RaCon: A gesture recognition approach via Doppler radar for intelligent human-robot interaction

Kaijie Zhang
School of Computer Science
Northwestern Polytechnical University
Xi'an, Shaanxi, China
zhangkj09@foxmail.com

Zhiwen Yu
School of Computer Science
Northwestern Polytechnical University
Xi'an, Shaanxi, China
zhiwenyu@nwpu.edu.cn

Dong Zhang
School of Computer Science
Northwestern Polytechnical University
Xi'an, Shaanxi, China
drower@mail.nwpu.edu.cn

Zhu Wang
School of Computer Science
Northwestern Polytechnical University
Xi'an, Shaanxi, China
wangzhu@nwpu.edu.cn

Bin Guo
School of Computer Science
Northwestern Polytechnical University
Xi'an, Shaanxi, China
guob@nwpu.edu.cn

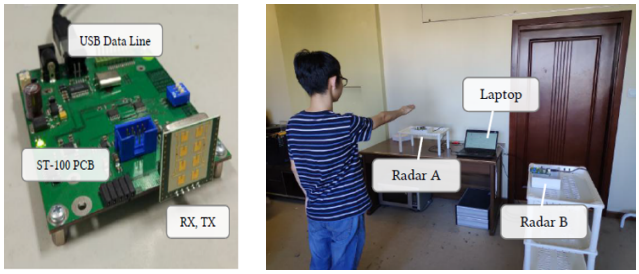
***Abstract*—As an important entrance for human-robot interaction, the hand gesture recognition based on wireless sensor has received great attention in recent years. By recognizing fine-grained arm movements, remotely deployed collaborative robot could work more accurately to satisfy human demands. Existing approaches mostly use wearable sensors or wireless devices to recognize human movement, which is with strict position requirements. In this paper, we propose a robust gesture recognition method based on double Doppler radars. Specifically, we use two Doppler radars to collect two sources of doppler signal of a gesture. Then 6 types of gestures with different angles between people and the radar were classified by employing an improved dynamic time warping (DTW) algorithm. Furthermore, we demonstrate the practicability of the proposed method by developing a cooperative robot control system and the average recognition accuracy is 96%.**

***Index Terms*—Wireless sensing; Gesture recognition; Human-robot interaction; Doppler radar; Signal synthesis analysis.**

I. Introduction

The emerging technology of wireless sensing and the prevalence of the internet of things have been effectively promoted to new ways of human-computer interaction. In prior studies, it is being actively used to interpret the hand gesture data based on computer vision [1], wearable sensor [2] [3] and sound wave [4]. Though high precision can be achieved, there are some disadvantages like inconvenience caused by wearing equipment and strict environmental requirements. For example, they need good lighting environment or relatively quiet environment. These lead to a number of open technical challenges, including (1) how to accurately identify different gestures caused by the target user from wireless signals, (2) how to analyze the tiny radio reflections and extract distinguishable features, and (3) how to analyze the fine-grained details of gestures, such as the angle of an arm movement.

To address the above challenges, we explored another contactless approach, i.e., radar. This paper shows RaCon, an arm gesture recognition system based on radar. Fig. 1(a) shows the radar device we used. Fig. 1(b) shows the working environment of the system. The main idea is to analyze and utilize doppler information generated by user actions on two radars. RaCon proposes a time-frequency spectrum feature extraction and synthesis method to obtain complete motion information. Based on this new time-frequency feature, RaCon can determine the gesture category and some additional information such as the direction and angle of gesture, with a higher accuracy.



(a) 24GHz radar sensor (b) Working environment

Fig. 1. Arm Gesture Recognition System

In this work, we have made the following contributions:

- A robust multi-radar gesture recognition system. RaCon uses Doppler radar sensors to capture and identify the types of human gestures as well as other details. RaCon provides a robust gesture recognition method by using two vertically placed radars to get rich information.
- Reliable signal synthesis analysis. We propose a signal time-frequency synthesis method to make a comprehensive exploration of the two radars. It provides complete motion information for any action to ensure accurate identification results.
- Accurate gesture recognition system. We demonstrate a human-robot interaction system in the real environment to test our method. The results show that RaCon achieves an average

success rate of 96% for gesture recognition.

The remainder of this paper is organized as follows. Section II reviews related work. Section III introduces the system architecture. Section IV provides the details of signal processing and gesture recognition algorithms. Section V presents experimental settings and the performance of the RaCon. We conclude our work and discuss future directions in the Section VII.

II. RELATED WORK

Existing work using wireless sensing for interactive behavior recognition can be divided into two categories: contact device based and non-contact device based. Contact device based approaches require a user to wear extra devices, such as magnetic field sensors [5] and mobile phone [6]. They are great work with high accuracy and high applicability despite the inconvenience of extra equipment. We think that non-contact approaches are more suitable for natural human computer interaction. Non-contact approaches can be mainly further divided into Wi-Fi based and radar based, which are described below.

Wi-Fi based approach: Human behavior recognition based on Wi-Fi signal is quite popular. In the early days, there are many researches use Received Signal Strength (RSS) as a fingerprint feature to achieve localization, tracking or recognition [7]. However, the recognition accuracy is relatively low because Wi-Fi signal is unevenly distributed in the air. Later, researchers shift their focus to the Channel State Information (CSI). By analyzing such information, CSI can be used for the recognition of fine-grained human behaviors. For example, WiFinger [8] and WiGeR [9] can identify finger motions like digital gestures or palm movements by using CSI information in a certain environment.

Radar based approach: Two types of radar are mainly used in human behavior recognition, i.e. Frequency Modulation Continuous Wave (FMCW) and Doppler. FMCW radar increases reliability by providing distance measurement along with speed measurement. Some subsequent studies that aim to recognize human behaviors such as gestures [10] and activities [11]. Some existing studies on Doppler radar choose to adopt Doppler radar to recognize human behaviors such as the RAM system [12] and the Tongue-n-Cheek system [13]. Some researchers choose to use the image recognition method to analyze the spectrogram of doppler radar for gesture recognition such as Zhang et al [14] and Kim [15]. Our previous work Gesture-Radar [16] used a single Doppler radar placed in front of the user to distinguish 5 different kinds of gestures.

III. SYSTEM OVERVIEW

RaCon is a wireless sensing system that utilizes 24GHz K-LC2 short range radar devices to recognize people's arm gestures. The architecture of RaCon is shown in Fig. 2, which includes three parts.

Signal Sampling Unit. The unit comprises two Doppler radars. The radars are deployed at about 1.5 meters above the ground and the distance between people and radar is within 2 meters (as shown in Fig. 1(b)). The radar facing the user is called "front radar", and the radar on the right of the user is called "side radar". The user stands in front of the radar and moves his/her arm in the air to make the specified gestures. Meanwhile, the quadrature signal will be captured by each radar and sent to the server with a sampling rate of 44.1KHz.

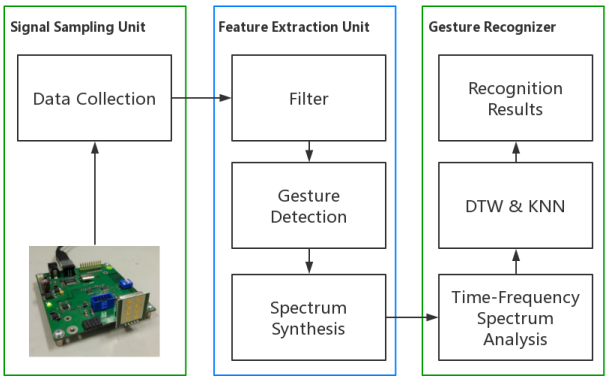


Fig. 2. System Overview

Feature Extraction Unit. RaCon system utilizes various de-noising methods to minimize the interference of environmental noise. We use a high-pass filter and double threshold method to extract "active" part of radar signal. At the end, there is a spectrum analysis method to arrange signals from two radars as a gesture profile. It includes two practical calculation methods: spectrum synthesis and feature extraction.

Gesture Recognizer. From each gesture profile, a special time-frequency feature vector is extracted. We use a k-Nearest Neighbor (kNN) algorithm for gesture classification, where the distance between people's movement feature vector and model vector is calculated by Restricted Dynamic Time Warping (R-DTW) algorithm. Finally, we utilize phase discrimination to obtain the direction of the gesture, as well as the angle of gesture from time-frequency feature.

IV. GESTURE RECOGNITION ALGORITHM

A. GESTURE DETECTION

The main effect of human arm gestures on received radar signal is a series of ups and downs or pausing. Unlike gesture signals, meaningless signals that represent no-action signal segments are usually gentle. So, we use a double threshold endpoint detection method to extract the signal segments corresponding to a gesture.

1) Gesture Segmentation: We use double-threshold endpoint detection and design an adaptive method to detect the start and end points of arm gestures. Before the segmentation, the filtered signal $S(t)$ is framed to $s(i)$ by the sliding window method. At 44.1KHz sampling rate, the frame length is 4410 samples, the frame shift is 2205 samples. The overlap is 50%.

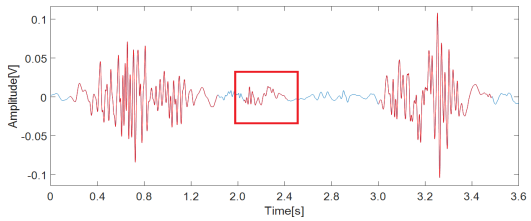


Fig. 3. Burst Noise

Therefore, we can calculate the short-term-crossing rate with threshold T as equation (1):

$$Z(i) = \sum_{j=1}^{N} \{abs(sign[s(j) - T] - sign[s(j-1) - T]) \\ -abs(sign[s(j) + T] - sign[s(j-1) + T])\} \quad (1)$$

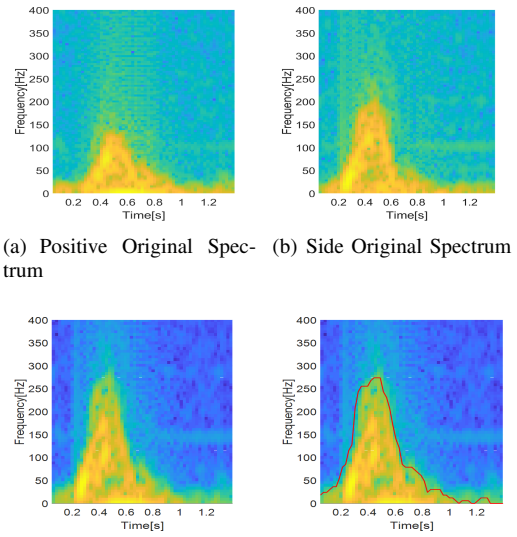
where $s(j), j < [1, N]$, is a data frame and the N is the frame length of $s(j)$. T is initially set as 0.015V based on the prior knowledge. $sign()$ is a symbolic function.

The purpose of threshold is to better obtain the arm gesture profile. A signal slice is considered to correspond to an arm gesture where $Z(i)$ continuously exceeds T_z. In particular, T_z is the threshold of the short-term-crossing rate, which is set as 50, and it is used to determine if the slice is likely to be a gesture. For example, if $Z(i)$ of 10 consecutive frames exceeds T_z, the signal of these consecutive frames would be considered as an arm gesture. What's more, considering human instability, we set a tolerance interval T_t, which means that even if $Z(i)$ is lower than T_z in a short time, the detection result will not change. T_t is set as 3 frames. In particular, no matter which radar detects a gesture, it will extract the signal. And the extraction method will acquire as many radar signals as possible.

2) Outlier Removal: In the process of different gesture conversions, a series of noises are inevitably generated due to the continuity of human motion. Some of the noises are difficult to filter out by the above low-pass filter, and thus may be erroneously detected by the motion detector. Fig. 3 shows the waveform of a burst noise between two arm gestures. Obviously, these outliers are not what we need. Given that outliers probably affect gesture recognition, RaCon utilizes short-term energy to eliminate these outliers. We suppose that if the energy of a signal segment is significantly low or even close to zero then it is a burst noise and it will be abandoned

B. FEATURE EXTRACTION

After gesture detection, RaCon obtains a movement profile for an arm gesture. Then RaCon applies short time fourier transform (STFT) to get the time-frequency spectrum like Fig. 4(a)(b), where the window length, noverlap and length of FFT are 4096, 75% and 10240. As the angle between radar and people increases, the Radar cross-section decreases. Then, the amplitudes of gesture signal on two radar devices are lower than the amplitudes on one radar devices when this gesture has a certain angle with the radar instead of facing it. Therefore, it is necessary to comprehensively use the information of the two radars.



(a) Positive Original Spectrum (b) Side Original Spectrum

(c) Synthetic Spectrum (d) Effective Spectral Feature Area

Fig. 4. Spectrum Feature Extraction

On the other hand, the velocity is proportional to the frequency shift in the Doppler principle as equation (2):

$$\Theta(t) = tan^{-1}\frac{Q(t)}{I(t)} = 2\pi * f_d t + \phi = \frac{4\pi v t}{\lambda} + \phi \quad (2)$$

where $Q(t)$ and $I(t)$ are orthogonal signals that make up $S(T)$, f_d is doppler shift and ϕ represents the initial phase resulting from the distance between the user and the radar. Based on this, we can refer to the vector nature of speed and synthesize two spectrums according to the formula $f_d = \sqrt{(f_{d1}^2 + f_{d2}^2)}$. As shown in Fig. 4, Fig. 4(a) and Fig. 4(b) are synthesized to Fig. 4(c).

According to above factors, RaCon calculates the spectrum from two radars to form a synthetic spectrum like Fig. 4(c). It's difficult to extract feature from spectrum because the time-frequency spectrum which we get from STFT is an energy hot image. Therefore, we make binary normalization of the obtained time-frequency map using the Minmax Scalar method. There are obvious differences between hotspots and non-hotspots in the binary map. To extract features, we smooth the result through a Gaussian filter with a Gaussian radius of 2 and use the Canny operator to make a first-order difference to calculate the direction and magnitude of the edge. Then maximum suppression and double threshold detection method are used to get features, where the lower threshold is 0.3 times of higher threshold in the double threshold detection method. The effective spectral feature area is finally obtained as shown in Fig. 4(d).

What's more, we assume that the ratio of the arm movement velocities observed by two radars are stable at any moment.

So Θ will be used to get gesture detail, the angle θ of gesture, as equation (3):

$$\theta = \frac{v_1}{v_2} = \frac{f_{d1}}{f_{d2}} = \frac{d\Theta_1/dt}{d\Theta_2/dt} \tag{3}$$

C. Classfication

RaCon adopts the kNN classifier to identify different arm gestures because of its intuition, strong applicability and high accuracy. The performance of a kNN classifier is mainly determined by the used similarity measure. To accurately calculate the similarity between two gesture waveforms, we propose to use the DTW with restricted paths. In particular, DTW is a method that calculates an optimal match between two given gesture sequence, which include the amplitude of above spectral feature area. To calculate the distance between test gesture $X = (x_1, x_2, ..., x_n)$ and sample gesture $Y = (y_1, y_2, ..., y_m)$ (n and m are the waveform lengths), DTW first constructs a matrix d(m, n) to preserve the distance of any pair of points in X and Y as $d(i, j) = [X(i) - Y(j)]^2$. Afterwards, DTW calculates a new matrix D(m, n) to represent the total distance, which are defined as equation (4):

$$D(i,j) = min \left\{ \begin{array}{c} [d(i,j) + D(i-1,j)] \\ [d(i,j) + D(i,j-1)] \\ [2 * d(i,j) + D(i-1,j-1)] \end{array} \right\} \tag{4}$$

Based on dynamic programming, we can obtain the final distance $D(m, n)$ of two waveforms and restore the optimal warping path. It is worth to mention that there is some sudden high-frequency noises when people do several gestures in a row. To filter out such wrong actions, we set an acceptance domain as $\eta = max[DTW(M', M_i)]$, where M_i is a gesture sequence of sample set and M' is the sample with the smallest distance from other samples. Let the distance between different samples as $R_i = \sum_{j=1}^{N} DTW(M_i, M_j)$ and choose the sample with the smallest R_i as M'. The maximum distance between M' and other sequences is taken as the DTW rejection threshold. In the DTW calculation process, if the R_k of the current gesture sequence M_k is greater than η, the algorithm rejects this result and determines that the action is untyped. At last, RaCon will apply kNN to get the final classification results.

V. EXPERIMENTS AND RESULTS

In the experiment, we use two 24GHz K-LC2 short range radars with ST-100 radar development kits and a dual core CPU, 8GB ROM laptop. The radar is placed at a height of 1.5m above the ground while the user stands about 1.5m away. When the sensor works, the transmitter will send pulse signals with a high rate of 24GHz to the user and the receiving antenna will capture the reflected signals. Then the sensor will transmit the quadrature signal to the laptop in the wav file with the sampling rate of 44.1KHz in real time.

A. Data Collection

We collected experimental data from 6 volunteers to feed the classification algorithm, who were first told how to perform

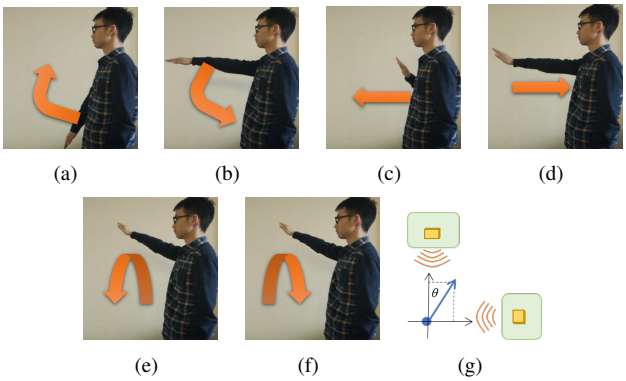


Fig. 5. Gesture Set

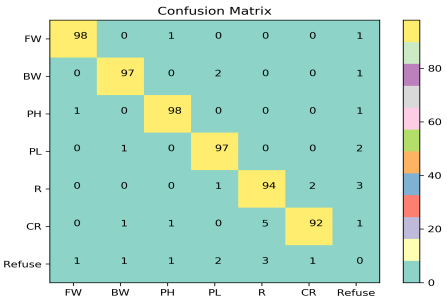


Fig. 6. Classification Results

the 6 target arm gestures with 5 different angles. The details of the gestures are shown in Fig. 5. Fig. 5(a) to Fig. 5(f) represent the 6 target arm gestures of Forward, Backward, Push, Pull, Rotate and CounterRotate. Fig. 5(g) shows that θ is the angle between the arm and the radar. Volunteers are free to take their own sequences of gestures, i.e. in any order and any angle they want. Specifically, the volunteers were asked to perform each gesture 5 times with an average time interval of one second. Each volunteer actually contributed 150 effective gesture samples. Finally, we take the 300 samples from first two volunteers as the training set, and the other 600 samples from the other four volunteers as the testing set.

B. Gesture Recognition Performance

1) Detection Accuracy: We evaluate the accuracy of the arm gesture detection algorithm in RaCon. The detection accuracy is defined as the ratio of total number of correctly extracted arm gesture to the total number of arm gesture we get. The 600 gestures were automatically segmented from radar data using the method described in Section III. The average accuracy of segmentation is 98%. The main reasons involved in the wrong segmentations are as follows: 1) some volunteers make the gesture too slowly due to the unfamiliar; and 2) there is not enough separation between two gestures in some data set.

2) Gesture Classification Accuracy: Fig. 6 presents a confusion matrix that illustrates the classification accuracy of the RaCon. The confusion matrix was constructed using all 600 gestures performed by the volunteers. The average accuracy

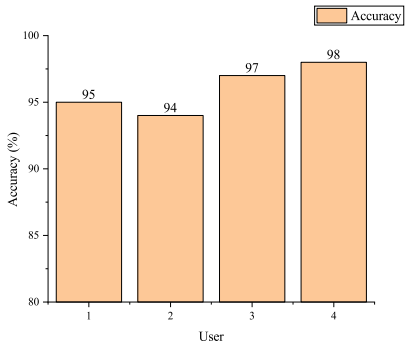


Fig. 7. Average accuracy of each user

of RaCon is 96%. For the training part, when the volunteers performed each gesture only once, the obtained system still performs well, as shown in Fig. 7. We observed that the classification accuracies of testing set are at a high level. Therefore, even if the training set is not large, the system is still with high availability.

What's more, we made a comparative experiment in our system:

1) TO: Use time domain feature only from front radar for classification with two-stage classification method[16], a classification method based on DTW-kNN.
2) TN: Use time domain feature from two radars for classification with two-stage classification method. But the gesture to be classified is non-standard, which means that there is an angle θ between arm and radar.
3) TF: Use time-frequency synthesis analysis method to classify the non-standard gesture.

The target set includes 6 different gestures with no angle between people and radar, and TO achieves a 96% accuracy.

The TN approach has an accuracy of 80%. The result obviously decreases where the target set has extra angle between arm and radar. The reason is that the time domain waveform amplitude of non-standard gestures is significantly lower than the standard gesture without angle between arm and radar. Lower amplitude leads DTW-KNN to make wrong judgments.

The TF method has a 96% accuracy, which indicates that the time-frequency synthesis analysis method can effectively overcome the impact of non-standard actions on the recognition system.

C. Gesture Details Profiling Performance

We use cameras and image observation method to calculate the actual moving angle of some gestures as testing set. RaCon analyze details of these gestures and Table II shows the experimental results of the average angle calculation. The recognition precision is set as 1 - (average deviation / 15°). Obviously, the calculation results are basically correct when the angle θ is not too large or too small. However, there is bigger deviation in the edge position. This is a shortcoming that needs our future improvement.

TABLE I
GESTURE DETAILS CALCULATION RESULTS

Actual Angle	Average Deviation	Precision
0°	9.8°	34.7%
15°	6.1°	59.3%
30°	4.4°	70.7%
45°	2.4°	84.0%
60°	4.6°	69.3%
75°	5.5°	63.3%
90°	10.0°	33.4%

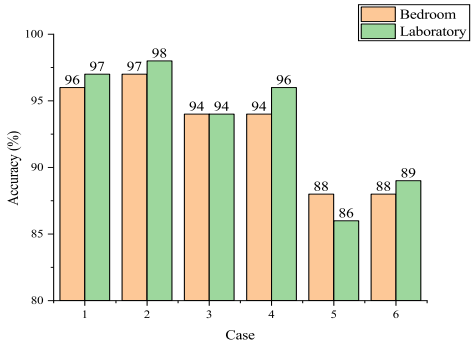


Fig. 8. Gesture Recognition Accuracy in Different Environment

D. Influence of Different Radar Positions

TABLE II
DIFFERENT ENVIRONMENTS

Case	Height of Two Radars	Angle
1	1.3	0
2	1.5	0
3	1.3	10
4	1.5	10
5	1.3	20
6	1.5	20

We tried different environment layouts during the experiments, including 2 rooms (a bedroom, and a laboratory), 6 different radar heights, and 3 angles between human and radar, as shown in Table I. In Fig. 8, we presented all the recognition accuracy in these experimental environments. To sum up, we observed that the accuracy was mainly affected by the angles between human and radar. The difference between the average recognition accuracy in two different rooms was not obvious, and there was no significant performance change when the device height changed from 1.3 to 1.5 meters. However, when the angle between a user and a device increases, the accuracy of direction detection would decrease significantly. Thus, the angle between a device and a user is particularly important for the arm gesture direction recognition system.

E. Results of Different Classification Approaches

Other three experiments were conducted to evaluate the recognition accuracy under different classification approaches.

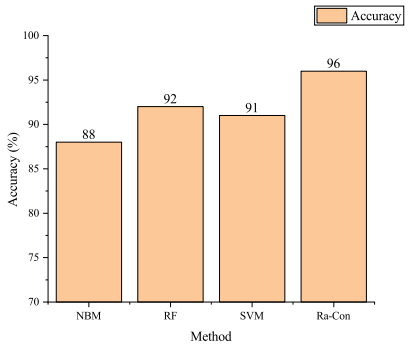


Fig. 9. Accuracy of different recognition models

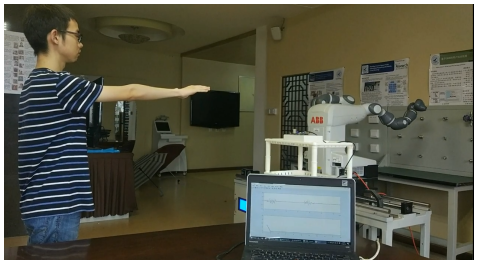


Fig. 10. Using arm gestures to control a robot arm

RAM uses two Doppler radars to recognize human activities, which is similar to our work. RAM extracted multiple features and classified them using a Random Forest model. Thus, we adopt the same experimental method for comparison. Specifically, we extract 6 features from filtered radar signal, including average amplitude, standard deviation, sample range, total energy, phase difference and velocity. Then, we use Naive Bayesian model (NBM), Random Forest (RF) and Support Vector Machine (SVM) to perform 10-fold cross validation and obtain the gesture type recognition results, as shown in Fig. 9. It is observed that the average recognition accuracy of RaCon is higher than that of the conventional classification method based on multi-feature extraction in the same experimental environment.

F. Application

We hope that this method can be used to achieve natural human-computer interaction, so we design experiments to control actual robotic arm (ABB IRB-14000) as in Fig. 10. We define the response time as the time from the gesture action to the start of the movement of the robotic arm, excluding the communication delay caused by the distance. The average response time of the control system is 0.7 seconds, which shows good real-time performance. This indicates that the method can be combined with a teleoperation system.

VI. CONCLUSION

In this paper, we present RaCon, a novel approach which can recognize interactive behaviors through two Doppler radar sensors. It does not require strict user location or room layout, and the results of our preliminary study show that RaCon is able to achieve high recognition performance where the target gesture set includes 6 different gestures with at least 5 different angles. Meanwhile, the proposed approach has certain anti-interference capability. Although there are still some shortcomings in details profiling and types of actions, we will provide better methods to achieve effective natural human-computer interaction.

REFERENCES

[1] M. Jacob, Y.-T. Li, G. Akingba, and J. P. Wachs, "Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room," *Journal of Robotic Surgery*, vol. 6, no. 1, pp. 53–63, 2012.

[2] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song, "E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pp. 260–273, ACM, 2011.

[3] B. Fang, N. D. Lane, M. Zhang, A. Boran, and F. Kawsar, "Bodyscan: Enabling radio-based sensing on wearable devices for contactless activity and vital sign monitoring," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 97–110, ACM, 2016.

[4] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the doppler effect to sense gestures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1911–1914, ACM, 2012.

[5] H. Ketabdar, P. Moghadam, B. Naderi, and M. Roshandel, "Magnetic signatures in air for mobile devices," in *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion*, pp. 185–188, ACM, 2012.

[6] S. Agrawal, I. Constandache, S. Gaonkar, R. Roy Choudhury, K. Caves, and F. DeRuyter, "Using mobile phones to write in air," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pp. 15–28, ACM, 2011.

[7] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, pp. 1472–1480, IEEE, 2015.

[8] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "Wifinger: talk to your smart devices with finger-grained gesture," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 250–261, ACM, 2016.

[9] M. Al-qaness and F. Li, "Wiger: Wifi-based gesture recognition system," *ISPRS International Journal of Geo-Information*, vol. 5, no. 6, p. 92, 2016.

[10] Z. Peng, C. Li, J.-M. Muñoz-Ferreras, and R. Gómez-García, "An fmcw radar sensor for human gesture recognition in the presence of multiple targets," in *2017 First IEEE MTT-S International Microwave Bio Conference (IMBIOC)*, pp. 1–3, IEEE, 2017.

[11] K. Chetty, Q. Chen, M. Ritchie, and K. Woodbridge, "A low-cost through-the-wall fmcw radar for stand-off operation and activity detection," in *Radar Sensor Technology XXI*, vol. 10188, p. 1018808, International Society for Optics and Photonics, 2017.

[12] M. A. A. H. Khan, R. Kukkapalli, P. Waradpande, S. Kulandaivel, N. Banerjee, N. Roy, and R. Robucci, "Ram: Radar-based activity monitor," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, IEEE, 2016.

[13] Z. Li, R. Robucci, N. Banerjee, and C. Patel, "Tongue-n-cheek: non-contact tongue gesture recognition," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pp. 95–105, ACM, 2015.

[14] B. Dekker, S. Jacobs, A. Kossen, M. Kruithof, A. Huizing, and M. Geurts, "Gesture recognition with a low power fmcw radar and a deep convolutional neural network," in *2017 European Radar Conference (EURAD)*, pp. 163–166, IEEE, 2017.

[15] Y. Kim and B. Toomajian, "Hand gesture recognition using micro-doppler signatures with convolutional neural network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[16] X. Lou, Z. Yu, Z. Wang, K. Zhang, and B. Guo, "Gesture-radar: Enabling natural human-computer interactions with radar-based adaptive and robust arm gesture recognition," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 4291–4297, IEEE, 2018.