Sparsity-based Dynamic Hand Gesture Recognition Using Micro-Doppler Signatures

Gang Li, Rui Zhang
Department of Electronic Engineering, Tsinghua University
Beijing, China
gangli@tsinghua.edu.cn, z-r12@mails.tsinghua.edu.cn

Matthew Ritchie, Hugh Griffiths
Department of Electronic and Electrical Engineering,
University College London
London WC1E 6BT, U.K.
{m.ritchie, h.griffiths}@ucl.ac.uk

***Abstract*—In this paper, a sparsity-driven method of micro-Doppler analysis is proposed for dynamic hand gestures recognition with radar sensors. The sparse representation of the radar signals in the time-frequency domain is achieved through the Gabor dictionary, and then the micro-Doppler features are extracted by using the orthogonal matching pursuit (OMP) algorithm and fed into classifiers for dynamic hand gesture recognition. The proposed method is validated with real data measured with a K-band radar. Experiment results show that the proposed method outperforms the principal component analysis (PCA) algorithm, with the recognition accuracy higher than 90%.**

Keywords—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). Numerous vision-based methods for dynamic hand gesture recognition have been developed in the past years [1]. However, these methods are sensitive to the illumination condition and cannot work in conditions of low visibility. In contrast, radar sensor is capable of detecting and classifying moving targets with high robustness to light conditions. Recently, radar-based approaches for dynamic hand gesture recognition have attracted much attention [2-6]. In [2], a Doppler radar system is developed for detecting three kinds of dynamic hand gestures. In [3], a portable radar sensor is employed to recognize dynamic hand gestures by using application-specific 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 [4, 5] model human hand as a non-rigid object and use a frequency modulated continuous wave (FMCW) radar to obtain the range-Doppler images of dynamic hand gestures of drivers' gestures. As presented in [5], radar echoes of dynamic hand gestures contain multiple components with time-varying frequency modulations, which are referred to as micro-Doppler signatures in radar jargon [6-7]. Micro-Doppler effect has been widely used for human activity classification [7-11], but micro-Doppler-based methods for hand gesture recognition have not been sufficiently investigated yet [5].

Most micro-Doppler-based methods for human activity classification contain two key phases: 1) feature extraction and 2) classification. At 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 [8], some empirical features such as the maximal instantaneous frequency and the period of human motion are extracted from the time-frequency spectrum. The techniques for dimension reduction, including PCA, empirical mode decomposition (EMD), linear predictive coding (LPC) and singular value decomposition (SVD) [9-11], have also been employed to extract micro-Doppler features. At 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 of kinds of classifiers, including support vector machine (SVM), Bayes classifier and deep convolutional neural networks [8-11], have been used for human activity classification. The experimental results in existing literatures show that the performances of these classifiers depend on applications.

The sparse signal processing technique provides a new perspective for radar data reduction without compromising performance and has been used to extract micro-Doppler features of vibrating or rotating targets [12-15]. In [12], 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 [13], which achieves the joint estimation of the spatial distribution of the scatterers on the target and the rotational speed of the target. In [14], sparse signal processing technique is combined with the time-frequency analysis to obtain high accuracy of helicopter classification. The methods proposed in [12-14] 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. The radar echoes reflected from dynamic hand gestures are mapped into time-frequency domain through the Gabor dictionary. Then sparse time-frequency features of the dynamic hand gestures are extracted via the OMP algorithm and fed into the SVM classifier for gesture recognition. Experiments with

This work was supported in part by the National Natural Science Foundation of China under Grants 61422110, 41271011 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 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. E-mail: gangli@tsinghua.edu.cn.

real data collected by a K-band radar show that the recognition accuracy produced by the proposed method exceeds 90%, which is higher than that yielded by the PCA-based methods.

The remainder of this paper is organized as follows. The radar data collection of the dynamic hand gestures is described in Section II. In Section III, the sparse representation of the radar echo is formulated and the sparsity-based feature extraction via the OMP algorithm is presented. In Section IV, the experimental results based on the measured data are provided. Section VI presents the conclusion.

II. MEASUREMENT OF DYNAMIC HAND GESTURES

The data analyzed in this paper are collected using 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. Data of four different dynamic hand gestures are collected, such as follows: (a) hand rotation, (b) calling, (c) snapping fingers and (d) flipping finger. An illustration of the performed gestures is shown in Fig.1. The descriptions of each activity are given in Table I. The data are collected from three people, including two males and one female. Each person repeats a particular 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 people)×(20 repeats)=240.

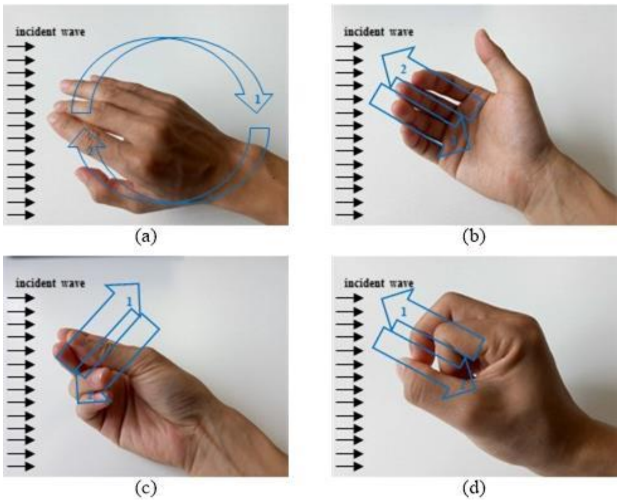

Fig. 1. Illustrations of four different dynamic hand gestures: (a) hand rotation; (b) calling; (c) snapping fingers; (d) flipping finger.

To visualize the time-varying characteristics of the dynamic hand gestures, the short time Fourier transform (STFT) with the Kaiser window is applied to the received signals to obtain the corresponding spectrograms. The resulting spectrograms of the four dynamic hand gestures from one person are shown in Fig. 2. It is clear from Fig. 2 that the time-frequency trajectories of these gestures are different from each other. The Doppler shifts corresponding to the gesture 'hand rotation' change continuously along the time-axis, because the velocity of the hand changes continuously during the rotation process.

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) Calling	The gesture of calling 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 sliming 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.

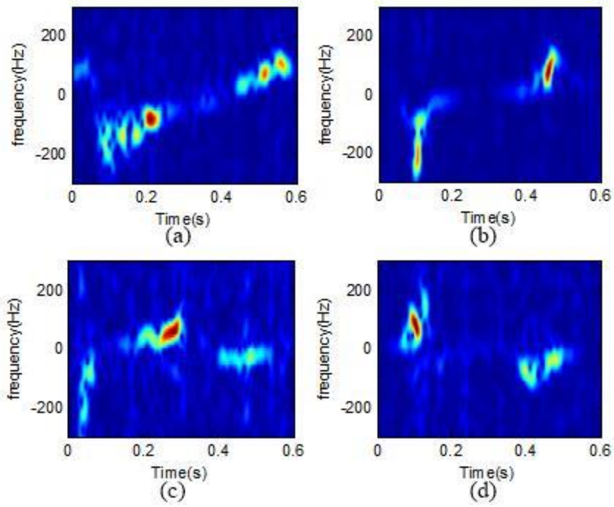

Fig. 2. Spectrograms of received signals corresponding to four dynamic hand gestures: (a) hand rotation; (b) calling; (c) snapping fingers; (d) flipping finger.

The echo of the gesture 'calling' 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 gesture 'snapping fingers' is larger than its positive Doppler shift, since the velocity corresponding to the flinging movement of middle finger is much larger than the velocity corresponding to the regressing movement. The time-frequency trajectory of the gesture 'flipping fingers' starts with a positive Doppler shift that corresponds to the middle finger flipping towards radar. The differences among the above time-frequency trajectories implies the potential to distinguish different dynamic hand gestures.

From Fig. 2 we can also see that, most 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.

III. SPARSITY-BASED MICRO-DOPPLER FEATURE EXTRACTION

In this paper, sparse signal processing technique is applied to extract the micro-Doppler features of dynamic hand gestures. Firstly, the received signal is mapped to a sparse vector in the time-frequency domain by using an over-complete Gabor dictionary. Then, the micro-Doppler features are extracted via the orthogonal matching pursuit (OMP) algorithm.

A. Sparse Representation with Gabor Dictionary

As discussed in Section II, the time-frequency distribution of the radar echo of the dynamic hand gesture is generally sparse. Denoting the received signal as an N×1 vector $\mathbf{y}$, the typical model of the sparse representation of $\mathbf{y}$ in time-frequency domain can be expressed as [12-15]

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

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 K non-zero entries in $\mathbf{x}$, $\mathbf{x}$ is called a K-sparse signal.

In this paper, the Gabor function, which is widely used in time-frequency analysis [16], is used to generate the dictionary $\mathbf{\Phi}$. The elements of the Gabor dictionary $\mathbf{\Phi}$ can be expressed as

$$\begin{aligned}\mathbf{\Phi}(n,m) &= Gabor\left(t_n \mid t_m, f_m, s_m\right)\\ &= \frac{1}{2^{1/4}}\frac{1}{\sqrt{s_m}}\exp\left\{-\left(\frac{t_n - t_m}{s_m}\right)^2\right\}\exp\left(jf_m n\right),\\ &\left(n=1,2,...,N;\ m=1,2,...,M.\right)\end{aligned} \tag{2}$$

where t_m, f_m, and s_m represent the time shift, the frequency shift and the scale factor, respectively, t_n is the n-th sampling instant, and $Gabor(\cdot)$ denotes the Gabor function. It is clear from (2) that each column of $\mathbf{\Phi}$, i.e. $\mathbf{\Phi}(:,m)$, is a Gabor basis signal. As described in [16], the parameters of the Gabor basis signals in dictionary $\mathbf{\Phi}$ are set as

$$\begin{aligned}&\left\{\left(t_m, f_m, s_m\right), m=1,2,...,M\right\}\\ &= \left\{\begin{matrix}\left(p\cdot 2^{j-1}, k\pi\cdot 2^{-j}, 2^j\right), j,p,k\in\mathbb{N},\\ 0<j<\log_2 N,\ 0\le p<N\cdot 2^{-j+1}, 0<k\le 2^{j+1}\end{matrix}\right\}.\end{aligned} \tag{3}$$

In this paper, the signal length N is 600, since the sampling frequency and the time duration of each dynamic hand gesture are 1 kHz and 0.6s, respectively. According to (3), the scale factor s_m of Gabor basis signal can be selected from {2, 4, 8, 16, …, 512}, and the time shift t_m and the frequency shift f_m are respectively selected from $\{0.5s_m, s_m, 1.5s_m, \ldots, 0.5s_m \times\lfloor 600/0.5s_m\rfloor\}$ and $\{\pi/s_m, 2\pi/s_m, 3\pi/s_m, \ldots, 2\pi\}$ under a certain scale factor s_m, where $\lfloor\cdot\rfloor$ is the round down function. Based on the above method, the scale factor s_m is selected from {16, 32} and a Gabor dictionary with size 600×4736 is designed in this paper.

According to the sparse signal processing theories [15], when $K \ll N < M$, the sparse representation vector $\mathbf{x}$ in (1) can be obtained by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2, \text{ s.t. } \|\hat{\mathbf{x}}\|_0 \le K, \tag{4}$$

where $\|\cdot\|_0$ and $\|\cdot\|_2$ denote the L_0 and L_2 norms, respectively. According to previous literatures, the solution for (3) can be obtained by greedy algorithms such as the orthogonal matching pursuit algorithm (OMP) [15] or linear programming after replacing L_0 norm with L_1 norm in (3). In this paper, we use the OMP algorithm to solve (3), 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.

B. Feature Extraction via OMP

Given a segment of the received signal $\mathbf{y}$, OMP algorithm is employed to obtain the sparse representation according to (1) and (3), and the sparse solution is denoted as

$$\hat{\mathbf{x}} = OMP\left(\mathbf{y}, \mathbf{\Phi}, K\right) = \left(0,...,0,x_{i_1},0,...,0,x_{i_2},0,...,0,x_{i_K},0,...,0\right)^T, \tag{5}$$

where $\hat{\mathbf{x}}$ is the K-sparse vector and the non-zero elements are x_{ik} (k = 1,2,…, K). According to (1) (2) and (5), the received signal $\mathbf{y}$ can be expressed as

$$\mathbf{y}\left(t_n\right) = \sum_{k=1}^{K} x_{i_k}\cdot Gabor\left(t_n \mid t_{i_k}, f_{i_k}, s_{i_k}\right) + \boldsymbol{\eta}\left(t_n\right), n=1,2,..,N. \tag{6}$$

Equation (6) implies that the time-frequency characteristics of $\mathbf{y}$ can be described by a group of Gabor basis signals. Inspired by this observation, we use the following feature vector to represent the micro-Doppler features of the received signal:

$$\mathbf{f}\left(\mathbf{y}\right) = \left(t_{i_1}, f_{i_1}, \left|x_{i_1}\right|, t_{i_2}, f_{i_2}, \left|x_{i_2}\right|, \ldots, t_{i_K}, f_{i_K}, \left|x_{i_K}\right|\right). \tag{7}$$

To explain the feature extraction processing clearer, the OMP algorithm is applied to analyze the measured signals presented in Fig.2. The sparsity K is set to be 10. We use the OMP algorithm to solve sparse vector $\hat{\mathbf{x}}$, and then compute the reconstructed signal $\mathbf{y}_{rec}=\mathbf{\Phi}\hat{\mathbf{x}}$. The time-frequency spectrograms of the reconstructed signals $\mathbf{y}_{rec}$ corresponding to different dynamic hand gestures are plotted in Fig. 3. By comparing Fig. 2 and Fig. 3, we can find that the reconstructed signals contain the majority part of the original time-frequency trajectories, which implies that the OMP algorithm succeeds to extract the micro-Doppler features of the dynamic hand gestures. 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 points (t_{ik}, f_{ik}) (k =1, 2,…, K) selected by the OMP algorithms are plotted on Fig. 4. By comparing Figs. 2 and 4, we can find that the selected time-frequency points are capable of representing the time-frequency trajectories of corresponding dynamic hand gestures.

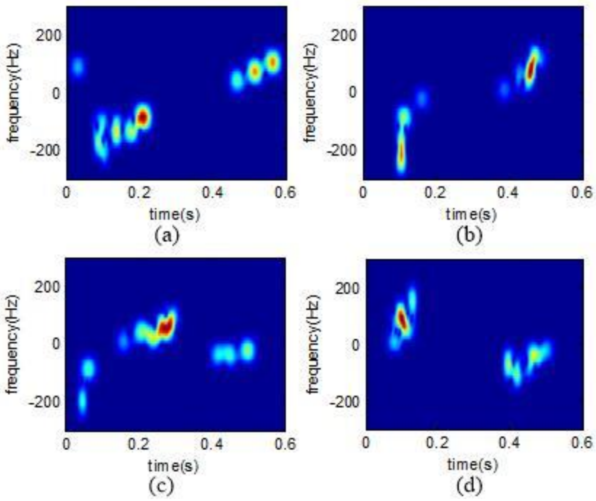

Fig. 3. Spectrograms of reconstructed signals yielded by the OMP algorithm with K=10: (a) hand rotation; (b) calling; (c) snapping fingers; (d) flipping finger.

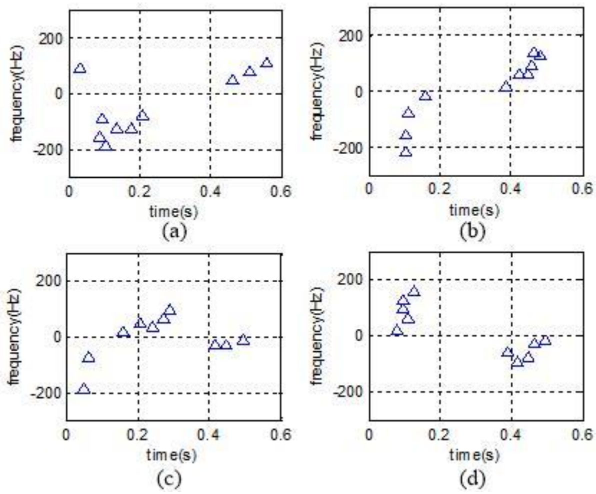

Fig. 4. Locations of Gabor basis signals selected by the OMP algorithms with K=10: (a) hand rotation; (b) calling; (c) snapping fingers; (d) flipping finger.

IV. DYNAMIC HAND GESTURE RECOGNITION

After the micro-Doppler feature extraction, the extracted features are fed into classifiers to determine the type of corresponding dynamic hand gesture. Four kinds of classifiers are considered in this paper, i.e. the naïve Bayes with kernel function estimators (NB), the nearest neighbor (NN), the nearest neighbors with three samples (NN3) and the support vector machine (SVM). These classifiers are described in more detail in [17]. For the training procedure, we use 33.3% data of all 3 testers as the training set, and the remaining 66.7% data as the validation set. The recognition accuracy is calculated by averaging the resulting recognition accuracies of 50 trials of cross validations.

TABLE II RECOGNITION PERFORMANCE

	Sparsity-based	PCA-based
NB	85.24%	75.00%
NN	87.57%	83.33%
NN3	85.39%	85.16%
SVM	91.46%	77.71%

TABLE III
CONFUSION MATRIX YIELDED BY SPARSE-BASED FEATURE EXTRACTION METHOD AND SVM CLASSIFIER.

	Hand rotation	Calling	Snapping fingers	Flipping fingers
Hand rotation	96.67%	6.67%	5.83%	0
Calling	0.83%	86.67%	9.17%	0
Snapping fingers	2.50%	6.67%	80.83%	0
Flipping fingers	0	0	4.17%	1

CONFUSION MATRIX YIELDED BY PCA-BASED FEATURE EXTRACTION METHOD AND NN3 CLASSIFIER.

	Hand rotation	Calling	Snapping fingers	Flipping fingers
Hand rotation	85.95%	28.40%	0.65%	1.90%
Calling	14.05%	66.30%	6.55%	1.50%
Snapping fingers	0	5.05%	92.25%	0.45%
Flipping fingers	0	0.25%	0.55%	96.15%

The performance of the proposed method is compared with that of the PCA-based methods [9]. With the PCA-based method, the micro-Doppler features of dynamic hand gestures are extracted by computing the principal components of the received signals as described in [9]. The feature vectors extracted by the PCA-based method and the proposed sparsity-based method with K=15 are fed into four classifiers, and the resulting recognition accuracies and confusion matrix are shown in TABLE II. For micro-Doppler features extracted by PCA, the highest recognition accuracy is obtained by the NN3 classifier, i.e. 85.16%. For micro-Doppler features extracted by the proposed method, SVM yields the highest recognition accuracy, i.e. 91.46%. It is clear that the proposed method outperforms the PCA-based method in terms of recognition accuracies.

V. CONCLUSION

In this paper, we have investigated the feasibility and performance of recognizing dynamic hand gestures based on micro-Doppler features using sparse signal processing techniques. The radar echoes reflected from dynamic hand gestures are mapped into time-frequency domain through the Gabor dictionary. Then sparse time-frequency features of the radar echo are extracted via the OMP algorithm and fed into four types of classifiers to recognize dynamic hand gestures. Real data of four dynamic hand gestures collected with a K-band CW radar are used to validate the proposed method and the resulting recognition accuracy exceed 90%. Experiment results show that the proposed method obtains higher recognition accuracy than the PCA-based methods.

REFERENCES

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

[2] 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, 2015, vol. 63, no. 12, pp. 4592-4602.

[3] 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, August 2014, pp. 6414-6417.

[4] 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, pp. 1-8.

[5] 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.

[6] 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.

[7] 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, 2013, vol. 2013, no. 1, pp. 1-18.

[8] 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, 2009, vol. 47, no. 5, pp. 1328-1337.

[9] A. Balleri, K. Chetty and K. Woodbridge, "Classification of personnel targets by acoustic micro-Doppler signatures," IET radar, sonar & navigation, 2011, vol. 5, no. 9, pp. 943-951.

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

[11] 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, 2015, vol. 12, no. 9, pp. 1933-1937.

[12] 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, 2013, vol. 7, no. 8, pp. 914-924.

[13] 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, 2014, vol. 7, no. 12, pp. 4937-4948.

[14] 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, pp. 1161-1165.

[15] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Transactions on information theory, 2007, vol. 53, no. 12, pp. 4655-4666.

[16] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Transactions on signal processing, 1993, vol. 41, no. 12, pp. 3397-3415.

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