Real-time detection of the change of human motion by analyzing millimeter-wave Doppler radar signals using deep learning techniques

Chien-Hung Lai
Department of Electronic Engineerning
National Taipei University of Technology
Taipei, Taiwan
laisan86@gmail.com

Yuh-Shyan Hwang
Department of Electronic Engineerning
National Taipei University of Technology
Taipei, Taiwan
yshwang@ntut.edu.tw

Sheng-Long Kao
Department of Transportation Science
National Taiwan Ocean University
Keelung, Taiwan
slkao@mail.ntou.edu.tw

***Abstract*—This paper presents a system based on a millimeter-wave radar module. After detecting the change of human motion, the changes of the point cloud obtained by analyzing the Doppler signal, and then the change of human motion is classified in real time through deep learning (DL) techniques, including long short-term memory (LSTM) and 1D time distributed convolutional neural network (CNN) two methods. The result did not only consider temporal continuity and, but also scalability.**

Measuring 100 mm wide, 40.8 mm long, and 52.8 mm high, this millimeter-wave radar module features Frequency Modulated Continuous Wave (FMCW) in the 60 to 64 GHz frequency range, has 3 transmit 15dBm and 4 receive 14dB antennas on a package (AOP), 120° azimuth field of view (FoV) and 40° elevation FoV. Additional 5V/2A DC power supply is required during operation, and 1843200bps communication is required through the USB serial port.

Keywords—FMCW, Doppler signals, point cloud, LSTM, CNN.

I. Introduction

The changes in human movements have received a great response in many aspects at this stage, such as analyze limb changes through machine vision. But whether technology will violate personal privacy has always been an issue of debate. In many rehabilitation settings, it is likely that the patients must perform certain movements, however these movements can be very insidious. For this reason, millimeter-wave radar is a good choice if we want to penetrate the medical curtain and also can distinguish changes in human motions.

In previous studies, classification methods basically relied on extracting features from acquired spectrograms after sampling through the speed, bandwidth, and stride of human motions, and these feature images contained signals of time-varying frequencies [1-3]. However, these methods identify the extracted features rather than the entire spectrum image, such as the methods mentioned in [4,5], which use the spectrum images to classify through CNN. These methods are organized, considering that the point cloud obtained by the Doppler signal in space will be affected by the difference in the actual environment, so each point will be affected accordingly. In addition, the human motion is continuous, which means that the change of each point obtained by the Doppler signal will also change with the movement time. Therefore, simply using one algorithm may limit the research results.

In [5], it is discussed whether to consider Constant False Positive Rate (CFAR) detection. The spectral changes of various parts of the human body are sampled without CFAR detection, by the pre-defined number of chirps and pulse repetition intervals to ensure Doppler resolution are sufficient to detect human motion behavior. However, compared with the effect of CFAR detection, the relative amount of data to be processed is too large to be easily distinguished in real time. Therefore, a series of consecutive Doppler frames are used in [5], which are then buffered and stored on the host as input for CNN to obtain features, so real-time motion detection and classification and be achieved.

In comparison, [1-3] does not use CNN to obtain features, but both [1-3] and [5] consider that human motion are continuous. Both [4] and [5] pass through CNN, because CNN has the characteristics of sharing kernels, the pooling layer operations can reduce the number of neurons. Compared with [1-3], there are fewer neuron connections and weight calculations, so CNN is more suitable for the identification of spectral images in terms of scalability and computational efficiency.

Based on [1-5], this paper proposes an improvement. First, using the LSTM method to analyze the Doppler signal. Thus, it is not necessary to buffer many spectrograms for processing like [5], and it also considers the temporal continuity of human motions like [1-3,5], and then use the characteristics of CNN like [4,5] to improve the efficiency of distinguishing changes in human motions. Such a concept has indeed been corroborated by articles such as [6,7].

II. Design

This section will first describe the basic hardware and firmware composition of the millimeter-wave radar module to facilitate understanding of the tools in this article, and then introduce the planning considerations for the two neural networks, LSTM and CNN, in the results of this article.

A. Millimeter-wave radar

In the design of the transmission mechanism of radio wave radar, devices are currently distinguished by different waveforms, such as pulse, continuous wave (CW), Frequency-Shift Keying (FSK) and Frequency-Modulated Continuous-Wave (FMCW), which is an improved application of CW radar.

At present, most of the radars with radio waves as the medium are mainly CW and FMCW. Both radars are easy to distinguish moving targets and suitable for detecting a single moving target. After comparing the two types of CW radars, the FMCW radar can change the frequency of the electromagnetic wave during the measurement operations.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

However, the frequency of the electromagnetic wave of the CW radar is fixed, it will make accurate timekeeping difficult without adding other mechanisms. For instance, a precise periodic rotation mechanism can be provided to change the phase. Due to the start and end periods of transmission and reception must be known. For the requirement of sensing the target object in FoV, the CW radar lacks the utilization period to calculate the time difference with no additional mechanisms help. So, based on the above analysis, this paper decided to use FMCW radar.

B. LSTM

The LSTM method is used in [6,7], which is actually one of the main improvements of this paper. The abstract of [6] mentions that CNN is used for deep feature extraction and LSTM is used for detection using the extracted features. [7] mentioned that gesture recognition is formulated as an irregular sequence recognition problem to capture long-term spatial correlations between point cloud sequences.

In fact, the method advocated by [7] is not only applicable to gestures, but also to the movement of the daily human body. For the millimeter-wave Doppler radar, the results of sensing in space will lead to changes in the point cloud. The principle is that there is relative motion between the radar and the human body, and the frequency measured on the human body is not the same as the frequency emitted by the radar.

If the radar emits a wave with frequency f_o, period T_o, wavelength λ_o, and wave speed V_o, the frequency measured by the human body is f and the motion speed v. If the human body is stationary, the wave number received in time t is $\frac{V_o \cdot t}{\lambda_o}$.

When the radar remains stationary, if the human body moves close to the radar at the speed of V, so at the side of human body will receive waves within the time t, the frequency f_{mov} measured by the side of human body can be expressed as (1).

$$f_{mov} = \frac{\frac{V_o \cdot t}{\lambda_o} + \frac{V \cdot t}{\lambda_o}}{t} = \frac{V_o + V}{V_o} f_o \tag{1}$$

From (1), the frequency measured at the human body side becomes higher. In the same way, when the human body moves away from the radar at a speed V, (2) can be obtained; from (1) and (2), in fact, regardless of the movement of any part of the human body, for a stationary radar, the obtained point cloud must be changed.

$$f_{mov} = \frac{\frac{V_o \cdot t}{\lambda_o} - \frac{V \cdot t}{\lambda_o}}{t} = \frac{V_o - V}{V_o} f_o \tag{2}$$

Therefore, in this paper, LSTM analysis is performed immediately after the point cloud data obtained by the millimeter-wave Doppler radar, so that the temporal continuity of human motions can be predicted as in [7].

C. CNN

As can be seen from the previous analysis, the results of this paper are mainly to use the time change of the point cloud in the space to analyze the human motions, so that the motions can be classified. Therefore, the design of the CNN is different from the one mentioned in [5]. After buffering many spectrograms, the spectrograms are classified one by one in a traditional 2D manner.

The method in this paper is changed to perform a 1D convolution operation for the records of each point over time. From a theoretical point of view, the point cloud obtained by the FMCW radar can be represented by the set formed by each point in the 3D space, such as (3).

$$\text{Point Cloud} = \{(x, y, z)^*\} \tag{3}$$

Therefore, when the human motion in the FoV changes, the obtained point cloud, that is, the set represented by (3), will also be different, as shown in Fig. 1.

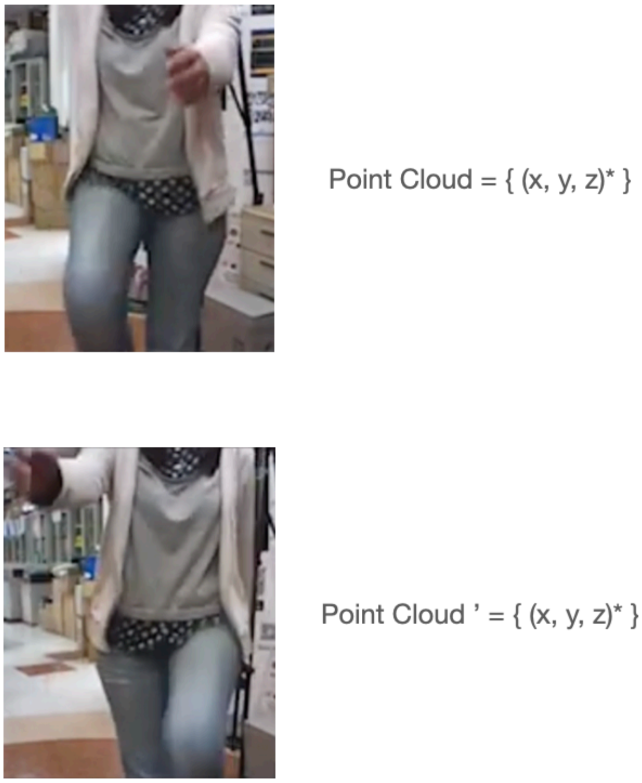


Fig. 1. Human motion variation in FoV makes point cloud different.

The human motion is different with time, which makes the point cloud obtained by the FMCW radar changed accordingly. For example, in this paper, the sampling period is set to 40 ms, so the Doppler effect corresponding to each point on the point cloud can be obtained, in meter per second (m/s). Therefore, the method of this paper is to collect the point cloud obtained over time under human motion and the records formed by the Doppler effect of each point, and then obtain the characteristics of human motion through convolution in time, as shown in (4).

$$\text{Records} = \{(x, y, z, Doppler\ effect\ in\ m/s)^*\} \tag{4}$$

The advantage of this paper after compared with [5] is that the results of [5] buffered the spectrograms in sequence, the continuous spectrograms may be the same or highly similar due to sampling, and thus contain a lot of redundant information. In addition to the high computational efficiency, the memory requirement is relatively less than that of 2D, which can confirm that the results of this paper are indeed an improvement for [5].

In this way, the time-varying CNN can perform classification calculations after the flatten layer is expanded and handed over to the bidirectional LSTM. Therefore, this paper adopts the architecture of CNN-LSTM mentioned in [6] and improves the method of [5], which is also simpler than the neural network architecture of [7], Because [7] requires a 3D convolution operation.

III. Implementation and Test

A. Millimeter-wave radar

Fig. 2 shows the planning of the geometric space during the operation of the radar planned in this paper, Fig. 3 is the infrastructure block diagram of this millimeter-wave radar.

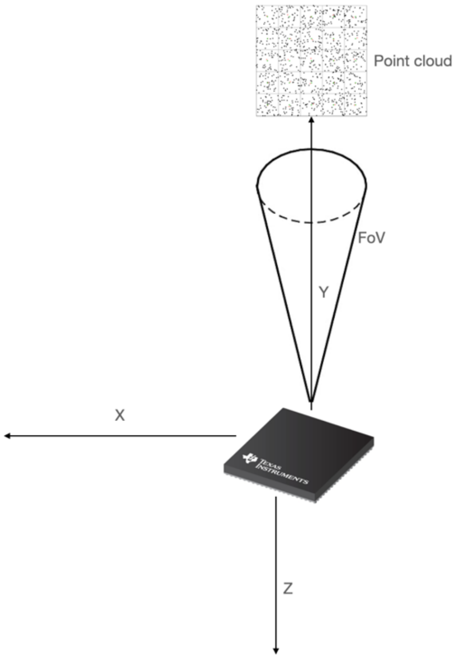


Fig. 2. Geometric space definitions of the radar.

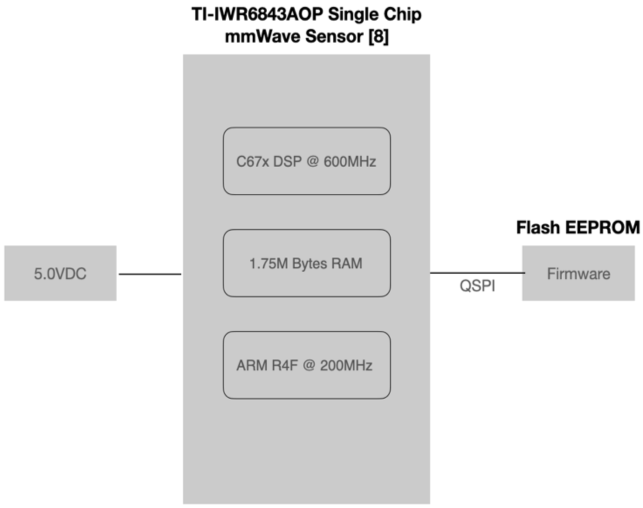
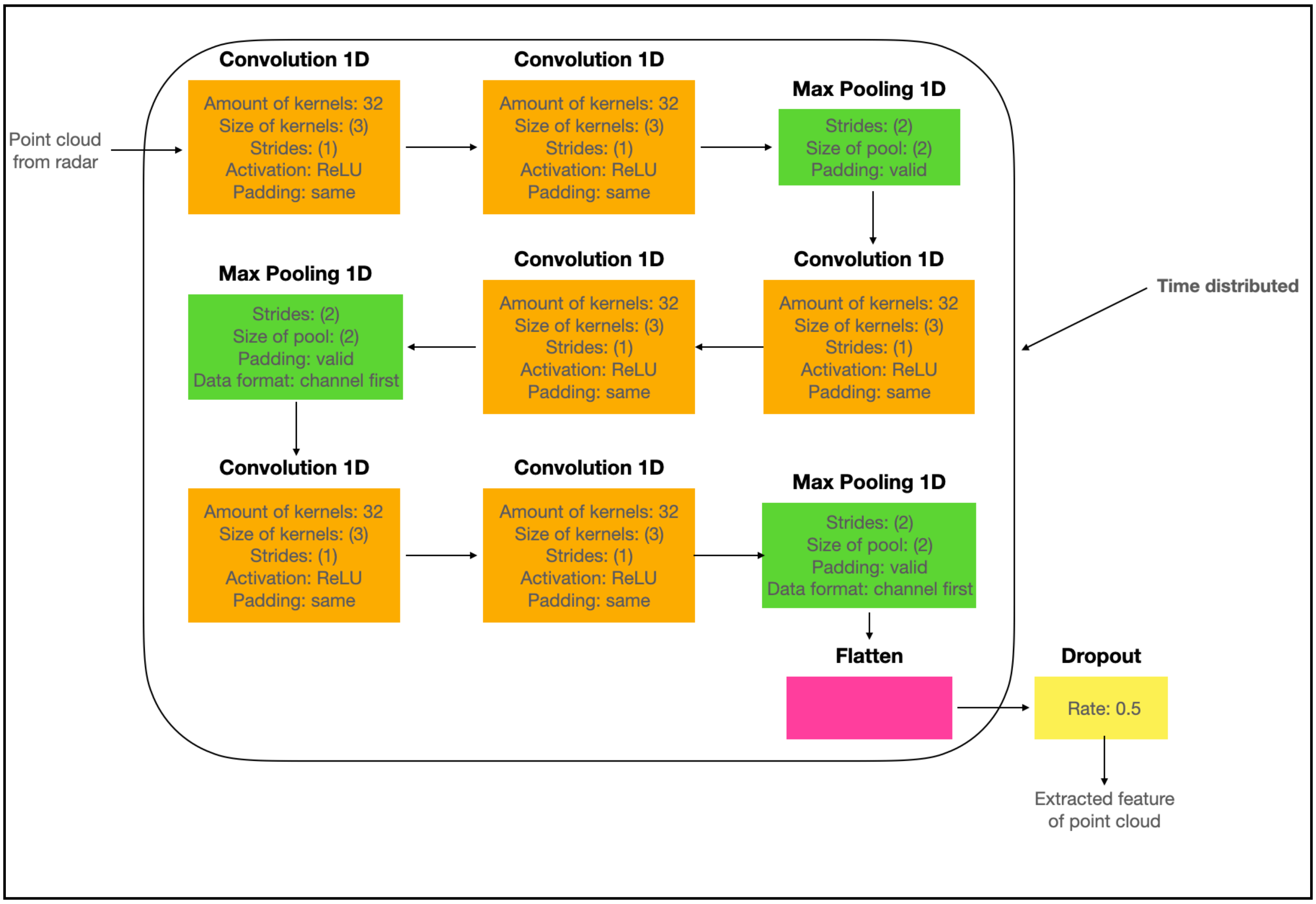


Fig. 3. Infrastructure block diagram of radar.

B. Neural network architecture

The neural network systems in this paper are all made with TensorFlow and Keras, which can be divided into two parts for reference: the first part is 1D CNN that performs convolution operations according to time distributed, shown in Fig. 4, the second part is bidirectional LSTM. Finally, it is fully expanded and classified by probability calculation with the SoftMax function. As shown in Fig. 5.

Fig. 4. 1D time distributed CNN (As below).



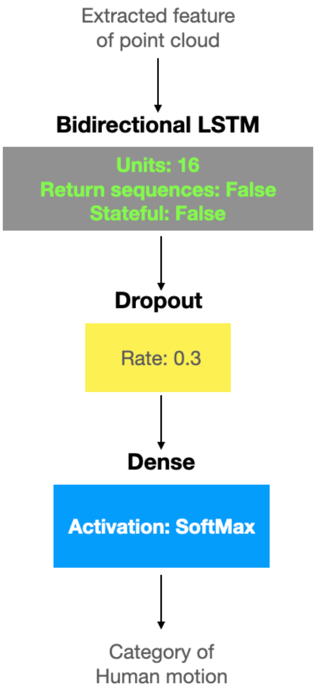


Fig. 5. Bidirectional LSTM and Dense Layer.

C. *Experiment*

- Contents

In the experiment, four categories are prepared for human motion:

1) null: In the FoV, the radar detected nothing.

2) quiet: the radar detects the object, but the object appears stationary.

3) Swing: The radar detects that the human body is swinging back and forth with both hands in situ.

4) Walking: The radar detects that the human body is walking in place.

For the above four kinds of movements, 50 times of records as shown in (4) are sampled on average each time, and then trained according to the neural network shown in Fig. 4 and Fig. 5, the epoch is 30.

- Demonstration

The millimeter-wave Doppler radar of the results of this paper is shown in Fig. 6; during the demonstration, the tester first kept still in the FoV, and then began to swing hands, as shown in Fig. 7. on the right side of the screen, the waveform generated by point cloud obtained by radar is different, and then start walking in place, as shown in Fig. 8, the radar can sense that the tester is walking in place.

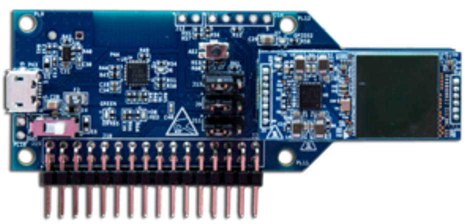


Fig. 6. The millimeter wave Doppler radar used in this paper.

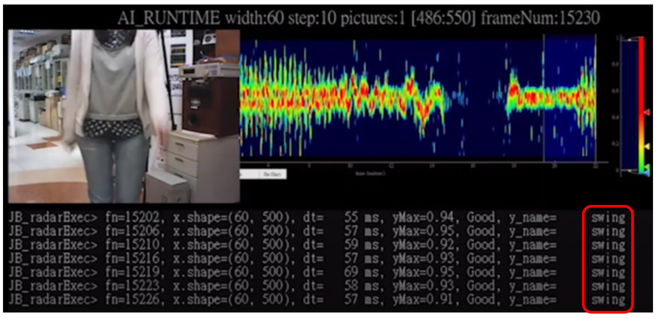


Fig. 7. Tester began to swing hands.

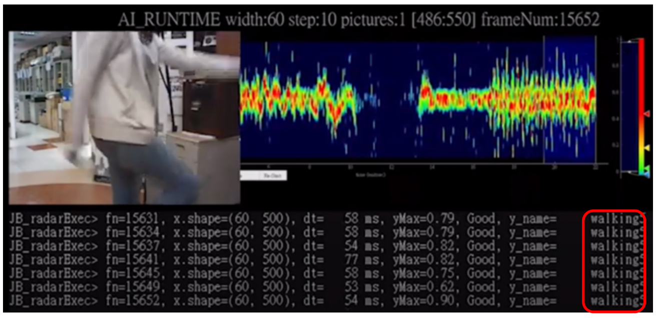


Fig. 8. Tester is walking in place.

The measurement data of each column appearing at the bottom of the screen can be compared with the definition of the coordinate axis in this article in Fig. 2, mainly about the two columns of dt and yMax. The FMCW radar in this paper computes maximum possible predicted rate for each frame, namely yMax, in the form of 22 windows per second over time, as described in (4). From Fig. 2, it can be seen that FoV is defined on the y-axis, and each point is sampled to predict through the models of Fig. 4 and Fig. 5, and the classification result of y_name on the far right can be obtained.

It can be seen from the above content that the results of this paper, after theoretical derivation and practical testing, have the effect of using millimeter-wave Doppler radar to classify human motions in real time. The detailed testing process can refer to the video of [9].

D. *Discussion*

The results of this paper are compared with other literatures as shown in Table 1. Compared with [5], the results of this paper can avoid buffering too much redundant information, which leads to the concern of memory wasting, but it requires relatively high computing performance of the processor. Comparing with [7], it can be seen that [7] has higher requirements for processor computing performance than this paper and [5], but the results of [7] and this paper can avoid redundant information due to buffering.

TABLE I. COMPARISON WITH OTHER PAPERS

	[5]	*[7]*	***This paper***
Neural network Architecture	2D CNN	3D LSTM + 3D CNN	1D time distributed CNN + LSTM
Memory usage	May have redundant information	Efficient	Efficient
Computational complexity	Low	High	Medium

IV. Conclusion

In this paper, the effectiveness of bidirectional LSTM for time series data processing is used, and the method proposed in the past literature is improved by buffering multiple continuous spectrograms and then performing real-time classification of human motion through 2D CNN one by one. At the same time, compared with the references listed so far, the CNN in this paper only needs 1D operations, and the required processor can be relatively low in computing performance.

So, the results of this paper can be regarded as a solution that can be used for real-time discrimination of human motion without special computer hardware.

Acknowledgment

This article is sponsored by the Taiwan Ministry of Science and Technology Project No. 110-2222-E-027-002, and I am very grateful to Taiwan Joybien Technology Co. Ltd., the partner of this article, for its help during this period.

References

[1] Otero, M., "Application of a continuous wave radar for human gait recognition," Defense and Security, International Society for Optics and Photonics, 2005, pp.538-548.

[2] Gurbuz, S. Z., Tekeli, B., Karabacak, C., and Yuksel, M., "Feature selection for classification of human micro-doppler," IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems (COMCAS), 2013, pp.1-5.

[3] Kim, Y. and Ling, H., "Human activity classification based on micro-doppler signatures using an artificial neural network," IEEE Antennas and Propagation Society International Symposium (AP-S), 2008, pp.1-4.

[4] Tyler S. Jordan, "Using convolutional neural networks for human activity classification on micro-Doppler radar spectrograms," Proc. SPIE 9825, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security, Defense, and Law Enforcement Applications XV, 982509, May, 2016.

[5] Renyuan Zhang, Siyang Cao, "Real-time Human Motion Behavior Detection via CNN using mmWave Radar," IEEE Sensors Letters, vol. 3, issue 2, Feb, 2019.

[6] Md. Zabirul Islam, Md. Milon Islam, Amanullah Asraf, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images," Informatics in Medicine Unlocked journal, vol.20, 100412, 2020.

[7] Yuecong Min, Yanxiao Zhang, Xiujuan Chai, Xilin Chen, "An Efficient PointLSTM for Point Clouds Based Gesture Recognition," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[8] Texas Instruments, "IWR6843AOP -- Single-chip 60-GHz to 64-GHz intelligent mmWave sensor with integrated antenna on package (AoP)", https://www.ti.com/product/IWR6843AOP [online].

[9] Chien-Hung Lai, "The mmWave Doppler effect senses the movement changes of the human body, and the first phase of the experiment was successful on 2021/12/03", https://www.youtube.com/watch?v=gO5ArlkdOh0 [online].