 더 높은 제스처 인식 확률을 얻기위해 이런 규칙들을 인공지능 모델의 학습 및 추론에 이용할 것을 제안하였다.



<그림 4> 각 동작에 대한 spectrogram image 차이

4. Multi-Headed Convolutional Neural Network

하나의 동작에 대해서 두 종류 spectrogram의 특징을 모두 포함하기 위해서 multi-headed CNN을 도입하였다. 학습에 사용한 모델의 구조가 그림5에 정리되어있다. 각각의 spectrogram이 input으로 동시에 사용되고, 각각의 convolutional layer 연산이 수행된다. 레이어 중간에 각각 진행된 레이어가 병합된다. 그 후에 한번의 convolution layer 연산을 더 거친 후 fully connected layer 연산이 수행된다.

Layer (Type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 224, 224, 3)]	0	
input_2 (InputLayer)	[(None, 224, 224, 3)]	0	
conv2d (Conv2D)	(None, 224, 224, 32)	896	input_1[0][0]
conv2d_3 (Conv2D)	(None, 224, 224, 32)	896	input_2[0][0]
max_pooling2d (MaxPooling2D)	(None, 112, 112, 32)	0	conv2d[0][0]
max_pooling2d_3 (MaxPooling2D)	(None, 112, 112, 32)	0	conv2d_3[0][0]
conv2d_1 (Conv2D)	(None, 112, 112, 64)	18496	max_pooling2d[0][0]
conv2d_4 (Conv2D)	(None, 112, 112, 64)	18496	max_pooling2d_3[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 64)	0	conv2d_1[0][0]
max_pooling2d_4 (MaxPooling2D)	(None, 56, 56, 64)	0	conv2d_4[0][0]
conv2d_2 (Conv2D)	(None, 56, 56, 128)	73856	max_pooling2d_1[0][0]
conv2d_5 (Conv2D)	(None, 56, 56, 128)	73856	max_pooling2d_4[0][0]
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 128)	0	conv2d_2[0][0]
max_pooling2d_5 (MaxPooling2D)	(None, 28, 28, 128)	0	conv2d_5[0][0]
concatenate (Concatenate)	(None, 56, 28, 128)	0	max_pooling2d_2[0][0] max_pooling2d_5[0][0]
conv2d_6 (Conv2D)	(None, 56, 28, 128)	147584	concatenate[0][0]
max_pooling2d_6 (MaxPooling2D)	(None, 28, 14, 128)	0	conv2d_6[0][0]
Flatten (Flatten)	(None, 50176)	0	max_pooling2d_6[0][0]
dropout (Dropout)	(None, 50176)	0	Flatten[0][0]
dense (Dense)	(None, 1024)	51381248	dropout[0][0]
dropout_1 (Dropout)	(None, 1024)	0	dense[0][0]
dense_1 (Dense)	(None, 512)	524800	dropout_1[0][0]
dropout_2 (Dropout)	(None, 512)	0	dense_1[0][0]
dense_2 (Dense)	(None, 4)	2052	dropout_2[0][0]

<그림 5> Multi-headed CNN structure

5. 실험 환경

기존 방법과 제안하는 방법에 대해서 각각 실험을 진행하였다. 각 방법에 해당하는 CNN 모델을 학습시켰고, 테스트셋을 통해 인식률을 평가하였다. 이때 기존방식과 제안하는 방식 모두에서 같은 레이더 수신데이터로 학습하고 테스트하였다. 실험에 사용된 세팅은 표 1과 같다.

RADAR model name	K-MC1 radar
Sampling rate	31.25 Kbps
Δf	9 MHz
손동작 종류(4가지)	‘상에서하’, ‘우에서좌’, ‘좌에서우’, ‘하에서상’
F_a, F_b	24.134GHz, 24.125GHz
학습, 테스트 데이터 수	520, 160
측정 참여 인원 수	6명

<표 1> 실험 환경

6. 실험 결과

표2에 실험결과를 정리하였다. 기존 방식에 비해서 논문에서 제안하는 방식의 인식 정확도는 1.88% 향상된 98.75%의 인식률을 보였다.

	기존방식	제안방식	정확도 차이
정확도	96.87 %	98.75 %	1.88 %

<표 2> 실험 결과

III. 결론

본 논문에서는 FSK radar를 적용한 손동작 인식 시스템을 제안했다. FSK radar의 각 data stream으로부터 얻은 spectrogram들의 차이점과 규칙을 이용해서 동작 인식확률을 향상시키는 기법을 제안했다. 두 종류의 spectrogram간의 특징과 규칙을 효과적으로 반영하기 위해서 multi-headed CNN 모델을 구성했다. 성능을 시험하기 위해 같은 레이더 수신 신호로 구성된 training set과 test set을 측정하였고, 기존방식과 제안방식의 학습 및 추론에 사용했다. 테스트 결과로 98.75%의 인식 정확도를 얻었고, 이는 기존방식 대비 1.88% 향상된 정확도이다. 결론적으로 각 동작별 스펙트로그램의 차이에 대한 패턴이 실제 학습에 유의미한 영향을 미쳤음을 확인했다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부의 재원으로 한국연구재단, 무인이동체원천기술개발사업단의 지원을 받은 무인이동체원천기술개발사업(No. 2023M3C1C1A01098414) 및 정부(교육부)의 재원으로 한국연구재단의 이공분야 대학중점연구지원사업의 지원(No. 2020R1A6A1A03038540)을 받아 수행하였으며, 검증을 위한 EDA 관련 툴은 IDEC의 지원을 받았음.

참고 문헌

- [1] G. Paravati and V. Gatteschi, “Human-Computer Interaction in Smart Environments,” *Sensors*, vol. 15, no. 8, pp. 19487 - 19494, Aug. 2015.
- [2] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, “Hand Gestures Recognition Using Radar Sensors for Human-Computer-Interaction: A Review,” *Remote Sensing*, vol. 13, no. 3, pp. 527, Jan. 2021.
- [3] Y. Kim, B. Toomajian, “Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network,” *IEEE Access*, vol. 4, pp. 7125-7130, 2016