Progress In Electromagnetics Research M, Vol. 79, 71–80, 2019

Body Gesture Recognition Based on Polarimetric Micro-Doppler Signature and Using Deep Convolutional Neural Network

Wenwu Kang[1, 2], Yunhua Zhang[1, 2, *], and Xiao Dong[1]

Abstract—Body gesture recognition can be applied not only to social security but also to rescue operations. In reality, body gesture can produce unique micro-Doppler signatures (MDSs), which can be used for identification. In this paper, we first acquired the echo signals of four body gestures via a Ka-band dual-polarization radar system under different angles and distances. The four gestures are respectively swinging arm up and down, swinging arm left and right, nodding, and shaking head. Then, time-frequency spectrograms were obtained by short-time Fourier transform, from which we can see that different body gestures have different polarimetric MDSs. Finally, we propose to classify four body gestures using the deep convolutional neural network (DCNN) method. It is shown that by combining HH and HV polarizations, about 92.7% recognition rate is achieved while only about 77.5% and 89.3% rates are obtained by using single HH polarization and single HV polarization, respectively.

1. INTRODUCTION

In recent years, recognition of body gestures has attracted wide attention in radar community for its potential applications in social security and rescue operations [1, 2]. When a moving target is illuminated by a radar, the Doppler frequency modulation is generated in the echo signal. Body gestures can cause Doppler frequency shifts known as micro-Doppler (m-D) frequencies [3–5]. Time-frequency (TF) representations are usually used to analyze micro-Doppler signatures (MDSs) among which short time Fourier transform (STFT) is mostly adopted. Different body gestures generate different MDSs, which can be explored to identify gestures conversely [6, 7].

During the last two years, deep convolutional neural network (DCNN) has been successfully applied to image recognition [8–12]. Compared with other traditional recognition algorithms, DCNN does not rely on handcrafted features. There are also a number of researches about recognition of body gestures reported based on MDSs using DCNN [8–12]. In [8], four targets, i.e., human, dog, horse, and car, were measured for human detection, and seven body gestures were measured for human activity recognition. In [9], eight human hand gestures, i.e., swiping from left to right, swiping from right to left, swiping from up to down, swiping from down to up, rotating clockwise, rotating counterclockwise, pushing, and holding, were measured. The feasibility of recognizing human hand gestures was investigated via using MDSs measured with a DCNN. In [10], transmission coefficient and reflection coefficient of on-body antenna were proposed to classify body gestures using DCNN. In [11], multistatic radar MDSs were applied to classify personnel and human gaits using DCNN. In [12], a K-band Doppler radar was applied to identify human based on MDSs using DCNN.

At present, most researches about body gesture recognition are conducted via single polarized radar (HH or VV), and only limited information on MDS can be obtained. Polarimetric MDS of a pedestrian has been studied [13], where simulated and experimental results show the same interesting phenomenon:

Received 15 November 2018, Accepted 24 January 2019, Scheduled 16 February 2019

* Corresponding author: Yunhua Zhang (zhangyunhua@mirslab.cn).

[1] CAS Laboratory of Microwave Remote Sensing, National Space Science Center, Chinese Academy of Sciences, Beijing, China.

[2] University of Chinese Academy of Sciences, Beijing, China.

HH m-D of arms is rising, while HV m-D is falling. Furthermore, most researches dealt with only a single viewing angle observation. However, in practical situations, the MDS of body gestures, such as the m-D frequency range and m-D strength are largely influenced by the viewing angle and distance, which means that observations at different angles and distances are necessary in experiment. Experimental polarimetric MDS analysis of small drones was researched in [14]. In this experiment, an interesting phenomenon was observed: the co-polarised antenna receives better signals when the aspect angle is 0°, whereas the cross-polarised antenna receives better signals when the aspect angle is 90°. In this paper, we focus on body gesture recognition based on polarimetric MDS using DCNN without restricting the viewing angle and distance, i.e., it is much closer to real application scenarios. First, we collected echo signals of four body gestures via a real Ka-band dual-polarimetric radar system. The four gestures are respectively swinging arm up and down, swinging arm left and right, nodding, and shaking head. Then, three kinds of TF spectrograms were obtained by STFT corresponding to HH polarization, HV polarization, and the combination of HH and HV polarizations, respectively. Finally, the four gestures were recognized using DCNN. The results show that about 92.7% recognition rate is achieved by using the time-frequency spectrogram of combination of HH and HV polarizations, while it is about 77.5% and 89.3% by using the time-frequency spectrograms of HH and HV polarizations, respectively. We further selected arbitrary three gestures and conducted the recognition experiment, and the recognition results show that the combination of HH and HV polarizations acquires a higher recognition rate than using single HH polarization or single HV polarization. Simultaneously, the SVM method has been used to recognize four body gestures for a comparison with the proposed method. The results show that the DCNN method outperforms the SVM method.

The remaining parts of this paper are organized as follows. In Section 2, the DCNN is introduced. In Section 3, real radar experiment is carried out using a Ka-band polarimetric radar. In Section 4, the recognition results of four human body gestures are presented using DCNN. As for comparison, the SVM method is used to process the same experiment data. Finally, we conclude the paper in Section 5.

2. DEEP CONVOLUTIONAL NEURAL NETWORKS

DCNN is one of the most successful deep learning algorithms, which has been very well applied to image recognition and identification in recent years. Compared with traditional machine learning algorithms, DCNN does not rely on handcrafted features and can attempt to learn mapping between input image and its corresponding label provided by human annotator. Inception-v3 is a well-performed DCNN model [15] widely used by many researchers, which was designed to reduce the computational cost while improving the recognition rate so that it could be even ported in mobile vision-related applications. In this paper, inception-v3 model is used to classify four body gestures based on TF spectrograms. Among the experimental data, 80% of the spectrograms of each body gesture were selected randomly as the training data set; 10% of the spectrograms of each body gesture were used as the validation data set; and the rest 10% were used as the test data set. During the training, the learning rate, training steps, and the number of nodes in the bottleneck layer were set to 0.01, 3000, and 2048, respectively. Figure 1

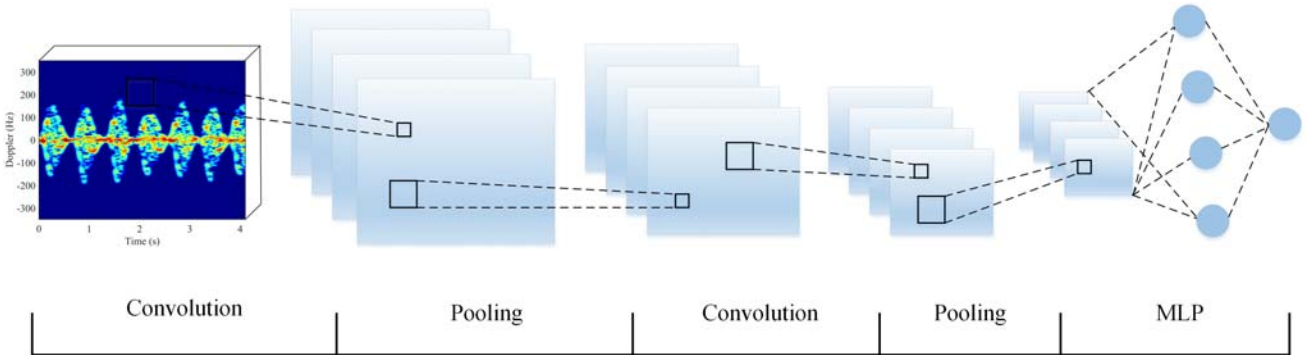


Figure 1. Architecture of convolutional neural network.

Table 1. The outline of the network architecture.

Type	Patch size/stride or remarks	Input size
conv	$3 * 3/2$	$299 * 299 * 3$
conv	$3 * 3/1$	$149 * 149 * 32$
conv padded	$3 * 3/1$	$147 * 147 * 32$
pool	$3 * 3/2$	$147 * 147 * 64$
conv	$3 * 3/1$	$73 * 73 * 64$
conv	$3 * 3/2$	$71 * 71 * 80$
conv	$3 * 3/1$	$35 * 35 * 192$
3 * Inception	As in Figure 2(a)	$35 * 35 * 288$
5 * Inception	As in Figure 2(b)	$17 * 17 * 768$
2 * Inception	As in Figure 2(c)	$8 * 8 * 1280$
pool	$8 * 8$	$8 * 8 * 2048$
linear	logits	$1 * 1 * 2048$
softmax	classifier	$1 * 1 * 4$

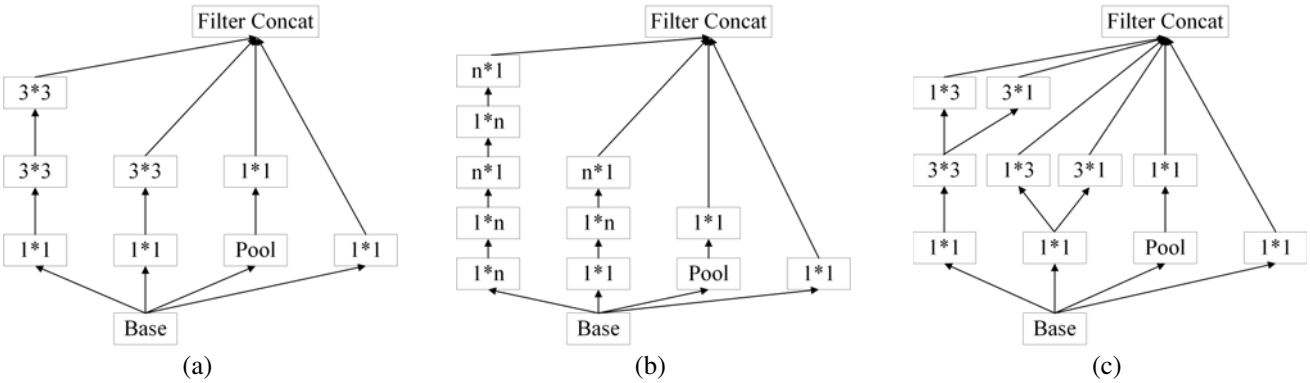


Figure 2. Inception models. (a) Inception models where each $5 * 5$ convolution is replaced by two $3 * 3$. (b) Inception models after the factorization of the $n * n$ convolutions. (c) Inception models with expanded filter bank outputs.

shows the architecture of convolutional neural network. The main structure of Inception-v3 model is show in Table 1 and Figure 2 [17].

3. EXPERIMENT

We conducted the experiment with a Ka-band dual-polarization radar system in an indoor laboratory environment. The transmitting antenna is a horn antenna of 26.2 dBi gain and 8.5° beamwidth, and the receiving antenna is an orthogonal mode horn antenna with two orthogonal polarization outputs. Figures 3(a) and (b) show the transmitting antenna and receiving antenna, respectively. During the experiment, the transmitting antenna transmitted horizontally linearly polarized signal, while the receiving antenna received horizontally and vertically linearly polarized echo signals simultaneously. So, HH and HV polarization measurements were obtained in this fashion. Stepped-frequency chirp signal (SFCS) is adopted, and the total bandwidth is 2 GHz with 20 sub-chirps, each of which has a bandwidth of 110 MHz. The carrier frequencies increase from 33 GHz to 35 GHz at a step of 100 MHz.

(a) (b)

Figure 3. System antennas. (a) The transmitting antenna. (b) The receiving antenna.

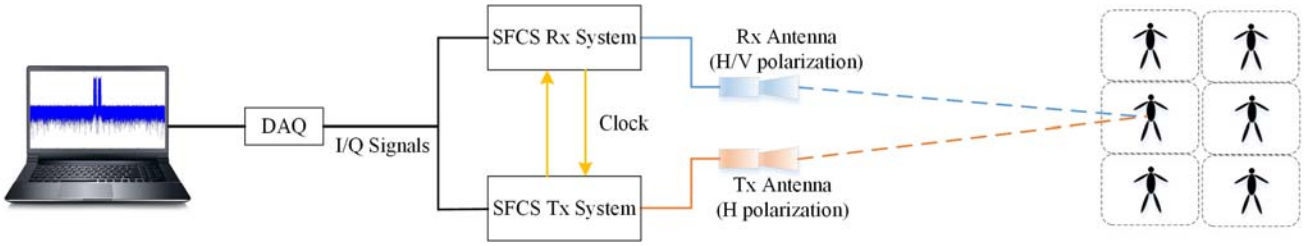


Figure 4. Measurement setup in indoor.

The interval between adjoin subchirps is 60 μs, and the burst repetition frequency (PRF) is 700 Hz. According to [16], we can calculate the maximum unambiguous Doppler velocity at 700 Hz PRF, and it is 3.18 m/s for a 35 GHz carrier frequency, which is an enough range for measuring the resulted velocity from body gestures. The indoor measurement setup is shown in Figure 4.

During the experiment, the four gestures, as shown in Figure 5, were observed at different angles ranging from $-90°$ to $90°$ and different distances ranging from 1 m to 5 m. The viewing angle is defined as the intersection angle of the facing direction and the radar line of sight. The four human body gestures are respectively swinging arm up and down, swinging arm left and right, nodding, and shaking head. A total of 100 experimental data sets were collected for each gesture, and thus 400 experimental data sets were obtained in total.

(a) (b) (c) (d)

Figure 5. Four human body gestures. (a) Swinging arm up and down. (b) Swing arm left and right. (c) Nodding. (d) Shaking head.

4. RECOGNITION RESULTS

4.1. The DCNN Method Recognition Results

TF spectrograms for visualizing the MDSs were obtained by STFT with a time window about 4.3 s. The total time of each collected experimental data is about 20 s, so about four or five TF spectrograms can be obtained for each experimental data. From the collected 400 experimental data sets, 460 images of swinging arm up and down, 432 images of swinging left and right, 440 images of nodding, and 447 images of shaking head are respectively formed. Figures 6(a), (b), and (c) show the TF spectrograms of swinging arm up and down for HH, HV, and the combination of HH and HV, respectively. Figures 7(a), (b), and (c) show the TF spectrograms of swinging arm left and right for HH, HV, and the combination of HH and HV, respectively. Figures 8(a), (b), and (c) show the TF spectrograms of nodding for HH, HV, and the combination of HH and HV, respectively. Figures 9(a), (b), and (c) show the TF spectrograms of shaking head for HH, HV, and the combination of HH and HV, respectively. As we know, every gesture has its MDSs which can be visualized in TF spectrogram. By comparing Figures 6(a) and (b), we can see that the m-D information of HH polarization is weaker than that of HV polarization for swinging arm up and down compared with the torso information. By comparing Figures 7(a) and (b), however, we can see that the m-D information of HH polarization is stronger than that of HV polarization for swinging arm left and right compared with the torso information. Figure 6 and Figure 7 show that the MDSs of swinging arm up and down and swinging arm left and right look like a little bit similar with each other. As we can see from Figure 8 and Figure 9, the MDSs of nodding and shaking head are quite different from each other, as well as quite different from that of swinging arm up and down and swinging arm left and right.

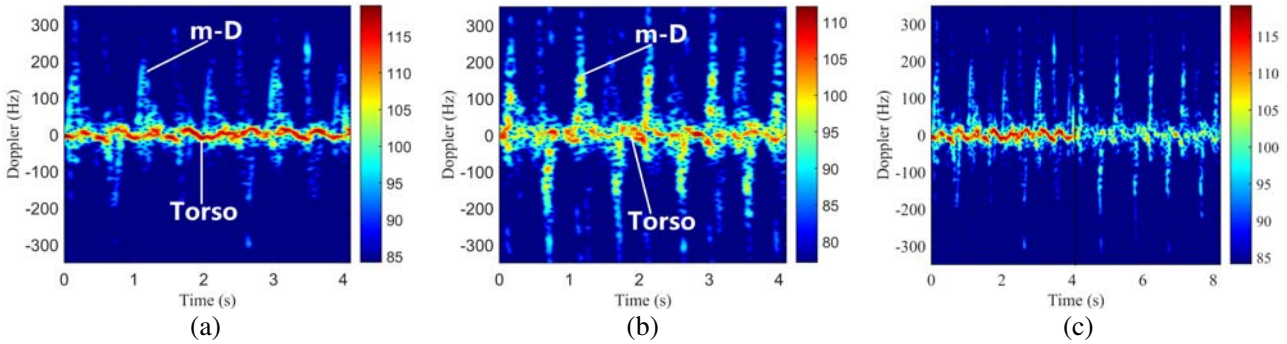


Figure 6. TF spectrograms of swinging arm up and down. (a) HH polarization. (b) HV polarization. (c) The combining of HH polarization and HV polarization.

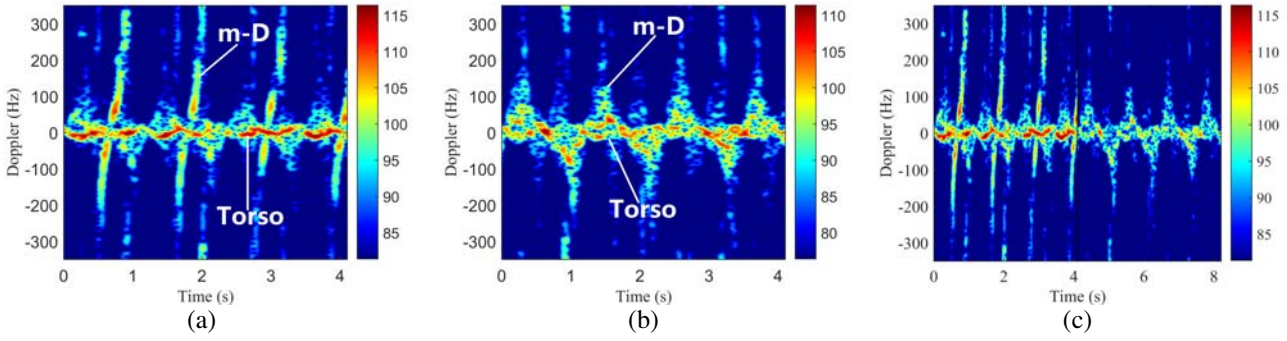


Figure 7. TF spectrograms of swinging arm left and right. (a) HH polarization. (b) HV polarization. (c) The combining of HH polarization and HV polarization.

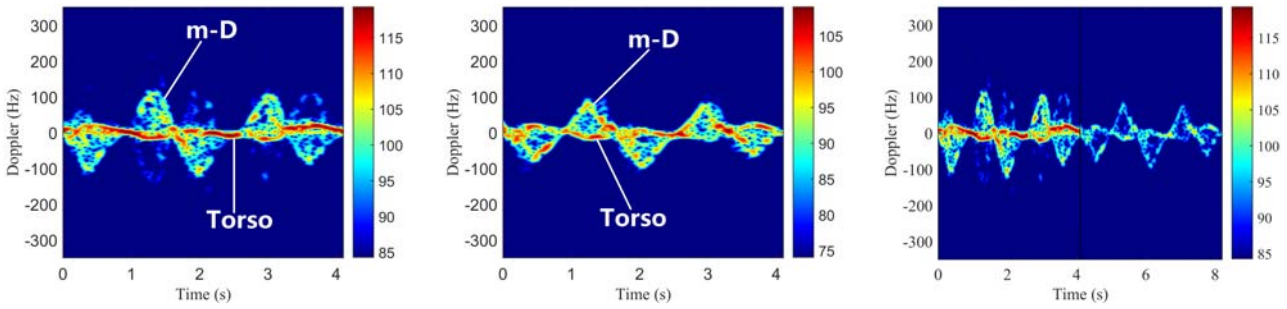


Figure 8. TF spectrograms of nodding. (a) HH polarization. (b) HV polarization. (c) The combining of HH polarization and HV polarization.

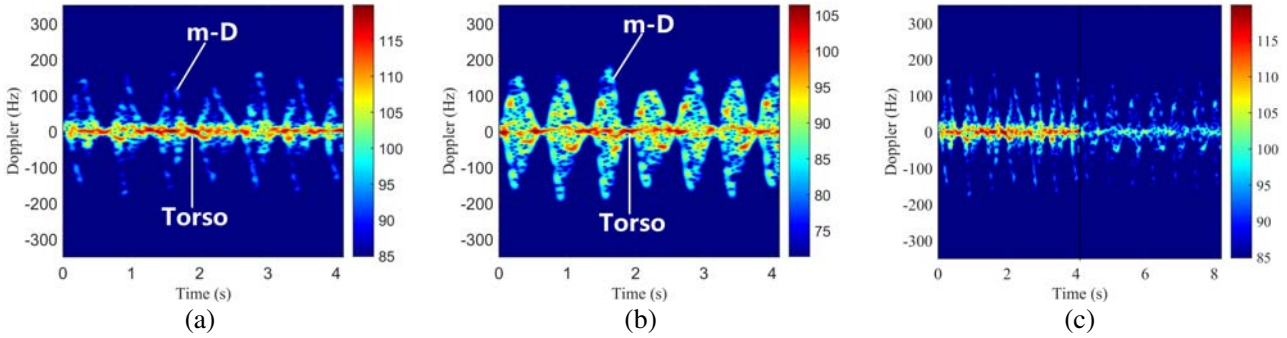


Figure 9. TF spectrograms of shaking head. (a) HH polarization. (b) HV polarization. (c) The combining of HH polarization and HV polarization.

Table 2 presents the recognition results of four body gestures using DCNN, where about 92.7% recognition rate is obtained based on dual-polarization time-frequency spectrograms, while about 77.5% and 89.3% recognition rates are respectively acquired based on time-frequency spectrograms of single HH polarization and single HV polarization. It means that the combination of HH and HV polarizations helps to improve the recognition rate. We can also see that the recognition rate of HV polarization is higher than that of HH polarization, and the reason is that the m-D information of HV polarization is clearer than that of HH polarization respective to the torso information, and thus, the influence of torso is weaker.

Table 2. Recognition results of four body gestures via the proposed method.

Polarization	HH	HV	$HH + HV$
Recognition rate	77.5%	89.3%	92.7%

Furthermore, we selected arbitrary three gestures and conducted the recognition experiment using DCNN with results presented in Table 3. Under circumstance of only considering swinging arm up and down, swinging arm left and right, and nodding, about 92.7% recognition rate is attained based on dual-polarization time-frequency spectrograms, while about 87.0% and 91.2% recognition rates are respectively obtained based on time-frequency spectrograms of single HH polarization and single HV polarization. Under the circumstance of only considering swinging arm up and down, swinging arm left and right, and shaking head, about 91.8% recognition rate is acquired based on dual-polarization

Table 3. Recognition results of arbitrary three body gestures via the proposed method.

Polarization	HH	HV	$HH + HV$
Swing arm up and down, Swing arm left and right, Nodding	87.0%	91.2%	92.7%
Swing arm up and down, Swing arm left and right, Shaking head	81.5%	87.1%	91.8%
Swing arm up and down, Nodding, Shaking head	86.5%	93.9%	97.7%
Swing arm left and right, Nodding, Shaking head	86.7%	96.6%	98.1%

time-frequency spectrograms, while about 81.5% and 87.1% recognition rates are respectively attained based on time-frequency spectrograms of single HH polarization and single HV polarization. Under the circumstance of only considering swinging arm up and down, nodding, and shaking head, about 97.7% recognition rate is achieved based on dual-polarization time-frequency spectrograms, while about 86.5% and 93.9% recognition rates are respectively reached based on time-frequency spectrograms of single HH polarization and single HV polarization. Under the circumstance of only considering swinging arm left and right, nodding, and shaking head, about 98.1% recognition rate is obtained based on dual-polarization time-frequency spectrograms, while about 86.7% and 96.6% recognition rates are respectively acquired based on time-frequency spectrograms of single HH polarization and single HV polarization. The recognition results mean that the combination of HH and HV polarizations is helpful for improving the recognition rate. According to Table 3, we can see that the recognition rate of HV polarization is higher than that of HH polarization, as the recognition results of four body gestures, and the reason is that the m-D information of HV polarization is clearer than that of HH polarization respective to the torso information, and thus, the influence of torso is weaker.

4.2. The SVM Method Recognition Results

In order to demonstrate the superiority of the proposed method compared with traditional machine learning methods, the SVM method is used to process the same experiment data. In this paper, three features, including the variance of the time-frequency spectrogram, the bandwidth of the targets' Doppler modulations in the time-frequency spectrogram, and the variance of the frequencies corresponding to the largest values in each column of the time-frequency spectrogram, are used to classify four body gestures via the SVM method [18]. These three features were also used in [6]. Assuming that $\tilde{S} = \left[\tilde{S}_{f,t}\right]_{f=1,t=1}^{N_F,N_T}$ is the time-frequency spectrogram, where f, t, N_F, and N_T respectively denote the frequency index, time index, frequency range, and time range. The variance of the time-frequency spectrogram can be written as follows

$$\text{Feature}\,(1) = \frac{1}{N_F \times N_T} \sum_{f=1}^{N_F} \sum_{t=1}^{N_T} \left(\tilde{S}_{f,t} - \bar{S}\right) \tag{1}$$

where $\bar{S}$ is the mean of the time-frequency spectrogram. Assume that $E = \left[\sum_{t=1}^{N_T} \tilde{S}_{f,t}\right]_{f=1}^{N_F} / N_T$ denotes the mean vector consisting of mean value of each frequency index; $f_{\max} = \arg\max \mathbf{E}$ is the index of the maximum value in $\mathbf{E}$; $\bar{E}$ is the mean of $\mathbf{E}$. The bandwidth of the targets' Doppler modulations in the time-frequency spectrogram can be expressed as follows

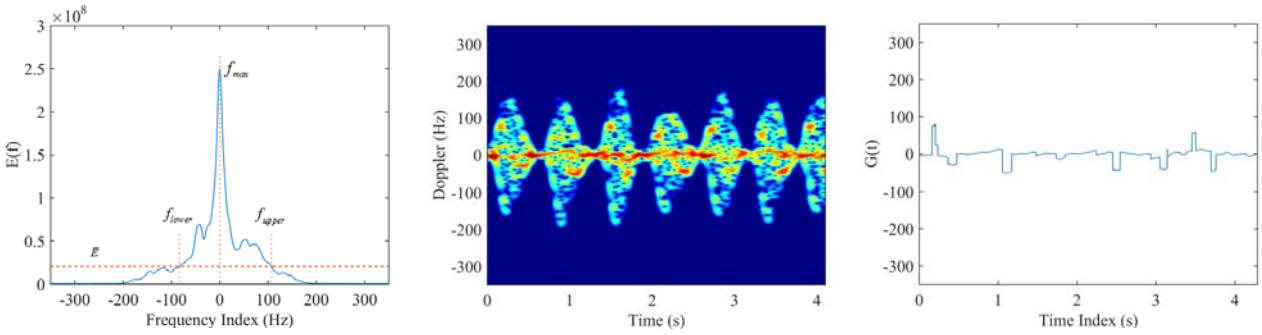
$$\text{Feature}\,(2) = f_{upper} - f_{lower} \tag{2}$$

where f_{upper} and f_{lower} respectively denote the indexes that stratify $E_{f_{upper}} \geq \bar{E}$ and $E_{f_{lower}-1} < \bar{E}$, which are the nearest indexes to $f_{\max}$ within the range $[f_{\max}, N_F]$ and the range $[1, f_{\max}]$. Assume that $G = \left[\arg\ \max_f \left(\left[\tilde{S}_{f,t}\right]_{f=1}^{N_F}\right)\right]_{t=1}^{N_T}$ denotes the frequency index of the largest value of each time index.

The variance of the frequencies corresponding to the largest values in each column of the time-frequency spectrogram can be written as follows

$$\text{Feature}\,(3) = \frac{1}{N_T}\sum_{t=1}^{N_T}\left(G_t - \bar{G}\right) \tag{3}$$

where $\bar{G}$ is the mean of G. Figures 10(a), (b), and (c) show $\mathbf{E}_f$, the time-frequency spectrogram, and $\mathbf{G}_t$, respectively.



(a) (b) (c)

Figure 10. Features extraction. (a) $\mathbf{E}_f$. (b) The time-frequency spectrogram. (c) $\mathbf{G}_t$.

Table 4 presents the recognition results of four body gestures by using SVM method, and the recognition rate is about 30.2% based on dual-polarization data, while the recognition rates are about 22.3% and 25.1%, respectively, based on single HH polarization data and single HV polarization data. It means that combination of HH and HV polarizations also helps to improve the recognition rate. That recognition rates of the SVM method are so low is because three features are largely influenced by the viewing angle and distance.

Furthermore, we selected arbitrary three gestures, conducted the recognition experiment, and list the results in Table 5, from which we can also see that combination of HH and HV polarizations realizes a higher classification accuracy than single HH polarization or single HV polarization, but the recognition rates are still so low as explained for last experiment.

By comparing the results of Table 2 with Table 4 and the results of Table 3 and Table 5, one can clearly see that our DCNN method outperforms the SVM method remarkably.

Table 4. Recognition results of four body gestures via the SVM method.

Polarization	HH	HV	$HH+HV$
Recognition rate	22.3%	25.1%	30.2%

Table 5. Recognition results of arbitrary three body gestures via the SVM method.

Polarization	HH	HV	$HH+HV$
Swing arm up and down, Swing arm left and right, Nodding	30.2%	33.9%	34.0%
Swing arm up and down, Swing arm left and right, Shaking head	30.5%	29.3%	35.4%
Swing arm up and down, Nodding, Shaking head	29.8%	30.6%	37.3%
Swing arm left and right, Nodding, Shaking head	33.4%	32.4%	39.0%

5. CONCLUSION

Four body gestures, i.e., swinging arm up and down, swinging arm left and right, nodding, and shaking head, were observed by a Ka-band dual-polarimetric radar system at different angles and distances, and the corresponding MDSs are obtained by STFT, based on which recognition experiments on different gestures were carried out using DCNN. It is demonstrated that by using the combination of HH and HV polarizations, higher recognition rate can be achieved than that by using single HH or HV polarization. It is clearly shown by the experiment results that our DCNN method outperforms the SVM method. Recognition results of arbitrary three body gestures also support the same conclusion. In the following research, other kinds of gestures will be focused on, e.g., combined gesture of nodding and swinging arms.

ACKNOWLEDGMENT

This work was support in part by the National Natural Science Foundation of China under Grant 41871274.

REFERENCES

1. Tekeli, B., S. Z. Gurbuz, and M. Yuksel, "Information-theoretic feature selection for human micro-Doppler signature classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54, No. 5, 2749–2762, 2016.
2. Garcia-Rubia, J. M., O. Kilic, V. Dang, Q. M. Nguyen, and N. Tran, "Analysis of moving human micro-doppler signature in forest environments," *Progress In Electromagnetics Research*, Vol. 148, 1–14, 2014.
3. Chen, V. C., F. Li, S. S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 42, No. 1, 2–21, 2006.
4. Chen, V. C., F. Li, S. S. Ho, and H. Wechsler, "Analysis of micro-Doppler signature," *IEE Proceeding — Radar Sonar and Navigation*. Vol. 150, No. 4, 271–276, 2003.
5. Li, L., Z. Huang, and W. Zhang, "Instantaneous frequency estimation methods of micro-doppler signal," *Progress In Electromagnetics Research C*, Vol. 58, 125–134, 2015.
6. Kim, Y. and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 47, No. 5, 1328–1337, 2009.
7. Fioranelli, F., M. Ritchie, and H. Griffiths, "Multistatic human micro-Doppler classification of armed/unarmed personnel," *IET Radar Sonar and Navigation*, Vol. 9, No. 7, 857–865, 2015.
8. Kim, Y. and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, Vol. 13, No. 1, 8–12, 2016.
9. Kim, Y. and B. Toomajian, "Hand gesture recognition using micro-doppler signatures with convolutional neural network," *IEEE Access*, Vol. 4, 7125–7130, 2016.
10. Kim, Y. and Y. Li, "Human activity classification with transmission and reflection coefficients of on-body antennas through deep convolutional neural networks," *IEEE Transactions on Antennas and Propagation*, Vol. 65, No. 5, 2764–2768, 2017.
11. Chen, Z., G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, 669–673, 2018.
12. Cao, P., W. Xia, M. Ye, J. Zhang, and J. Zhou, "Radar-ID: human identification based on radar micro-Doppler signatures using deep convolutional neural networks," *IET Radar Sonar and Navigation*, Vol. 12, No. 7, 729–734, 2018.

13. Kang, W., Y. Zhang, and X. Dong, "Polarimetric MDS of pedestrian," *Electronics Letters*, Vol. 54, No. 17, 1051–1053, 2018.
14. Kim, B. K., H. Kang, and S. Park, "Experimental analysis of small drone polarimetry based on micro-Doppler signature," *IEEE Geoscience and Remote Sensing Letters*, Vol. 14, No. 10, 1670–1674, 2017.
15. Available at http://download.tensorflow.org/models/inception_v3_2016_08_28.tar.gz/.
16. Chen, V. C., *The Micro-Doppler Effect in Radar*, Artech House, Boston, 2011.
17. Li, J., P. Wang, Y. Li, Y. Zhao, X. Liu, and K. Luan, "Transfer learning of pre-trained Inception-v3 model for colorectal cancer lymph node metastasis classification," *2018 IEEE Inter. Conf. on Mechatronics and Automation*, 1650–1654, Changchun, China, Aug. 5–8, 2018.
18. Du, L., L. Li, B. Wang, and J. Xiao, "Micro-doppler feature extraction based on time-frequency spectrogram for ground moving targets classification with low-resolution radar," *IEEE Sensors Journal*, Vol. 16, No. 10, 3756–3763, 2016.