

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*

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

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

12 I. INTRODUCTION

13 Dynamic hand gesture recognition has been regarded as an effective approach for human-computer interaction (HCI)
14 [1]. Numerous vision-based methods for dynamic hand gesture recognition have been developed [2]. However, these
15 methods are sensitive to the illumination condition and cannot work in conditions of low visibility. In contrast, a radar
16 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.

17 approaches for dynamic hand gesture recognition have attracted much attention [3-8]. In [3], a Doppler radar system is
18 developed for detecting three kinds of dynamic hand gestures. In [4], a portable radar sensor is employed to recognize
19 dynamic hand gestures using application-dependent features and principal component analysis (PCA), and the results
20 illustrate the potential of radar-based dynamic hand gesture recognition for smart home applications. The authors of [5,
21 6] use a frequency modulated continuous wave (FMCW) radar and analyze the range-Doppler images of dynamic hand
22 gestures of drivers. As presented in [6], radar echoes of dynamic hand gestures contain multiple components with
23 time-varying frequency modulations, which are referred to as micro-Doppler signatures [7-22]. In recent years, the use
24 of micro-Doppler analysis for the hand gesture recognition has attracted growing attention. In [7], the feasibility of
25 hand gestures recognition using micro-Doppler signatures with a deep convolutional neural network (DCNN) is
26 investigated, and the recognition accuracy is found to be 93.1% for seven gestures. In [8], the empirical micro-Doppler
27 features are fed into support vector machine (SVM) to accomplish dynamic hand gesture recognition.

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

41 The sparse signal processing technique [23] provides a new perspective for radar data reduction without
42 compromising performance, and this technique has been used to extract micro-Doppler features of vibrating or rotating
43 targets [24-27]. In [24], the micro-Doppler signatures induced by rotating scatterers in radar imaging applications are

44 extracted by the orthogonal matching pursuit (OMP) algorithm. A pruned OMP algorithm is developed in [25], which
45 achieves the joint estimation of the spatial distribution of the scatterers on the target and the rotational speed of the
46 target. In [26-27], the sparse signal processing technique is combined with the time-frequency analysis to obtain high
47 accuracy of helicopter classification. The methods proposed in [24-27] are based on the analytic expressions of the
48 micro-Doppler signals and cannot be used for dynamic hand gesture analysis, because it is difficult to analytically
49 formulate the radar echoes of dynamic hand gestures. To the best of our knowledge, the combination of sparse signal
50 representation and the micro-Doppler analysis for dynamic hand gesture recognition has not been sufficiently
51 investigated yet.

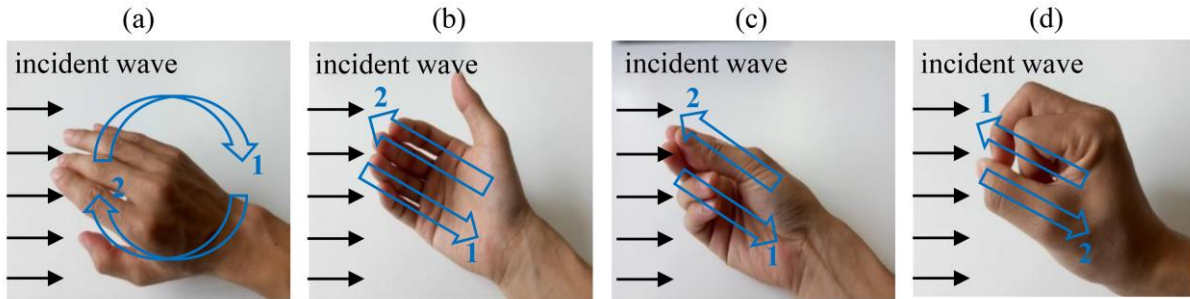
52 In this paper, we propose a sparsity-driven method of micro-Doppler analysis for dynamic hand gesture recognition.
53 Firstly, the radar echoes reflected from dynamic hand gestures are mapped into the time-frequency domain through the
54 Gaussian-windowed Fourier dictionary. Then, the micro-Doppler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

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

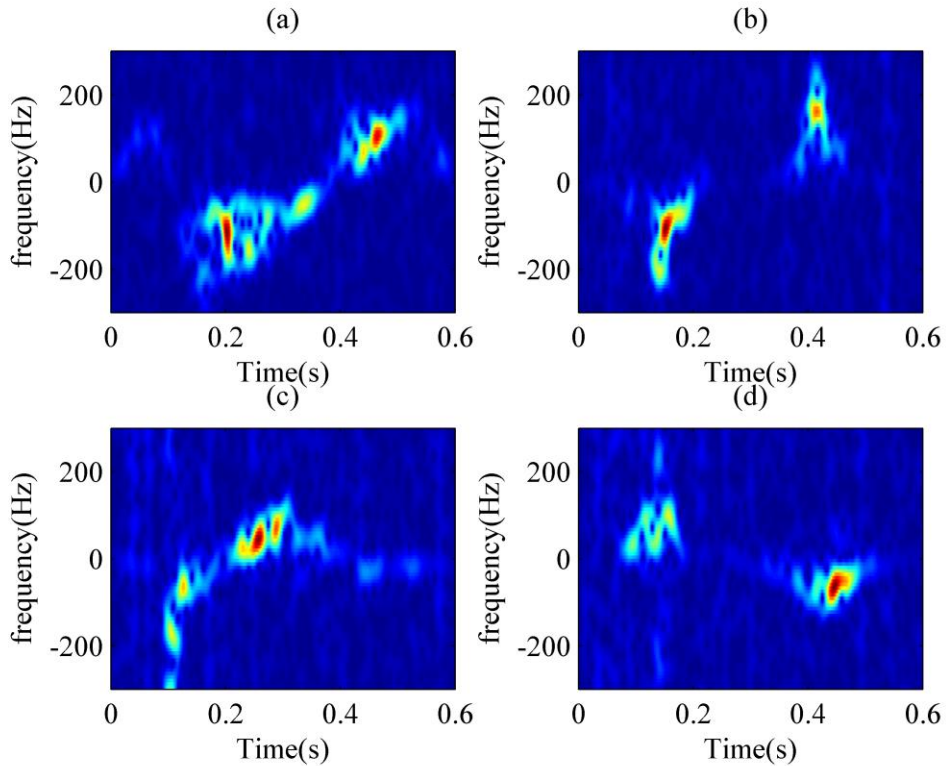
65 II. MEASUREMENT OF DYNAMIC HAND GESTURES

66 The data analyzed in this paper are collected by a K-band continuous wave (CW) radar system. The carrier frequency
67 and the base-band sampling frequency are 25 GHz and 1 kHz, respectively. The radar antenna is oriented directly to the
68 human hand at a distance of 0.3 m. The following four dynamic hand gestures are considered: (a) hand rotation, (b)
69 beckoning, (c) snapping fingers and (d) flipping fingers. The illustrations and descriptions of the four dynamic hand

70 gestures are shown in Fig.1 and Table I, respectively. The data are collected from three personnel targets: two males
71 and one female. Each person repeats a particular dynamic hand gesture for 20 times. Each 0.6s time interval containing
72 a complete dynamic hand gesture is recorded as a signal segment. The total number of the signal segments is (4
73 gestures) \times (3 personnel targets) \times (20 repeats) = 240.



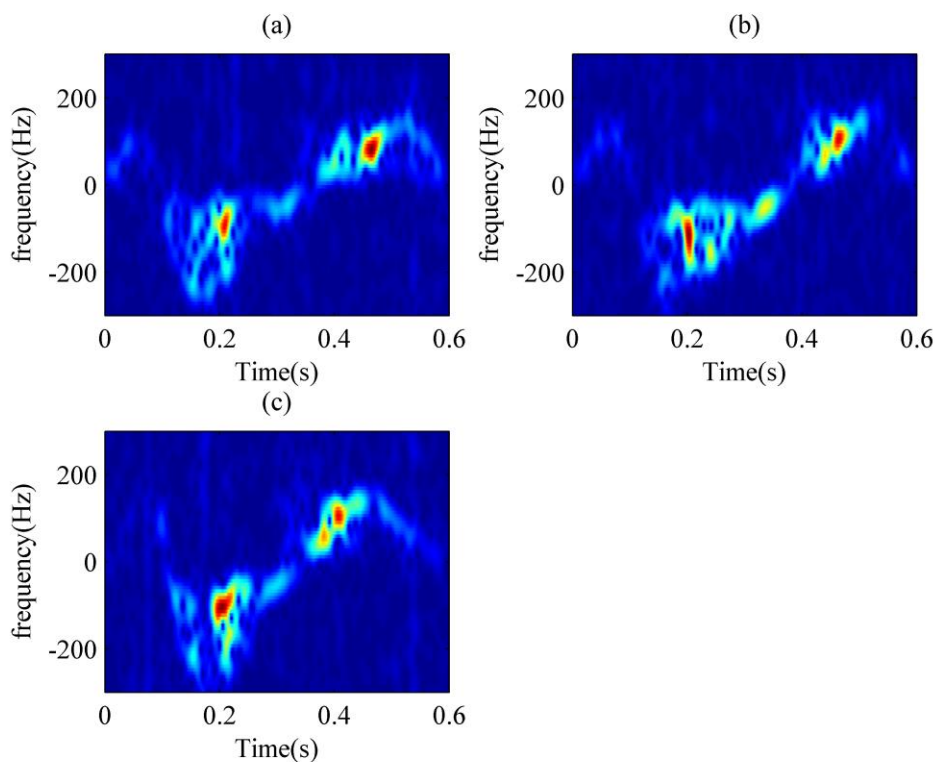
74
75 Fig. 1. Illustrations of four dynamic hand gestures: (a) hand rotation; (b) beckoning; (c) snapping fingers; (d) flipping fingers.



76
77 Fig. 2. Spectrograms of received signals corresponding to four dynamic hand gestures from one personnel target: (a) hand rotation; (b) beckoning;
78 (c) snapping fingers; (d) flipping fingers.

79 To visualize the time-varying characteristics of the dynamic hand gestures, the short time Fourier transform (STFT)
80 with a Kaiser window is applied to the received signals to obtain the corresponding spectrograms. The resulting
81 spectrograms of the four dynamic hand gestures from one personnel target are shown in Fig.2. It is clear that the
82 time-frequency trajectories of these dynamic hand gestures are different from each other. The Doppler shifts

83 corresponding to the gesture ‘hand rotation’ continuously change along the time-axis, because the velocity of the hand
84 continuously changes during the rotation process. The echo of the gesture ‘beckoning’ contains a negative Doppler
85 shift and a positive Doppler shift, which are corresponding to the back and forth movements of the fingers, respectively.
86 The negative Doppler shift of the gesture ‘snapping fingers’ is larger than its positive Doppler shift, since the velocity
87 corresponding to the retreating movement of the fingers is much larger than the velocity corresponding to the returning
88 movement. The time-frequency trajectory of the gesture ‘flipping fingers’ starts with a positive Doppler shift that
89 corresponds to the middle finger flipping towards the radar. The differences among the time-frequency trajectories
90 imply the potential to distinguish different dynamic hand gestures. From Fig. 2 we can also see that, most of the power
91 of the dynamic hand gesture signals is distributed in limited areas in the time-frequency domain. This allows us to use
92 sparse signal processing techniques to extract micro-Doppler features of dynamic hand gestures. Fig. 3 shows the
93 spectrograms of received signals corresponding to dynamic hand gesture ‘hand rotation’ from three personnel targets.
94 It can be seen that the time-frequency spectrograms of the same gesture from different personnel targets have similar
95 patterns.



96
97
98 Fig. 3. Spectrograms of received signals corresponding to dynamic hand gesture ‘hand rotation’ from three personnel targets: (a) Target 1; (b)
99 Target 2; (c) Target 3.

102
103

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.

104

105

III. SPARSITY-DRIVEN DYNAMIC HAND GESTURE RECOGNITION

106

107

108

109

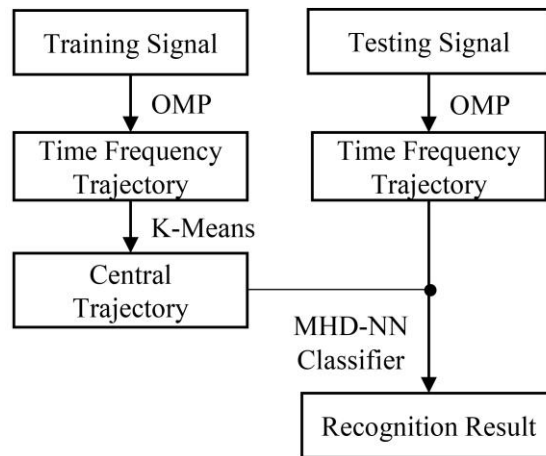
110

111

112

113

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.



114

115

Fig. 4. The scheme the proposed method.

116

117 A. *Extracting Time-Frequency Trajectory*

118 As discussed in Section II, the time-frequency distributions of the dynamic hand gesture signals are generally sparse.
 119 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
 120 can be expressed as [23],

$$\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta}, \quad (1)$$

121 where Φ 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
 122 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
 123 signal, which is widely used in time-frequency analysis [30], to generate the dictionary Φ . The m -th column of
 124 dictionary Φ can be expressed as

$$\Phi[:,m] = [\phi_m(1), \phi_m(2), \dots, \phi_m(N)]^T \quad (2)$$

125 where

$$\begin{aligned} \phi_m(n) &\triangleq \phi(n|t_m, f_m) \\ &= \frac{1}{2^{\frac{1}{4}}\sqrt{\sigma}} \exp\left[-\frac{(n-t_m)^2}{\sigma^2}\right] \exp(-j2\pi f_m n), \\ n &= 1, \dots, N, \end{aligned} \quad (3)$$

126 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
 127 Gaussian window. As discussed in [30], for a certain variance σ of the Gaussian window, the value sets of the time shift
 128 t_m and the frequency shift f_m can be set to be $\{0.5\sigma, \sigma, 1.5\sigma, \dots, 0.5\sigma \times \lfloor N/(0.5\sigma) \rfloor\}$ and $\{1\pi/\sigma, 2\pi/\sigma, 3\pi/\sigma, \dots, 2\pi\}$,
 129 respectively, where $\lfloor \cdot \rfloor$ is the round down function.

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

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

132 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
 133 such as the orthogonal matching pursuit algorithm (OMP) [31] or linear programming after replacing L_0 norm with L_1
 134 norm in (4). In this paper, we use the OMP algorithm to solve (4), which first finds the sparse support of \mathbf{x} iteratively
 135 and then determines the nonzero coefficients of the sparse solution by the least square estimator. The sparse solution is
 136 denoted as

$$\hat{\mathbf{x}} = \text{OMP}(\mathbf{y}, \Phi, P) = (0, \dots, \hat{x}_{i_1}, 0, \dots, \hat{x}_{i_2}, 0, \dots, \hat{x}_{i_P}, \dots)^T, \quad (5)$$

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

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

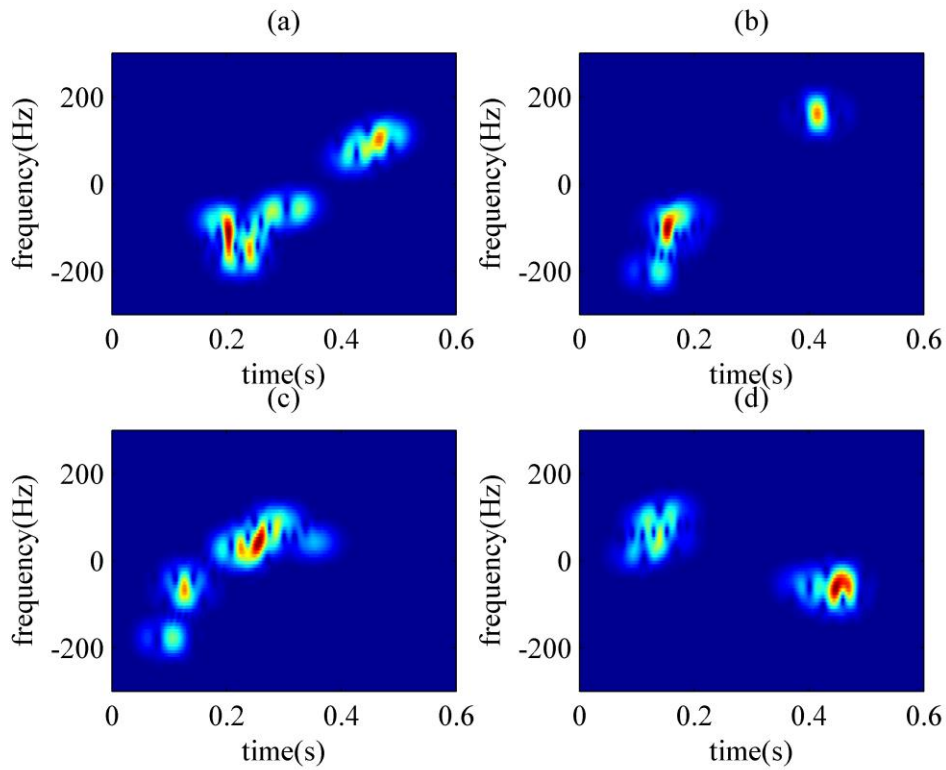
$$\begin{aligned} \mathbf{y}(n) &= \sum_{p=1}^P \hat{x}_{i_p} \phi(n | t_{i_p}, f_{i_p}) + \boldsymbol{\eta}(n) \\ n &= 0, 1, \dots, N-1 \end{aligned} \quad (6)$$

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

$$\mathbf{T}(\mathbf{y}) = \{(t_{i_p}, f_{i_p}, A_{i_p}), p = 1, 2, \dots, P\}. \quad (7)$$

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

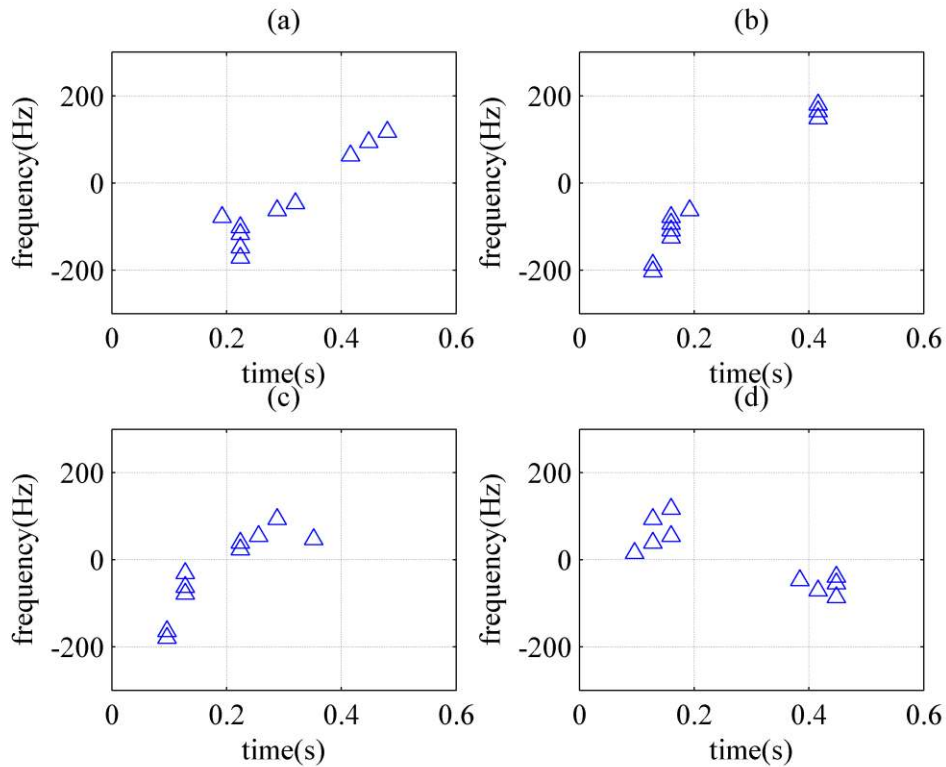
143 To explain the sparse signal representation clearer, the OMP algorithm is applied to analyze the measured signals
 144 presented in Fig.2. The length of each dynamic hand gesture signal is 0.6 s and the sampling frequency is 1 kHz, which
 145 means that the value of N is 600. The sparsity P is set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



154

155
156

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.



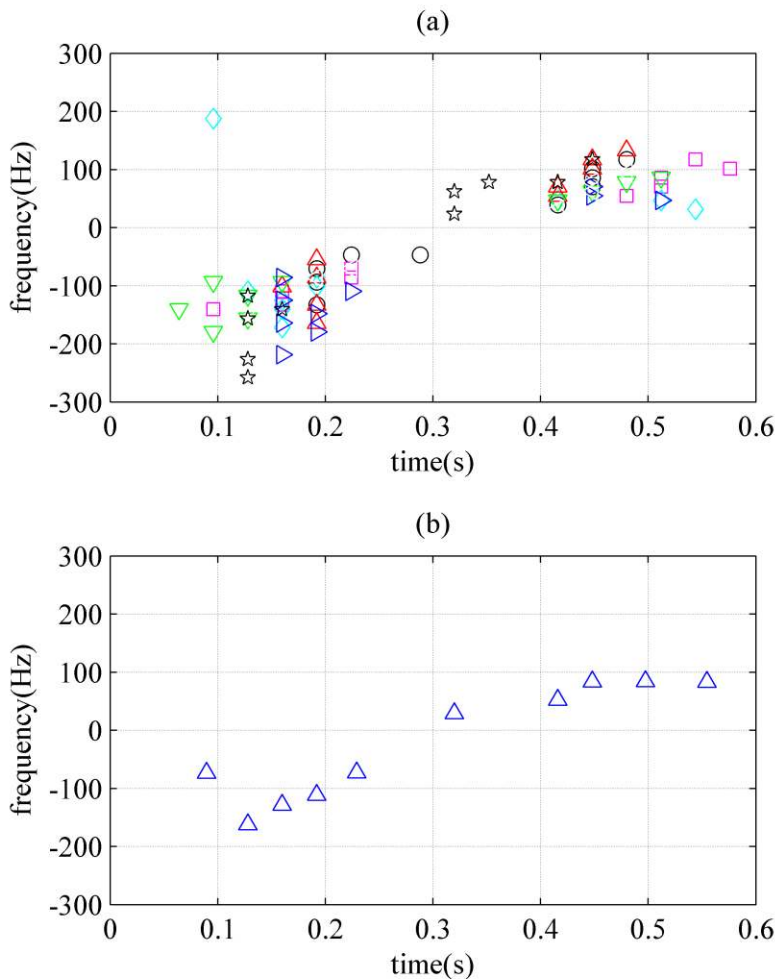
157

158
159
160

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

161 *B. Clustering for Central Time-Frequency Trajectory*

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



165
166 Fig. 7. (a) Locations of time-frequency trajectories extracted from 8 segments of signals corresponding to the gesture ‘snapping fingers’ with
167 $P=10$, where each type of marker indicates the time-frequency trajectory of a certain signal segment. (b) Locations of the clustered
168 time-frequency trajectory generated by the *K*-means algorithm.
169

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

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

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

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

194 *C. Nearest Neighbor Classifier Based on Modified Hausdorff Distance*

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

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

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

201 Secondly, the modified Hausdorff distances between $T(\mathbf{y}^{(*)})$ and the central time-frequency trajectories $T_{c,g}$ ($g =$
202 $1, 2, \dots, G$) are computed according to the following formula [28],

$$\text{MHD}(T(\mathbf{y}^{(*)}), T_{c,g}) = \sum_{\tau^{(*)} \in T(\mathbf{y}^{(*)})} d_H(\tau^{(*)}, T_{c,g}) \quad (9)$$

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

$$d_H(\tau^{(*)}, T_{c,g}) = \min_{\tau \in T_{c,g}} \|\tau^{(*)} - \tau\|_2 \quad (10)$$

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

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

$$g^{(*)} = \underset{g \in \{1, 2, \dots, G\}}{\text{argmin}} \text{MHD}(T(\mathbf{y}^{(*)}), T_{c,g}) \quad (11)$$

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

208 IV. EXPERIMENTAL RESULTS

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

212 A. Analysis about the Recognition Accuracies and the Sparsity

213 In this experiment, we evaluate the recognition accuracies of the proposed method with different values of sparsity P .
214 The performance of the proposed method is compared with that of the Sparse-SVM method proposed by the same
215 authors [34]. With the Sparse-SVM method, the time-frequency trajectories of the dynamic hand gestures are extracted
216 by the OMP algorithm as described in Section III-A and inputted into SVM for recognition. The sparsity P is varied
217 from 7 to 21 with a step size of 2, and the recognition accuracies are computed using cross-validation. For each value of
218 sparsity P , we randomly select a certain proportion of measured signals for training, and the remaining data are used for
219 testing. The recognition accuracies are averaged over 50 trails with randomly selected training data. The variance σ of

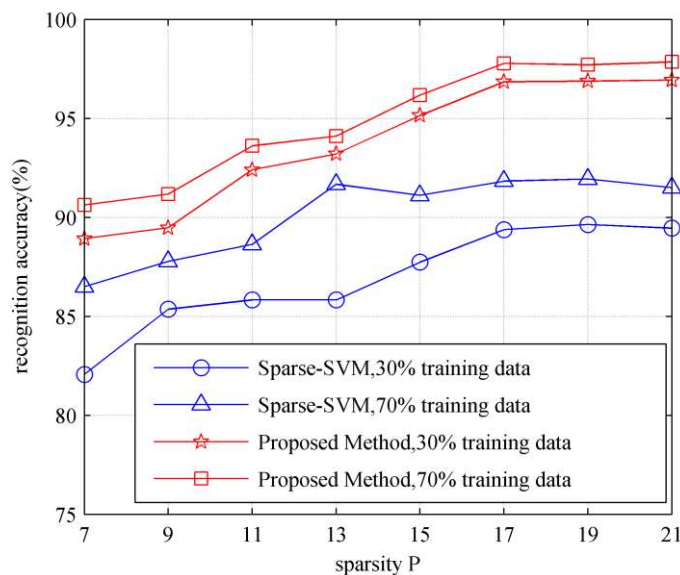
220 the Gaussian window is set to be 32. The recognition accuracies yielded by the proposed method and the Sparse-SVM
 221 method using 30% and 70% of data for training are illustrated in Fig. 8, and the confusion matrix yielded by the
 222 proposed method with $P=17$ with 30% training data is presented in Table II.

223
 224

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%

225



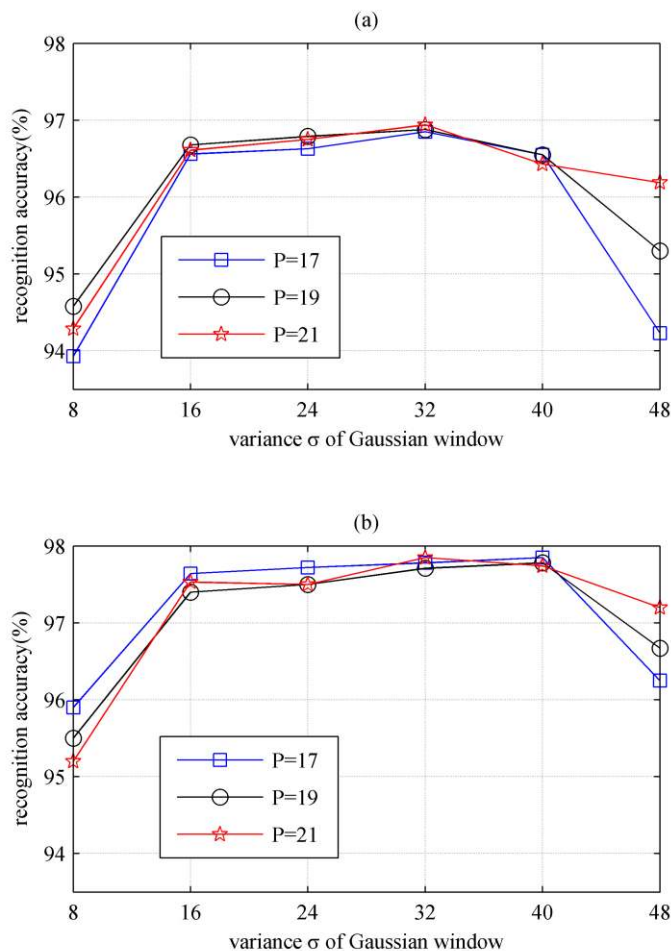
226

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

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

237 values of sparsity P by conducting multi-fold validation within the training dataset off-line, and the value of sparsity P
238 corresponding to the highest recognition accuracy is selected in the final recognition system.

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



242
243 Fig. 9. Recognition accuracies yielded by the proposed method versus different variances of the Gaussian window: (a) using 30% of data for
244 training; (b) using 70% of data for training.

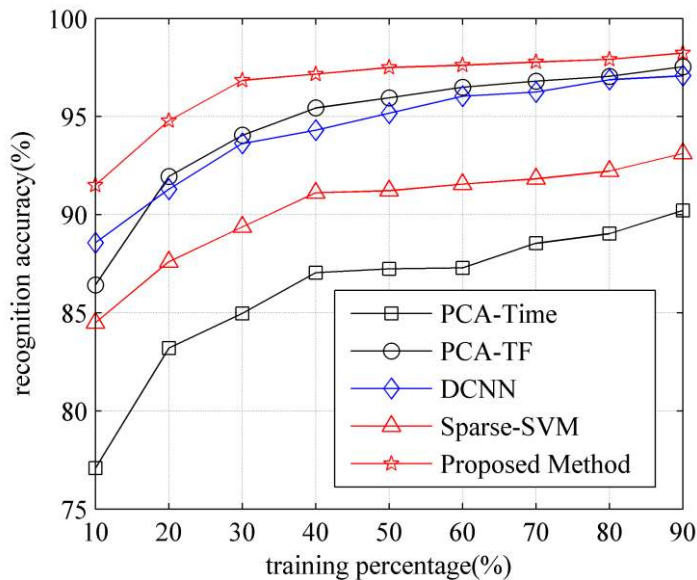
245 B. Analysis about the Recognition Accuracy and the Variance of the Gaussian Window

246 In this experiment, the performance of the proposed method versus the window size of the time-frequency dictionary
247 is evaluated under different proportions of training data and different values of sparsity. The variance σ of the Gaussian
248 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

249 method obtains the highest recognition accuracy when the variance σ of the Gaussian window is 32 as presented in
250 Section IV-A. The recognition accuracies yielded by the proposed method using 30% and 70% of data for training are
251 illustrated in Figs. 9(a) and (b), respectively. It can be seen that the proposed method achieves the best performance and
252 the recognition accuracy changes slightly when the variance σ of the Gaussian window is in [16, 40], which implies
253 that the proposed method is quite robust to the window size of the time-frequency dictionary. The performance of the
254 proposed method declines when the variance σ is less than 16 or larger than 40, this is because the frequency resolution
255 or the time resolution of the Gaussian-windowed Fourier dictionary are poor when the variance of the Gaussian
256 window is too small or too large [30], respectively, which leads to the quality reduction of micro-Doppler feature
257 extraction.

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

259 In this experiment, the performance of the proposed method is analyzed with different sizes of training dataset. We
260 compare the recognition accuracies yielded by the proposed method with that yielded by the Sparse-SVM method, the
261 PCA-based methods, and the DCNN-based method. With the PCA-based methods, the micro-Doppler features of
262 dynamic hand gestures are obtained by extracting the principal components of the received signals and inputted into
263 SVM for recognition. Two kinds of PCA-based methods are considered here: 1) the PCA in the time-domain, which
264 extracts the features in the time-domain data [18]; 2) the PCA in the time-frequency domain, which extracts the features
265 in the time-frequency domain as presented in [16, 17]. As for the DCNN-based method, the time-frequency
266 spectrograms are fed into a deep convolutional neural network, where the micro-Doppler features are extracted using
267 convolutional filters and the recognition is performed through fully connected perceptron functions. The structure of
268 the DCNN used in this paper is similar with that used in [7]. The proportions of training data are set to be varied from
269 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
270 be 32. The resulting recognition accuracies are shown in Fig. 10, where the proposed method obtains the highest
271 recognition accuracies under different sizes of training set. In addition, the advantages of the proposed method over the
272 PCA-based and the DCNN-based methods are remarkable especially when the proportion of the training data is less
273 than 50%. This implies the proposed method is more applicable than the PCA-based and the DCNN-based methods
274 when the training set is small.



275

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

279 *D. Analysis of Recognition Accuracy and Noise Level*

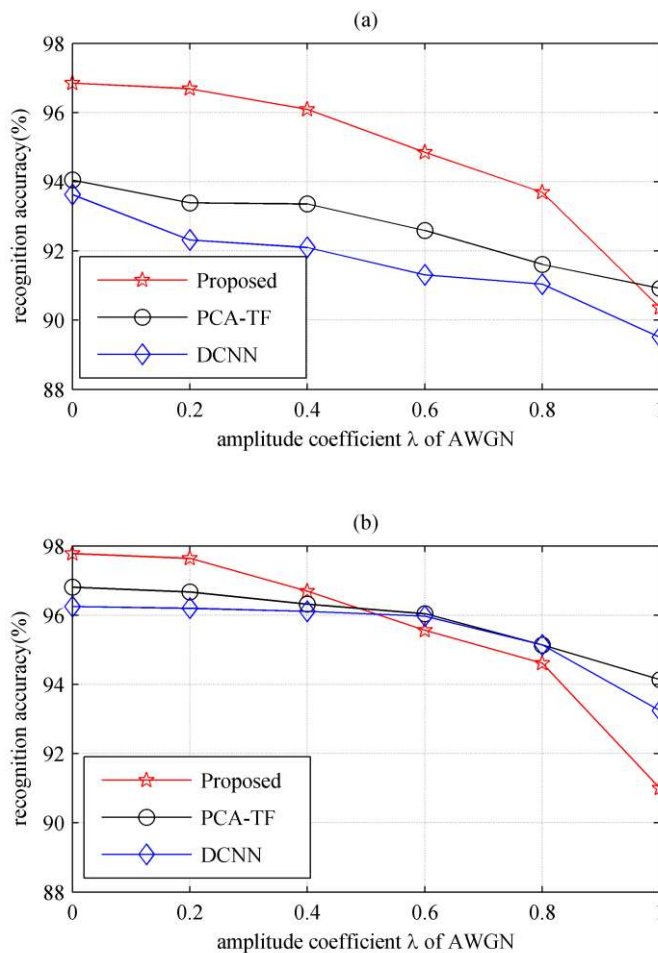
280 In this experiment, the proposed method, the PCA-based and the DCNN-based methods are validated under different
 281 levels of additive Gaussian white noise (AWGN). The signal received by the radar are mixed with simulated AWGN
 282 according to the following expression,

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

283 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
 284 $\boldsymbol{\varepsilon}$ is an AWGN with zero mean and unit variance. According to (12), the ratio between the power of the received signal
 285 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
 286 mixed signal \mathbf{s} is not $1/\lambda^2$, because the received signal \mathbf{y} also contains noise components as depicted in Fig. 2.

287 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
 288 each method are measured by averaging over 100 trials of cross-validations. Here the sparsity P is set to be 17, and the
 289 variance σ of the Gaussian window is set to be 32. The experimental results corresponding to 30% and 70% of data for
 290 training are depicted in Figs. 11 (a) and (b), respectively. It can be seen that the recognition accuracy yielded by the
 291 proposed sparsity-driven method is higher than 90% when the value of λ is less than 1. Moreover, the proposed method
 292 outperforms the PCA-based and the DCNN-based methods in conditions of small training dataset under moderate noise.

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



296
297 Fig. 11. Recognition accuracies yielded by the proposed method under different levels of additive Gaussian white noise: (a) using 30% of data for
298 training; (b) using 70% of data for training.

299 E. Recognition Accuracy for Unknown Personnel Targets

300 As described in Section II, the dynamic hand gesture signals are measured from three personnel targets, denoted as
301 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,
302 and a part of the data are used for training and the remaining data are used for testing. In this experiment, the data
303 measured from one of Target 1, 2 and 3 are used for training, and the data measured from the other two personnel
304 targets are used for testing. This experiment aims to validate the proposed method in condition of recognizing the
305 dynamic hand gestures of unknown personnel targets. Cross-validation is employed in this experiment. We randomly

306 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
307 two personnel targets for testing. The recognition accuracies are averaged over 50 trails with randomly selected
308 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.
309 The resulting recognition accuracies are listed in Table III. It can be seen that the proposed method obtains the highest
310 recognition accuracies under all conditions. This implies that the proposed method is superior to the Sparse-SVM
311 method, the PCA-based methods and the DCNN-based method in terms of recognizing dynamic hand gestures of
312 unknown personnel targets.

313
314

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%

315

316 *F. Analysis about the Time Consumption*

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

322 The computational time consumed by the dynamic hand gesture recognition methods are measured in this subsection.
323 The hardware platform is a laptop with an Intel(R) Core(TM) i5-4200M CPU inside, and the CPU clock frequency and
324 the memory size are 2.5 GHz and 3.7 GB, respectively. The software platform is MATLAB 2014a and the operation
325 system is Windows 10. For each method, the running time for training and classifying are measured by averaging over
326 100 trials. The results of running time are presented in Table IV. It can be seen that the sparsity-driven method and the
327 DCNN-based method consume the longest time for classifying and training dynamic hand gesture signals, respectively,
328 among all the tested approaches. In realistic applications, the dynamic hand gesture recognition needs to be real-time
329 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.

- 357 [5] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," In *Proceeding of 2015 11th*
358 *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, May 2015, vol. 1, pp. 1-8.
- 359 [6] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Short-range FMCW monopulse radar for hand-gesture sensing," In *Proceeding of 2015*
360 *IEEE Radar Conference*, May 2015, pp. 1491-1496.
- 361 [7] Y. Kim and B. Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network," *IEEE*
362 *Access*, pp. 7125 - 7130, Oct. 2016.
- 363 [8] S. Zhang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, "Dynamic Hand Gesture Classification Based on Radar Micro-Doppler
364 Signatures," In *Proceedings of 2016 CIE International Conference on Radar*, Oct. 2016, pp. 1977-1980.
- 365 [9] L. Yang, G. Li, M. Ritchie, F. Fioranelli, and H. Griffiths, "Gait Classification Based on Micro-Doppler Features," In *Proceedings of 2016*
366 *CIE International Conference on Radar*, Oct. 2016, pp. 1977-1980.
- 367 [10] F. Fioranelli, M. Ritchie, S. Gürbüz, and H. D. Griffiths, "Feature diversity for optimized human micro-doppler classification using
368 multistatic radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 2, pp. 640-654, 2017.
- 369 [11] D. P. Fairchild and R. M. Narayanan, "Multistatic micro-doppler radar for determining target orientation and activity classification," *IEEE*
370 *Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 512-521, 2016.
- 371 [12] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina, "A novel algorithm for radar classification based on doppler
372 characteristics exploiting orthogonal Pseudo-Zernike polynomials," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1,
373 pp. 417-430, 2015.
- 374 [13] T. Thayaparan, S. Abrol, E. Riseborough, L. J. Stankovic, D. Lamothe, and G. Duff, "Analysis of radar micro-Doppler signatures from
375 experimental helicopter and human data," *IET Radar, Sonar & Navigation*, vol. 1, no. 4, pp. 289-299, 2007.
- 376 [14] C. Clemente, A. Balleri, K. Woodbridge, and J. J. Soraghan, "Developments in target micro-Doppler signatures analysis: radar imaging,
377 ultrasound and through-the-wall radar," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1-18, 2013.
- 378 [15] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Transactions*
379 *on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328-1337, 2009.
- 380 [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,"
381 *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 164-172, 2015.
- 382 [17] B. Jokanovic, M. G. Amin, F. Ahmad, and B. Boashash, "Radar fall detection using principal component analysis," In *SPIE Defense+*
383 *Security. International Society for Optics and Photonics*, May 2016, pp. 982919.
- 384 [18] A. Balleri, K. Chetty and K. Woodbridge, "Classification of personnel targets by acoustic micro-Doppler signatures," *IET Radar, Sonar &*
385 *Navigation*, vol. 5, no. 9, pp. 943-951, 2011.
- 386 [19] D. P. Fairchild and R. M. Narayanan, "Classification of human motions using empirical mode decomposition of human micro-Doppler
387 signatures," *IET Radar, Sonar & Navigation*, vol. 8, no. 5, pp. 425-434, 2014.
- 388 [20] R. J. Javier and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-Doppler signatures," *IEEE*
389 *Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1831-1834, 2014.
- 390 [21] F. Fioranelli, M. Ritchie and H. Griffiths, "Classification of unarmed/armed personnel using the NetRAD multistatic radar for
391 micro-Doppler and singular value decomposition features," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1933-1937,
392 2015.
- 393 [22] Y. Kim and T. Moon, "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural
394 Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 8-12, 2016.
- 395 [23] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.
- 396 [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
397 orthogonal matching pursuit decomposition," *IET Radar, Sonar & Navigation*, vol. 7, no. 8, pp. 914-924, 2013.
- 398 [25] G. Li, and P. K. Varshney, "Micro-Doppler parameter estimation via parametric sparse representation and pruned orthogonal matching
399 pursuit," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, pp. 4937-4948, 2014.

- 400 [26] D. Gaglione, C. Clemente, F. Coutts, G. Li, and J. J. Soraghan, "Model-based sparse recovery method for automatic classification of
401 helicopters," In *Proceeding of 2015 IEEE Radar Conference*, May 2015, pp. 1161-1165.
- 402 [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
403 classification," In *Proceeding of 2015 IEEE International Conference on Digital Signal Processing*, Jul. 2015, pp. 90-94.
- 404 [28] M. P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," In *Proceedings of the International Conference on*
405 *Pattern Recognition (ICPR '94)*, Oct. 1994, pp. 566-568.
- 406 [29] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Transactions on*
407 *Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, 1993.
- 408 [30] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12,
409 pp. 3397-3415, 1993.
- 410 [31] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on*
411 *Information Theory*, vol. 53, no. 12, pp. 4655-4666, 2007.
- 412 [32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin, Germany:
413 Springer Series in Statistics, 2009.
- 414 [33] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm:
415 Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881-892, 2002.
- 416 [34] G. Li, R. Zhang, M. Ritchie, and H. Griffiths, "Sparsity-based Dynamic Hand Gesture Recognition Using Micro-Doppler Signatures", In
417 *Proceeding of 2017 IEEE Radar Conference*, May 2017, pp. 0928-0931.
- 418 [35] P. Molchanov, *Radar target classification by micro-Doppler contributions*, Tampereen teknillinen yliopisto. Julkaisu-Tampere University
419 of Technology. Publication; 1255, 2014.
420