Sparsity-Driven Micro-Doppler Feature Extraction for Dynamic Hand Gesture Recognition

Gang Li, *Senior Member, IEEE*, Rui Zhang, Matthew Ritchie, and Hugh Griffiths, *Fellow, IEEE*

***Abstract*—In this paper, a sparsity-driven method of micro-Doppler analysis is proposed for dynamic hand gesture recognition with radar sensors. Firstly, sparse representations of the echoes reflected from dynamic hand gestures are achieved through the Gaussian-windowed Fourier dictionary. Secondly, the micro-Doppler features of dynamic hand gestures are extracted using the orthogonal matching pursuit (OMP) algorithm. Finally, the nearest neighbor classifier is combined with the modified Hausdorff distance to recognize dynamic hand gestures based on the sparse micro-Doppler features. Experiments with real radar data show that 1) the recognition accuracy produced by the proposed method exceeds 96% under moderate noise, and 2) the proposed method outperforms the approaches based on principal component analysis and deep convolutional neural network with small training dataset.**

***Index Terms*—dynamic hand gesture recognition, micro-Doppler analysis, sparse signal representation**

I. Introduction

Dynamic hand gesture recognition has been regarded as an effective approach for human-computer interaction (HCI) [1]. Numerous vision-based methods for dynamic hand gesture recognition have been developed [2]. However, these methods are sensitive to the illumination condition and cannot work in conditions of low visibility. In contrast, a radar sensor is capable of detecting and classifying moving targets independent of light conditions. Recently, radar-based

This work was supported in part by the National Natural Science Foundation of China under Grants 61422110 and 61661130158, and in part by the National Ten Thousand Talent Program of China (Young Top-Notch Talent), and in part by the Royal Society Newton Advanced Fellowship, and in part by Shenzhen Fundamental Research Program, and in part by the Tsinghua National Laboratory for Information Science (TNList), and in part by the Tsinghua University Initiative Scientific Research Program, and in part by the IET A. F. Harvey Prize awarded to Hugh Griffiths in 2013 and the Engineering and Physical Sciences Research Council [EP/G037264/1].Corresponding author: Gang Li. Email: gangli@tsinghua.edu.cn.

Gang Li and Rui Zhang are with the Department of Electronic Engineering, Tsinghua University, Beijing, China. G. Li is also with The Research Institute of Tsinghua University in Shenzhen, Shenzhen, China.

Matthew Ritchie and Hugh Griffiths are with the Department of Electronic and Electrical Engineering, University College London, London WC1E 6BT, UK.

approaches for dynamic hand gesture recognition have attracted much attention [3-8]. In [3], a Doppler radar system is developed for detecting three kinds of dynamic hand gestures. In [4], a portable radar sensor is employed to recognize dynamic hand gestures using application-dependent features and principal component analysis (PCA), and the results illustrate the potential of radar-based dynamic hand gesture recognition for smart home applications. The authors of [5, 6] use a frequency modulated continuous wave (FMCW) radar and analyze the range-Doppler images of dynamic hand gestures of drivers. As presented in [6], radar echoes of dynamic hand gestures contain multiple components with time-varying frequency modulations, which are referred to as micro-Doppler signatures [7-22]. In recent years, the use of micro-Doppler analysis for the hand gesture recognition has attracted growing attention. In [7], the feasibility of hand gestures recognition using micro-Doppler signatures with a deep convolutional neural network (DCNN) is investigated, and the recognition accuracy is found to be 93.1% for seven gestures. In [8], the empirical micro-Doppler features are fed into support vector machine (SVM) to accomplish dynamic hand gesture recognition.

Most micro-Doppler-based methods for human activity classification contain two key phases: 1) feature extraction and 2) classification. In Phase 1), a feature vector, which usually has lower dimension than the raw radar data, is derived from the received signal via certain feature extraction techniques. In [15], some empirical features such as the maximal instantaneous frequency and the period of motion are extracted from the time-frequency spectrogram. The techniques for dimension reduction, including PCA [16-18], empirical mode decomposition [19], linear predictive coding [20] and singular value decomposition [21], have also been employed to extract micro-Doppler features. In Phase 2), the micro-Doppler features extracted in Phase 1) are inputted into a trained classifier to determine the type of the observed human activity. A variety kinds of classifiers, including *k*-nearest neighbor, SVM [15] and Bayes classifier [21], have been used for human activity classification. Recently, DCNN have been used in human activity classification [7, 22], which extracts micro-Doppler features from time-frequency spectrograms using convolutional filters and performs classification via fully connected perceptron functions. The experimental results in existing literatures show that the performances of these classifiers depend on applications, though generally the choice of correct features is more important than which classifier is used.

The sparse signal processing technique [23] provides a new perspective for radar data reduction without compromising performance, and this technique has been used to extract micro-Doppler features of vibrating or rotating targets [24-27]. In [24], the micro-Doppler signatures induced by rotating scatterers in radar imaging applications are extracted by the orthogonal matching pursuit (OMP) algorithm. A pruned OMP algorithm is developed in [25], which achieves the joint estimation of the spatial distribution of the scatterers on the target and the rotational speed of the target. In [26-27], the sparse signal processing technique is combined with the time-frequency analysis to obtain high accuracy of helicopter classification. The methods proposed in [24-27] are based on the analytic expressions of the micro-Doppler signals and cannot be used for dynamic hand gesture analysis, because it is difficult to analytically formulate the radar echoes of dynamic hand gestures. To the best of our knowledge, the combination of sparse signal representation and the micro-Doppler analysis for dynamic hand gesture recognition has not been sufficiently investigated yet.

In this paper, we propose a sparsity-driven method of micro-Doppler analysis for dynamic hand gesture recognition. Firstly, the radar echoes reflected from dynamic hand gestures are mapped into the time-frequency domain through the Gaussian-windowed Fourier dictionary. Then, the micro-Doppler features of the dynamic hand gestures are extracted via the OMP algorithm and fed into the modified-Hausdorff-distance-based nearest neighbor (NN) classifier for recognition. Experiments with real data collected by a K-band radar show that 1) the recognition accuracy produced by the proposed method exceeds 96% under moderate noise, and 2) the proposed method outperforms the PCA-based and the DCNN-based methods in conditions of small training dataset. In addition, the proposed method is expected to achieve real-time processing in practical applications with optimized code and accelerated hardware. The main contribution of this paper lies in the combination of the sparsity-aware feature extraction and the modified-Hausdorff-distance-based classifier for dynamic hand gesture recognition.

The reminder of this paper is organized as follows. The radar data collection is described in Section II. In Section III, we present the details about the sparsity-driven micro-Doppler feature extraction and dynamic hand gesture recognition. In Section IV, the experimental results based on the measured data are provided. Section V presents the conclusion.

II. Measurement of Dynamic Hand Gestures

The data analyzed in this paper are collected by a K-band continuous wave (CW) radar system. The carrier frequency and the base-band sampling frequency are 25 GHz and 1 kHz, respectively. The radar antenna is oriented directly to the human hand at a distance of 0.3 m. The following four dynamic hand gestures are considered: (a) hand rotation, (b) beckoning, (c) snapping fingers and (d) flipping fingers. The illustrations and descriptions of the four dynamic hand gestures are shown in Fig.1 and Table I, respectively. The data are collected from three personnel targets: two males and one female. Each person repeats a particular dynamic hand gesture for 20 times. Each 0.6s time interval containing a complete dynamic hand gesture is recorded as a signal segment. The total number of the signal segments is (4 gestures)×(3 personnel targets)×(20 repeats) = 240.

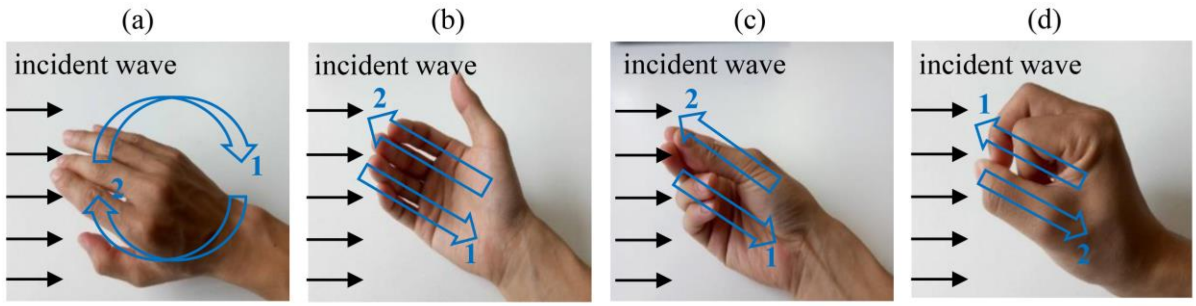

Fig. 1. Illustrations of four dynamic hand gestures: (a) hand rotation; (b) beckoning; (c) snapping fingers; (d) flipping fingers.

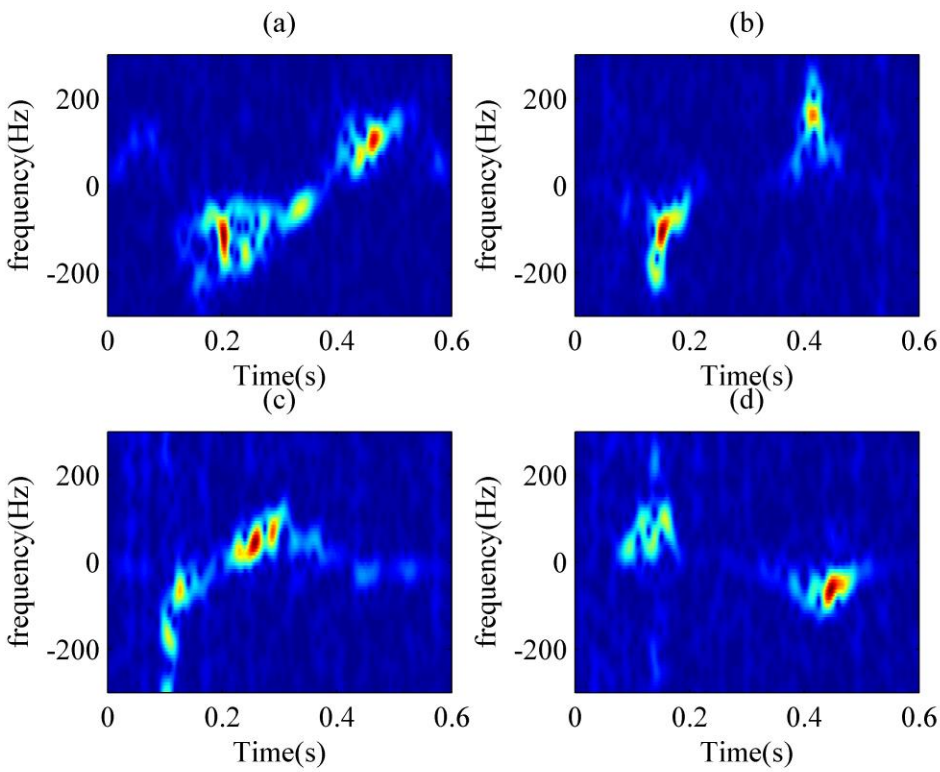

Fig. 2. Spectrograms of received signals corresponding to four dynamic hand gestures from one personnel target: (a) hand rotation; (b) beckoning; (c) snapping fingers; (d) flipping fingers.

To visualize the time-varying characteristics of the dynamic hand gestures, the short time Fourier transform (STFT) with a Kaiser window is applied to the received signals to obtain the corresponding spectrograms. The resulting spectrograms of the four dynamic hand gestures from one personnel target are shown in Fig.2. It is clear that the time-frequency trajectories of these dynamic hand gestures are different from each other. The Doppler shifts corresponding to the gesture ‘hand rotation’ continuously change along the time-axis, because the velocity of the hand continuously changes during the rotation process. The echo of the gesture ‘beckoning’ contains a negative Doppler shift and a positive Doppler shift, which are corresponding to the back and forth movements of the fingers, respectively. The negative Doppler shift of the gesture ‘snapping fingers’ is larger than its positive Doppler shift, since the velocity corresponding to the retreating movement of the fingers is much larger than the velocity corresponding to the returning movement. The time-frequency trajectory of the gesture ‘flipping fingers’ starts with a positive Doppler shift that corresponds to the middle finger flipping towards the radar. The differences among the time-frequency trajectories imply the potential to distinguish different dynamic hand gestures. From Fig. 2 we can also see that, most of the power of the dynamic hand gesture signals is distributed in limited areas in the time-frequency domain. This allows us to use sparse signal processing techniques to extract micro-Doppler features of dynamic hand gestures. Fig. 3 shows the spectrograms of received signals corresponding to dynamic hand gesture ‘hand rotation’ from three personnel targets. It can be seen that the time-frequency spectrograms of the same gesture from different personnel targets have similar patterns.

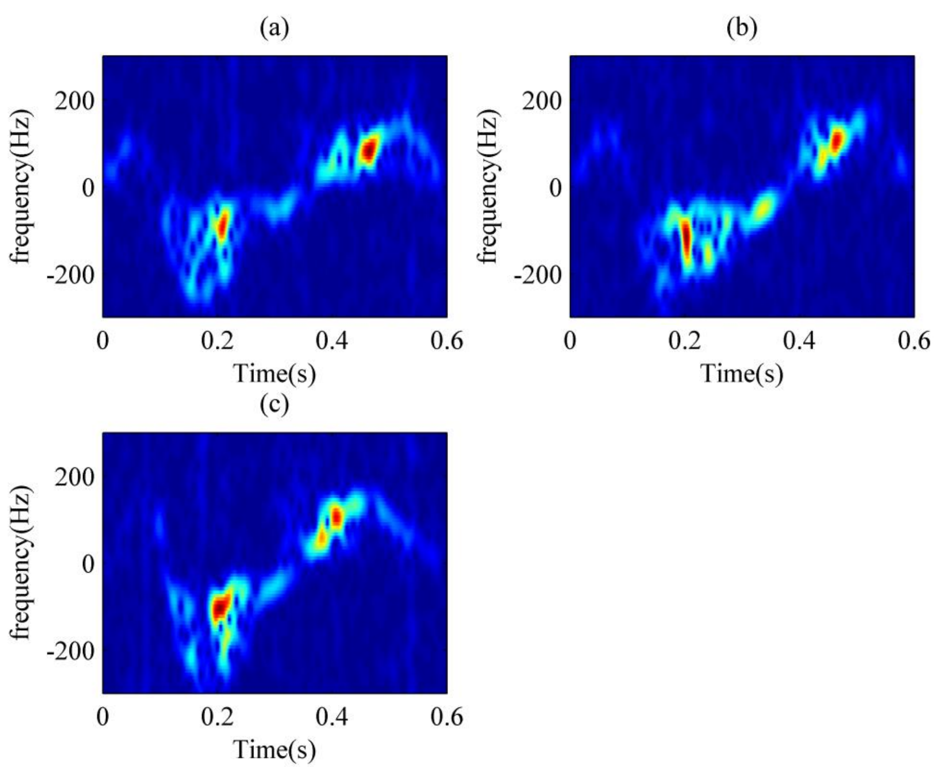

Fig. 3. Spectrograms of received signals corresponding to dynamic hand gesture ‘hand rotation’ from three personnel targets: (a) Target 1; (b) Target 2; (c) Target 3.

TABLE I
FOUR DYNAMIC HAND GESTURES UNDER STUDY

Gesture	Description
(a) Hand rotation	The gesture of rotating the right hand for a cycle. The hand moves away from the radar in the first half cycle and towards the radar in the second half.
(b) Beckoning	The gesture of beckoning someone with the fingers swinging back and forth for one time.
(c) Snapping fingers	The gesture of pressing the middle finger and the thumb together and then flinging the middle finger onto the palm while the thumb sliding forward quickly. After snapping fingers, pressing the middle finger and the thumb together again.
(d) Flipping fingers	The gesture of bucking the middle finger under the thumb and then flipping the middle finger forward quickly. After flipping fingers, bucking the middle finger under the thumb again.

III. SPARSITY-DRIVEN DYNAMIC HAND GESTURE RECOGNITION

The scheme of the proposed method is illustrated in Fig.4. This method contains two sub-processes, i.e., the training process and the testing process. The training process is composed of two steps. Firstly, the time-frequency trajectory of each training signal is extracted using the OMP algorithm. Secondly, the K-means algorithm is employed to cluster the time-frequency trajectories of all training signals and generate the central trajectory corresponding to each dynamic hand gesture. In the testing process, the modified Hausdorff distances [28,29] between the time-frequency trajectory of the testing signal and the central trajectories of dynamic hand gestures are computed and inputted into the nearest neighbor classifier to determine the type of the dynamic hand gesture under test. The details of the proposed method are presented as below.

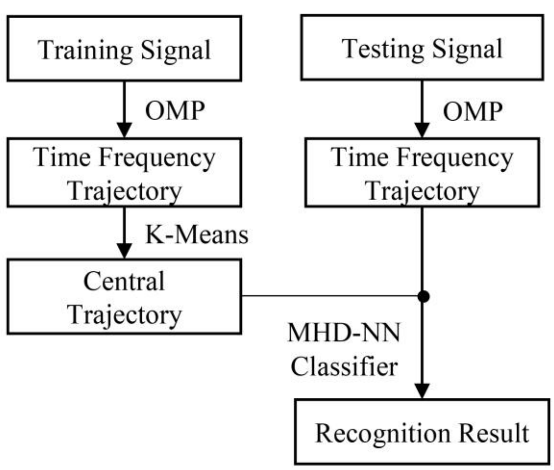

Fig. 4. The scheme the proposed method.

A. Extracting Time-Frequency Trajectory

As discussed in Section II, the time-frequency distributions of the dynamic hand gesture signals are generally sparse. Denoting the received signal as an $N\times 1$ vector $\mathbf{y}$, the model of the sparse representation of $\mathbf{y}$ in time-frequency domain can be expressed as [23],

$$\mathbf{y} = \mathbf{\Phi x} + \boldsymbol{\eta}\ , \tag{1}$$

where $\mathbf{\Phi}$ is an $N\times M$ time-frequency dictionary, $\mathbf{x}$ is an $M\times 1$ sparse vector, and $\boldsymbol{\eta}$ is an $N\times 1$ noise vector. When there are only P non-zero entries in $\mathbf{x}$, $\mathbf{x}$ is called a P-sparse signal. In this paper, we use the Gaussian-windowed Fourier basis signal, which is widely used in time-frequency analysis [30], to generate the dictionary $\mathbf{\Phi}$. The m-th column of dictionary $\mathbf{\Phi}$ can be expressed as

$$\mathbf{\Phi}[:,m] = \left[\phi_{\mathrm{m}}(1), \phi_{\mathrm{m}}(2), \ldots, \phi_{\mathrm{m}}(N)\right]^{\mathrm{T}} \tag{2}$$

where

$$\begin{aligned} \phi_{\mathrm{m}}(n) &\triangleq \phi\left(n \middle| t_m, f_m\right) \\ &= \frac{1}{2^{\frac{1}{4}}\sqrt{\sigma}} \exp\left[-\frac{(n\text{-}t_m)^2}{\sigma^2}\right] \exp\left(-\mathrm{j}2\pi f_m n\right), \\ n &= 1, \ldots, N, \end{aligned} \tag{3}$$

where t_m and f_m represent the time shift and the frequency shift of the basis signal, respectively, σ is the variance of the Gaussian window. As discussed in [30], for a certain variance σ of the Gaussian window, the value sets of the time shift t_m and the frequency shift f_m can be set to be $\{0.5\sigma, \sigma, 1.5\sigma, \ldots, 0.5\sigma \times \lfloor N/(0.5\sigma) \rfloor\}$ and $\{1\pi/\sigma, 2\pi/\sigma, 3\pi/\sigma, \ldots, 2\pi\}$, respectively, where $\lfloor\cdot\rfloor$ is the round down function.

Based on the sparse signal processing technique [23], when $P \ll N < M$, the sparse representation vector $\mathbf{x}$ in (1) can be obtained by

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{\Phi x}\|_2\ ,\ s.t.\ \|\mathbf{x}\|_0 \leq P, \tag{4}$$

where $\|\cdot\|_0$ and $\|\cdot\|_2$ denote the L_0 and L_2 norms, respectively. The solution for (4) can be obtained by greedy algorithms such as the orthogonal matching pursuit algorithm (OMP) [31] or linear programming after replacing L_0 norm with L_1 norm in (4). In this paper, we use the OMP algorithm to solve (4), which first finds the sparse support of $\mathbf{x}$ iteratively and then determines the nonzero coefficients of the sparse solution by the least square estimator. The sparse solution is denoted as

$$\hat{\mathbf{x}} = \mathrm{OMP}(\mathbf{y},\boldsymbol{\Phi},P) = \left(0,\ldots,\hat{x}_{i_1},0,\ldots,\hat{x}_{i_2},0,\ldots,\hat{x}_{i_P},\ldots\right)^{\mathrm{T}}, \tag{5}$$

where $\hat{\mathbf{x}}$ is the P-sparse vector and the non-zero elements are $\hat{x}_{i_p}(p = 1,2,\ldots, P)$.

According to (1), (3) and (5), the received signal $\mathbf{y}$ can be expressed as

$$\begin{array}{rcl} \mathbf{y}(n) & = & \sum_{p=1}^{P} \hat{x}_{i_p} \boldsymbol{\phi}\left(n \,\middle|\, t_{i_p}, f_{i_p}\right) + \boldsymbol{\eta}(n) \\ n & = & 0, 1,\ldots, N\text{-}1 \end{array}. \tag{6}$$

Equation (6) implies that the time-frequency characteristics of $\mathbf{y}$ can be described by a group of basis signals with time-frequency parameters $\left(t_{i_p}, f_{i_p}, \hat{x}_{i_p}\right)$ (p = 1, 2, …, P). Based on this observation, we define the time-frequency trajectory of $\mathbf{y}$ as,

$$\mathrm{T}(\mathbf{y}) = \left\{\left(t_{i_p}, f_{i_p}, A_{i_p}\right), p = 1,2,\ldots, P\right\}. \tag{7}$$

where $A_{i_p} \triangleq \left|\hat{x}_{i_p}\right|$ indicates the intensity at the time-frequency position $\left(t_{i_p}, f_{i_p}\right)$ (p = 1, 2, …, P).

To explain the sparse signal representation clearer, the OMP algorithm is applied to analyze the measured signals presented in Fig.2. The length of each dynamic hand gesture signal is 0.6 s and the sampling frequency is 1 kHz, which means that the value of N is 600. The sparsity P is set to be 10. The variance σ of the Gaussian window is set to be 32. The dictionary $\boldsymbol{\Phi}$ is designed as discussed above and its size is 600×2400. The OMP algorithm is used to solve sparse vector $\hat{\mathbf{x}}$, and then the reconstructed signal is obtained by $\mathbf{y}_{\mathrm{rec}} = \boldsymbol{\Phi}\hat{\mathbf{x}}$. The time-frequency spectrograms of the reconstructed signals $\mathbf{y}_{\mathrm{rec}}$ are plotted in Fig.5. By comparing Figs. 2 and 5, we can find that the reconstructed signals contain the majority part of the original time-frequency features. In addition, it is clear that the noise energy has been significantly suppressed in the reconstructed signals, which is beneficial to dynamic hand gesture recognition. The locations of time-frequency trajectory, i.e. $\left(t_{i_p}, f_{i_p}\right)$ (p = 1, 2, …, P), extracted by the OMP algorithms are plotted in Fig. 6. By comparing Figs. 2 and 6, we can see that the extracted time-frequency trajectories are capable of representing the time-frequency patterns of corresponding dynamic hand gestures.

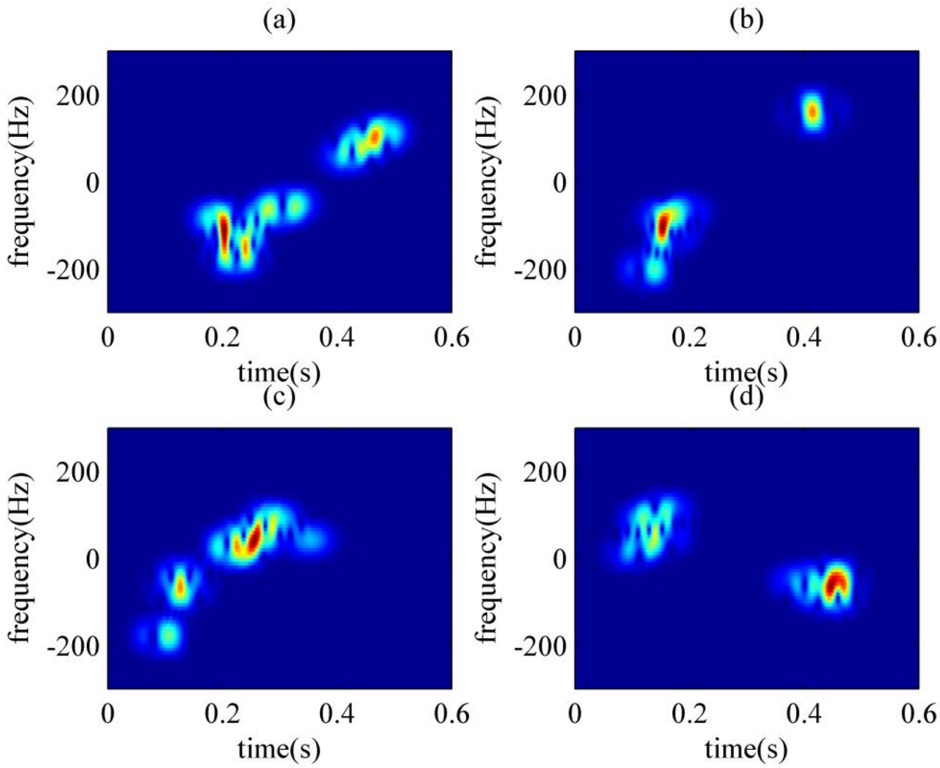

Fig. 5. Spectrograms of reconstructed signals yielded by the OMP algorithm with P=10: (a) hand rotation; (b) beckoning; (c) snapping fingers; (d) flipping fingers.

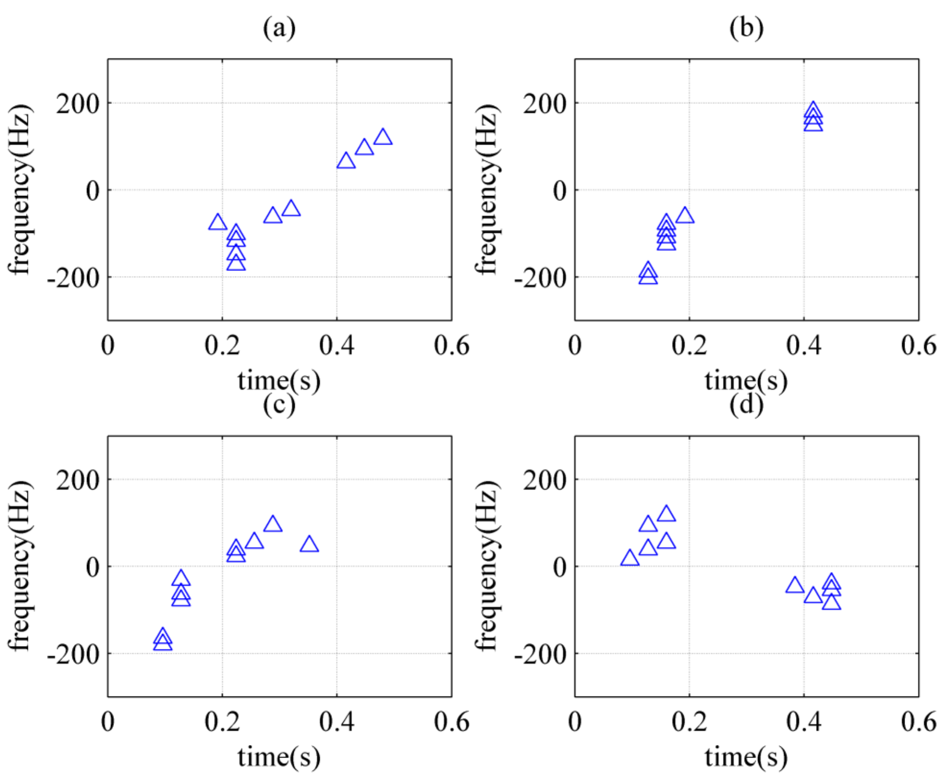

Fig. 6. Locations of time-frequency trajectories $\left(t_{i_p}, f_{i_p}\right)$ (p = 1, 2, …, P) extracted by the OMP algorithms with P=10: (a) hand rotation; (b) beckoning; (c) snapping fingers; (d) flipping fingers.

B. Clustering for Central Time-Frequency Trajectory

In the training process, a central time-frequency trajectory is clustered for each dynamic hand gesture using the *K*-means algorithm based on the time-frequency trajectories of training signals. The details of clustering process are presented as below.

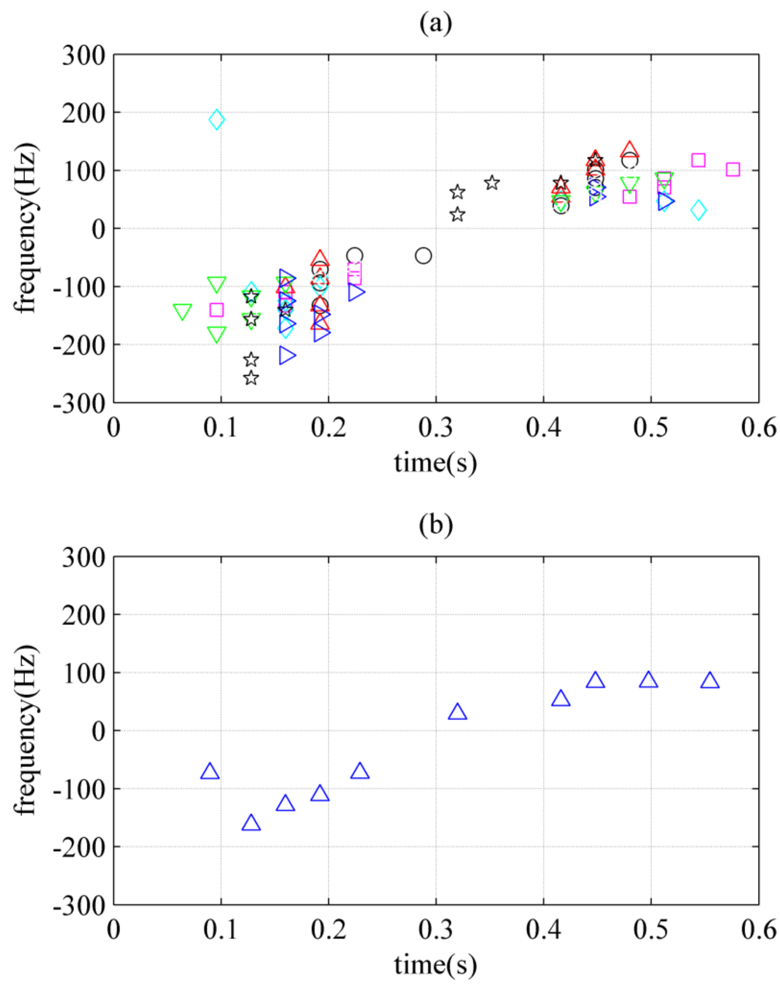

Fig. 7. (a) Locations of time-frequency trajectories extracted from 8 segments of signals corresponding to the gesture ‘snapping fingers’ with *P*=10, where each type of marker indicates the time-frequency trajectory of a certain signal segment. (b) Locations of the clustered time-frequency trajectory generated by the K-means algorithm.

We assume there are *S* segments of training signals for each dynamic hand gesture, denoted as $\mathbf{y}_g^{(s)}$ (s =1,2,…, S), where s and g are the indexes of training segments and dynamic hand gestures, respectively. The time-frequency trajectory of $\mathbf{y}_g^{(s)}$ is denoted as $\mathrm{T}\left(\mathbf{y}_g^{(s)}\right)$, which is composed of *P* time-frequency positions as presented in (7). In ideal case, different realizations of a certain dynamic hand gesture are expected to have the same time-frequency trajectory. However, in realistic scenarios, a human can hardly repeat one dynamic hand gesture in a completely same way.

Therefore, there are minor differences among the time-frequency trajectories extracted from different realizations of one dynamic hand gesture. In order to explain this phenomenon more clearly, we plot the locations of the time-frequency trajectories extracted from 8 segments of signals corresponding to the gesture ‘snapping fingers’ in Fig.7(a) for an example. It is clear that the time-frequency trajectories of different signal segments are distributed closely to each other with slight differences. In order to extract the main pattern from the time-frequency trajectories of training data, the K-means algorithm, which is a clustering technique widely used in pattern recognition [32,33], is employed to generate the central time-frequency trajectory of each dynamic hand gesture. The inputs of the K-means algorithm are the time-frequency positions on the time-frequency trajectories of S training signals and the total number of input time-frequency positions is $P\times S$. With the K-means algorithm, P central time-frequency positions are produced to minimize the mean squared distance from each input time-frequency position to its nearest central position. The K-means algorithm is capable of compressing data and suppressing disturbances while retaining the major pattern of input data. More details about the K-means algorithm can be found in [33]. We denote the central time-frequency trajectory of dynamic hand gesture g generated by the K-means algorithm as:

$$\begin{aligned} \mathrm{T}_{\mathrm{c,g}} &= \text{K-means}\left(\mathrm{T}\left(\mathbf{y}_{\mathrm{g}}^{(1)}\right), \mathrm{T}\left(\mathbf{y}_{\mathrm{g}}^{(2)}\right), \ldots, \mathrm{T}\left(\mathbf{y}_{\mathrm{g}}^{(\mathrm{S})}\right)\right) \\ &= \left\{\left(t_{\mathrm{c,g}}^{(p)}, f_{\mathrm{c,g}}^{(p)}, A_{\mathrm{c,g}}^{(p)}\right),\ p=1,2,\ldots,P\right\} \end{aligned} \tag{8}$$

where $t_{c,\mathrm{g}}^{(p)}$, $f_{c,\mathrm{g}}^{(p)}$, and $A_{c,\mathrm{g}}^{(p)}$ denote the time shift, the frequency shift and the magnitude of the p-th time-frequency position on the central time-frequency trajectory, respectively, and the superscript g is the dynamic hand gesture index. Fig.7(b) shows the location of the central time-frequency trajectory generated by the K-means algorithm using the time-frequency positions in Fig.7(a). It is clear that the majority of time-frequency positions in Fig.7(a) are located around the central time-frequency trajectory in Fig.7(b), which implies that the central time-frequency trajectory is capable of representing the major time-frequency pattern of a dynamic hand gesture.

C. Nearest Neighbor Classifier Based on Modified Hausdorff Distance

In the testing process, the type of dynamic hand gesture corresponding to a given testing signal is determined by the NN classifier. The modified Hausdorff distance, which is widely used in the fields of pattern recognition [28, 29], is used to measure the similarity between the time-frequency trajectory of the testing signal and the central time-frequency trajectory of each dynamic hand gesture.

For a testing signal $\mathbf{y}^{(*)}$, the classification process can be divided into three steps.

Firstly, the time-frequency trajectory $\mathrm{T}(\mathbf{y}^{(*)})$ is extracted using the OMP algorithm as described in Section III-A.

Secondly, the modified Hausdorff distances between $\mathrm{T}(\mathbf{y}^{(*)})$ and the central time-frequency trajectories $\mathrm{T}_{\mathrm{c,g}}$ ($g = 1,2,\ldots,G$) are computed according to the following formula [28],

$$\mathrm{MHD}\left(\mathrm{T}\left(\mathbf{y}^{(*)}\right),\mathrm{T}_{\mathrm{c,g}}\right)=\sum_{\tau^{(*)}\epsilon\mathrm{T}\left(\mathrm{y}^{(*)}\right)}\mathrm{d_H}\left(\tau^{(*)},\mathrm{T}_{\mathrm{c,g}}\right) \tag{9}$$

where $\tau^{(*)}$is an element in set $\mathrm{T}(\mathbf{y}^{(*)})$, i.e., $\tau^{(*)}$is a parameter set composed of the time shift, the frequency shift and the amplitude as described in (7), and $\mathrm{d_H}(\cdot,\cdot)$ represents the Hausdorff distance, which is defined as [29],

$$\mathrm{d_H}\left(\tau^{(*)},\mathrm{T}_{\mathrm{c,g}}\right)=\min_{\tau\epsilon\mathrm{T}_{\mathrm{c,g}}}\left\|\tau^{(*)}-\tau\right\|_2 \tag{10}$$

where τ is an element in set $\mathrm{T}_{\mathrm{c,g}}$. More details about modified Hausdorff distance can be found in [29].

Thirdly, the type of the dynamic hand gesture corresponding to $\mathbf{y}^{(*)}$ is determined by the following NN classifier,

$$g^{(*)}=\underset{g\in\{1,2,\ldots,G\}}{\operatorname{argmin}}\ \mathrm{MHD}\left(\mathrm{T}\left(\mathbf{y}^{(*)}\right),\mathrm{T}_{\mathrm{c,g}}\right) \tag{11}$$

where G represents the total number of dynamic hand gestures and $g^{(*)}$ indexes the recognition result.

IV. Experimental Results

In this section, the real data measured with the K-band radar are used to validate the proposed method in terms of recognition accuracy, which is defined as the proportion of correctly recognized dynamic hand gesture signals among all the testing signals.

A. Analysis about the Recognition Accuracies and the Sparsity

In this experiment, we evaluate the recognition accuracies of the proposed method with different values of sparsity P. The performance of the proposed method is compared with that of the Sparse-SVM method proposed by the same authors [34]. With the Sparse-SVM method, the time-frequency trajectories of the dynamic hand gestures are extracted by the OMP algorithm as described in Section III-A and inputted into SVM for recognition. The sparsity P is varied from 7 to 21 with a step size of 2, and the recognition accuracies are computed using cross-validation. For each value of sparsity P, we randomly select a certain proportion of measured signals for training, and the remaining data are used for testing. The recognition accuracies are averaged over 50 trails with randomly selected training data. The variance σ of the Gaussian window is set to be 32. The recognition accuracies yielded by the proposed method and the Sparse-SVM method using 30% and 70% of data for training are illustrated in Fig. 8, and the confusion matrix yielded by the proposed method with P=17 with 30% training data is presented in Table II.

TABLE II.
CONFUSION MATRIX YIELDED BY THE PROPOSED METHOD WITH P=17 AND 30% OF DATA FOR TRAINING

	Hand rotation	Beckon -ing	Snapping fingers	Flipping fingers
Hand rotation	96.67%	2.32%	0.72%	0
Beckoning	3.21%	95.42%	3.45%	0
Snapping fingers	0.12%	2.26%	95.71%	0
Flipping fingers	0	0	0.12%	100%

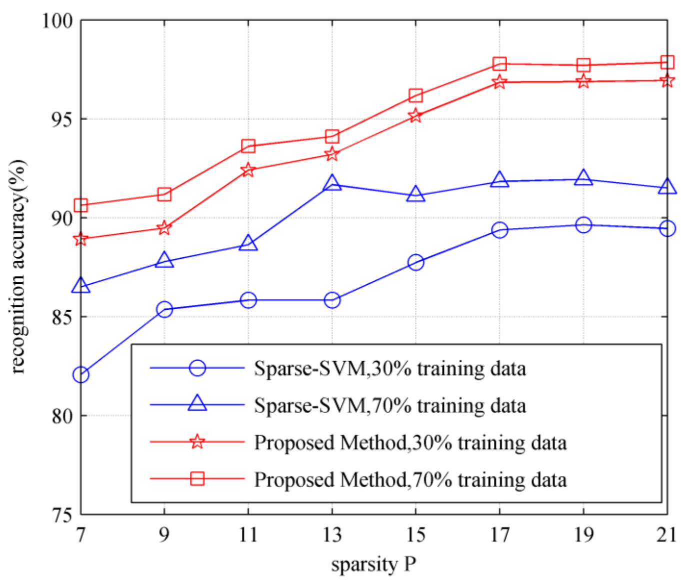

Fig. 8. Recognition accuracies of the proposed method and the Sparse-SVM method versus different values of sparsity P.

It is clear from Fig. 8 that the recognition accuracies of the proposed method increases as the sparsity P increases when $P \leq 15$. This is because more features of the dynamic hand gestures are extracted as the sparsity P increases. The recognition accuracies change slightly as the sparsity P increases when $P \geq 17$. This is because no more useful features can be extracted in this condition. Therefore, the sparsity P is selected to be larger than 15 for the experimental dataset in this paper to achieve satisfying recognition accuracy. If the proposed method is applied to other dataset, the sparsity P should be selected large enough to extract the micro-Doppler features sufficiently. However, a too large value of sparsity P results in more computational burden. In order to determine the proper value of sparsity P, we suggest to employ the training scheme proposed in [35]. With this training scheme, the training dataset is used for selecting the value of sparsity P, i.e., the sparsity-driven method for dynamic hand gesture recognition is evaluated under different values of sparsity P by conducting multi-fold validation within the training dataset off-line, and the value of sparsity P corresponding to the highest recognition accuracy is selected in the final recognition system.

In addition, it can be seen from Fig. 8 that the proposed method outperforms the Sparse-SVM method under each value of sparsity. Furthermore, the recognition accuracies corresponding to 70% of training data are higher than that corresponding to 30% training data.

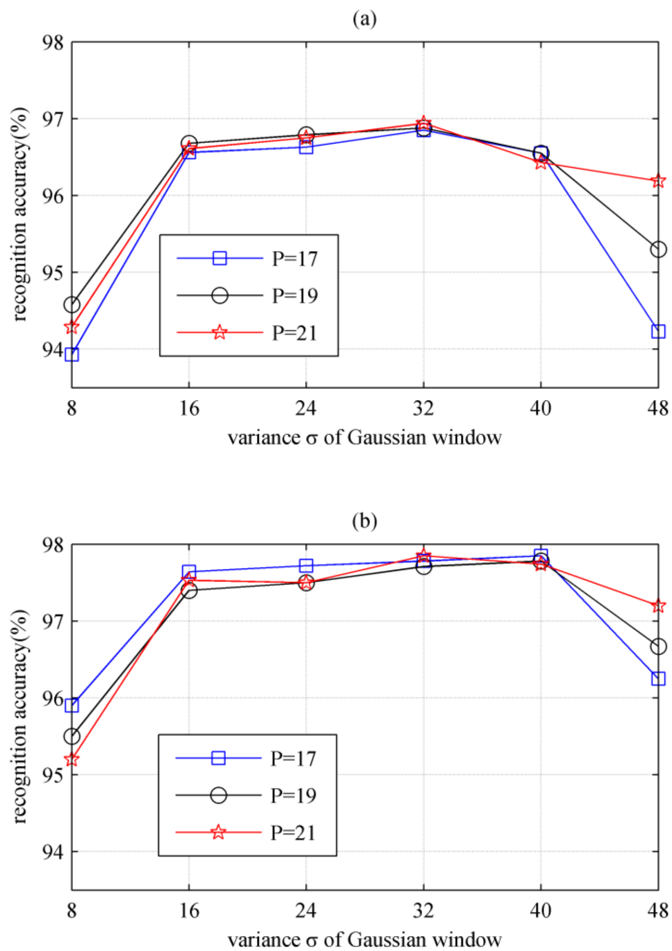

Fig. 9. Recognition accuracies yielded by the proposed method versus different variances of the Gaussian window: (a) using 30% of data for training; (b) using 70% of data for training.

B. Analysis about the Recognition Accuracy and the Variance of the Gaussian Window

In this experiment, the performance of the proposed method versus the window size of the time-frequency dictionary is evaluated under different proportions of training data and different values of sparsity. The variance σ of the Gaussian window is varied from 8 to 48 with a step size of 8. The sparsity P is chosen from {17, 19, 21}, with which the proposed method obtains the highest recognition accuracy when the variance σ of the Gaussian window is 32 as presented in Section IV-A. The recognition accuracies yielded by the proposed method using 30% and 70% of data for training are illustrated in Figs. 9(a) and (b), respectively. It can be seen that the proposed method achieves the best performance and the recognition accuracy changes slightly when the variance σ of the Gaussian window is in [16, 40], which implies that the proposed method is quite robust to the window size of the time-frequency dictionary. The performance of the proposed method declines when the variance σ is less than 16 or larger than 40, this is because the frequency resolution or the time resolution of the Gaussian-windowed Fourier dictionary are poor when the variance of the Gaussian window is too small or too large [30], respectively, which leads to the quality reduction of micro-Doppler feature extraction.

C. Analysis about the Recognition Accuracy and the Size of Training Dataset

In this experiment, the performance of the proposed method is analyzed with different sizes of training dataset. We compare the recognition accuracies yielded by the proposed method with that yielded by the Sparse-SVM method, the PCA-based methods, and the DCNN-based method. With the PCA-based methods, the micro-Doppler features of dynamic hand gestures are obtained by extracting the principal components of the received signals and inputted into SVM for recognition. Two kinds of PCA-based methods are considered here: 1) the PCA in the time-domain, which extracts the features in the time-domain data [18]; 2) the PCA in the time-frequency domain, which extracts the features in the time-frequency domain as presented in [16, 17]. As for the DCNN-based method, the time-frequency spectrograms are fed into a deep convolutional neural network, where the micro-Doppler features are extracted using convolutional filters and the recognition is performed through fully connected perceptron functions. The structure of the DCNN used in this paper is similar with that used in [7]. The proportions of training data are set to be varied from 10% to 90% with a step size of 10%, the sparsity P is set to be 17, and the variance σ of the Gaussian window is set to be 32. The resulting recognition accuracies are shown in Fig. 10, where the proposed method obtains the highest recognition accuracies under different sizes of training set. In addition, the advantages of the proposed method over the PCA-based and the DCNN-based methods are remarkable especially when the proportion of the training data is less than 50%. This implies the proposed method is more applicable than the PCA-based and the DCNN-based methods when the training set is small.

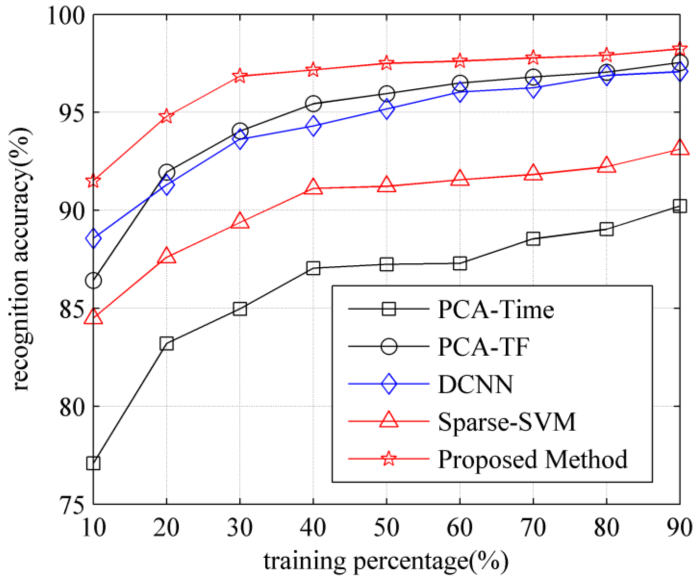

Fig. 10. Recognition accuracies yielded by the proposed method, the Sparse-SVM method, the time-domain PCA-based method (denoted as PCA-Time), the time-frequency domain PCA-based method (denoted as PCA-TF) and the DCNN-based method (denoted as DCNN) under different sizes of training set.

D. Analysis of Recognition Accuracy and Noise Level

In this experiment, the proposed method, the PCA-based and the DCNN-based methods are validated under different levels of additive Gaussian white noise (AWGN). The signal received by the radar are mixed with simulated AWGN according to the following expression,

$$\mathbf{s} = \frac{\mathbf{y}}{\sqrt{\|\mathbf{y}\|_2}} + \lambda \boldsymbol{\varepsilon}, \tag{12}$$

where $\mathbf{y}$ represents the received signal, $\|\mathbf{y}\|_2$ represents the L_2 norm of $\mathbf{y}$, λ is a non-negative amplitude coefficient, and $\boldsymbol{\varepsilon}$ is an AWGN with zero mean and unit variance. According to (12), the ratio between the power of the received signal and the AWGN equals to $1/\lambda^2$ in the mixed signal $\mathbf{s}$. It is worth emphasizing that the signal to noise ratio (SNR) of the mixed signal $\mathbf{s}$ is not $1/\lambda^2$, because the received signal $\mathbf{y}$ also contains noise components as depicted in Fig. 2.

The value of λ is varied from 0 to 1 with a step of 0.2. Under each value of λ, the recognition accuracies yielded by each method are measured by averaging over 100 trials of cross-validations. Here the sparsity P is set to be 17, and the variance σ of the Gaussian window is set to be 32. The experimental results corresponding to 30% and 70% of data for training are depicted in Figs. 11 (a) and (b), respectively. It can be seen that the recognition accuracy yielded by the proposed sparsity-driven method is higher than 90% when the value of λ is less than 1. Moreover, the proposed method outperforms the PCA-based and the DCNN-based methods in conditions of small training dataset under moderate noise.

In addition, the descent speed of recognition accuracies along the λ-axis of the proposed method is faster than that of the PCA-based and the DCNN-based methods, which implies that the performance of the proposed method may be worse than the PCA-based and the DCNN-based methods in conditions of serious noise.

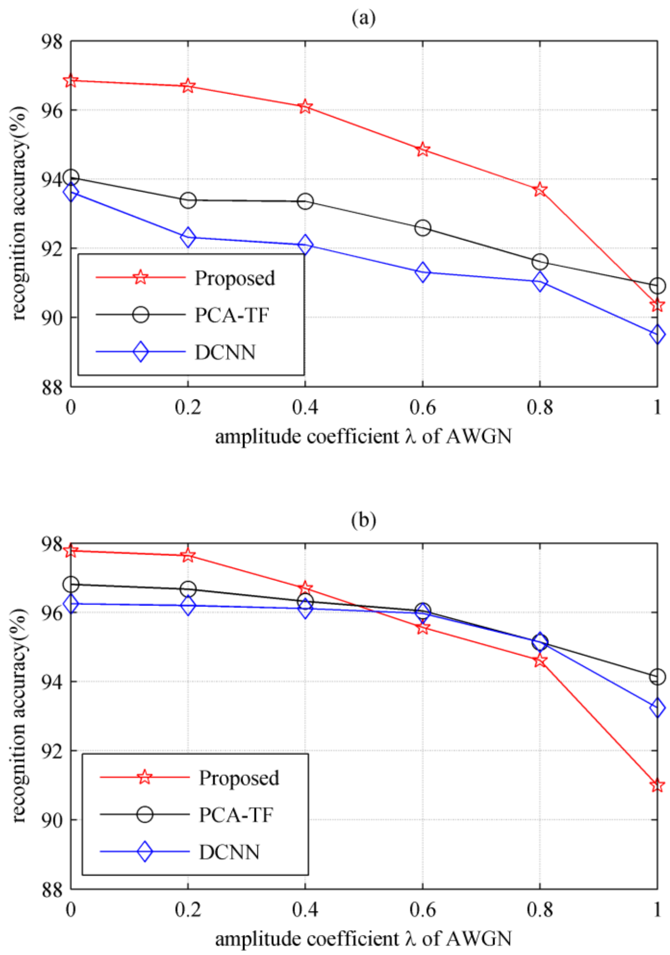

Fig. 11. Recognition accuracies yielded by the proposed method under different levels of additive Gaussian white noise: (a) using 30% of data for training; (b) using 70% of data for training.

E. Recognition Accuracy for Unknown Personnel Targets

As described in Section II, the dynamic hand gesture signals are measured from three personnel targets, denoted as Target 1, 2 and 3, respectively. In Sections IV-A, B and C, the data measured from Target 1, 2 and 3 are mixed together, and a part of the data are used for training and the remaining data are used for testing. In this experiment, the data measured from one of Target 1, 2 and 3 are used for training, and the data measured from the other two personnel targets are used for testing. This experiment aims to validate the proposed method in condition of recognizing the dynamic hand gestures of unknown personnel targets. Cross-validation is employed in this experiment. We randomly select 70% of the data measured from one of Target 1, 2 and 3 for training and 70% of the data measured from the other two personnel targets for testing. The recognition accuracies are averaged over 50 trails with randomly selected training data and testing data. The sparsity P is set to be 17, and the variance σ of the Gaussian window is set to be 32. The resulting recognition accuracies are listed in Table III. It can be seen that the proposed method obtains the highest recognition accuracies under all conditions. This implies that the proposed method is superior to the Sparse-SVM method, the PCA-based methods and the DCNN-based method in terms of recognizing dynamic hand gestures of unknown personnel targets.

TABLE III.
RECOGNITION ACCURACIES FOR UNKNOWN PERSONNEL TARGETS

	Training data from		
	Target 1	Target 2	Target 3
Proposed Method	96.96%	96.88%	95.48%
Sparse-SVM	90.54%	90.54%	88.21%
DCNN	94.38%	95.25%	91.87%
PCA-TF	92.14%	91.80%	91.14%
PCA-Time	84.68%	84.59%	81.07%

F. Analysis about the Time Consumption

In the sparsity-driven method for dynamic hand gesture recognition, the OMP algorithm performs P inner iterations for each received signal to extract the micro-Doppler feature, where P is the sparsity of the received signal. As presented by the experimental results in Section IV-A, the sparsity P can be selected less than 21 to achieve satisfying recognition accuracies. Therefore, the number of inner iterations of the OMP algorithm is less than 21, and the time consumption is controllable.

The computational time consumed by the dynamic hand gesture recognition methods are measured in this subsection. The hardware platform is a laptop with an Intel(R) Core(TM) i5-4200M CPU inside, and the CPU clock frequency and the memory size are 2.5 GHz and 3.7 GB, respectively. The software platform is MATLAB 2014a and the operation system is Windows 10. For each method, the running time for training and classifying are measured by averaging over 100 trials. The results of running time are presented in Table IV. It can be seen that the sparsity-driven method and the DCNN-based method consume the longest time for classifying and training dynamic hand gesture signals, respectively, among all the tested approaches. In realistic applications, the dynamic hand gesture recognition needs to be real-time processing. Considering that the training process can be accomplished off-line, the bottleneck of real-time processing is the time consumption for classifying. Since the running time for classifying one hand gesture by the proposed sparsity-driven method with the non-optimized Matlab code is only 0.22 second, it is promising to achieve real-time processing with optimized code on DSP or GPU platforms in practical applications.

TABLE IV.
TIME CONSUMPTION OF THE DYNAMIC HAND GESTURE RECOGNITION METHODS

	Training time for one sample	Testing time for one sample
Proposed Method	650 ms	220 ms
Sparse-SVM	130 ms	154 ms
DCNN	850 ms	18 ms
PCA-TF	4 ms	11 ms
PCA-Time	0.1 ms	0.5 ms

V. CONCLUSION

In this paper, we have investigated the feasibility and performance of sparsity-driven micro-Doppler extraction method for dynamic hand gesture recognition. Taking advantage of the sparse properties of radar echoes reflected from dynamic hand gestures, the OMP algorithm was used to extract the micro-Doppler features of dynamic hand gesture signals. The extracted features were inputted into modified-Hausdorff-distance-based NN classifier to determine the type of dynamic hand gestures. Real data collected by a K-band CW radar are used to validate the proposed method. Experimental results show that the proposed method obtains recognition accuracy higher than 96% and outperforms the PCA-based and the DCNN-based methods in conditions of small training dataset under moderate noise. In addition, the proposed method is expected to achieve real-time processing with optimized code and accelerated hardware. Application of the proposed method to a larger database with more types of dynamic hand gestures will be included in future work.

REFERENCES

[1] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C* (*Applications and Reviews*), vol. 37, no. 3, pp. 311-324, 2007.

[2] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1-54, 2015.

[3] F. K. Wang, M. C. Tang, Y. C. Chiu, and T. S. Horng, "Gesture Sensing Using Retransmitted Wireless Communication Signals Based on Doppler Radar Technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4592-4602, 2015.

[4] Q. Wan, Y. Li, C. Li, and R. Pal, "Gesture recognition for smart home applications using portable radar sensors," In *Proceeding of 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2014, pp. 6414-6417.

[5] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, “Multi-sensor system for driver's hand-gesture recognition,” In *Proceeding of 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, May 2015, vol. 1, pp. 1-8.

[6] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, “Short-range FMCW monopulse radar for hand-gesture sensing,” In *Proceeding of 2015 IEEE Radar Conference*, May 2015, pp. 1491-1496.

[7] Y. Kim and B. Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network," *IEEE Access*, pp. 7125 - 7130, Oct. 2016.

[8] S. Zhang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, “Dynamic Hand Gesture Classification Based on Radar Micro-Doppler Signatures,” In *Proceedings of 2016 CIE International Conference on Radar*, Oct. 2016, pp. 1977-1980.

[9] L. Yang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, “Gait Classification Based on Micro-Doppler Features,” In *Proceedings of 2016 CIE International Conference on Radar*, Oct. 2016, pp. 1977-1980.

[10] F. Fioranelli, M. Ritchie, S. Gürbüz, and H. D. Griffiths, “Feature diversity for optimized human micro-doppler classification using multistatic radar,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 2, pp. 640-654, 2017.

[11] D. P. Fairchild and R. M. Narayanan, “Multistatic micro-doppler radar for determining target orientation and activity classification,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 512-521, 2016.

[12] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina, “A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal Pseudo-Zernike polynomials,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 417-430, 2015.

[13] T. Thayaparan, S. Abrol, E. Riseborough, L. J. Stankovic, D. Lamothe, and G. Duff, “Analysis of radar micro-Doppler signatures from experimental helicopter and human data,” *IET Radar, Sonar & Navigation*, vol. 1, no. 4, pp. 289-299, 2007.

[14] C. Clemente, A. Balleri, K. Woodbridge, and J. J. Soraghan, “Developments in target micro-Doppler signatures analysis: radar imaging, ultrasound and through-the-wall radar,” *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1-18, 2013.

[15] Y. Kim 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, pp. 1328-1337, 2009.

[16] Q. Wu, Y. D. Zhang, W. Tao, and M. G. Amin, “Radar-based fall detection based on Doppler time–frequency signatures for assisted living,” *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 164-172, 2015.

[17] B. Jokanovic, M. G. Amin, F. Ahmad, and B. Boashash, “Radar fall detection using principal component analysis,” In *SPIE Defense+ Security. International Society for Optics and Photonics*, May 2016, pp. 982919.

[18] A. Balleri, K. Chetty and K. Woodbridge, “Classification of personnel targets by acoustic micro-Doppler signatures,” *IET Radar, Sonar & Navigation*, vol. 5, no. 9, pp. 943-951, 2011.

[19] D. P. Fairchild and R. M. Narayanan, “Classification of human motions using empirical mode decomposition of human micro-Doppler signatures,” *IET Radar, Sonar & Navigation*, vol. 8, no. 5, pp. 425-434, 2014.

[20] R. J. Javier and Y. Kim, “Application of linear predictive coding for human activity classification based on micro-Doppler signatures,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1831-1834, 2014.

[21] F. Fioranelli, M. Ritchie and H. Griffiths, “Classification of unarmed/armed personnel using the NetRAD multistatic radar for micro-Doppler and singular value decomposition features,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933-1937, 2015.

[22] Y. Kim 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, pp. 8-12, 2016.

[23] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.

[24] Y. Luo, Q. Zhang, C. Qiu, S. Li, and T. S. Yeo, “Micro-Doppler feature extraction for wideband imaging radar based on complex image orthogonal matching pursuit decomposition,” *IET Radar, Sonar & Navigation*, vol. 7, no. 8, pp. 914-924, 2013.

[25] G. Li, and P. K. Varshney, “Micro-Doppler parameter estimation via parametric sparse representation and pruned orthogonal matching pursuit,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4937-4948, 2014.

[26] D. Gaglione, C. Clemente, F. Coutts, G. Li, and J. J. Soraghan, “Model-based sparse recovery method for automatic classification of helicopters,” In *Proceeding of 2015 IEEE Radar Conference*, May 2015, pp. 1161-1165.

[27] F. K. Coutts, D. Gaglione, C. Clemente, G. Li, I. K. Proudler, and J. J. Soraghan, “Label Consistent K-SVD for sparse micro-Doppler classification,” In *Proceeding of 2015 IEEE International Conference on Digital Signal Processing*, Jul. 2015, pp. 90-94.

[28] M. P. Dubuisson and A. K. Jain, “A modified Hausdorff distance for object matching,” In *Proceedings of the International Conference on Pattern Recognition (ICPR’94)*, Oct. 1994, pp. 566–568.

[29] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, “Comparing images using the Hausdorff distance,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, 1993.

[30] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397-3415, 1993.

[31] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655-4666, 2007.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin, Germany: Springer Series in Statistics, 2009.

[33] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithm: Analysis and implementation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881-892, 2002.

[34] G. Li, R. Zhang, M. Ritchie, and H. Griffiths, “Sparsity-based Dynamic Hand Gesture Recognition Using Micro-Doppler Signatures”, In *Proceeding of 2017 IEEE Radar Conference*, May 2017, pp. 0928-0931.

[35] P. Molchanov, *Radar target classification by micro-Doppler contributions*, Tampereen teknillinen yliopisto. Julkaisu-Tampere University of Technology. Publication; 1255, 2014.